UNIVERSAL
LIBRARY
OU_158061
UNIVERSAL
LIBRARY

OSMANIA UNIVERSITY LIBRARY

Call No. 519/K34A Accession No. 47236

Author Kendall, Maurice G. and Stuart, A.

Title Advanced Theory of Statistics

d.2d.Vol.1.

This book should be returned on or before the date last marked below.

THE ADVANCED THEORY OF STATISTICS

VOLUME 1

DISTRIBUTION THEORY

OTHER GRIFFIN BOOKS ON STATISTICS, &c.

Combinatorial chance	F. N. DAVID and D. E. BARTON
Games, gods and gambling	F. N. DAVID
A statistical primer	F. N. DAVID
An introduction to the theory of statistics	G. U. YULE and M. G. KENDALL
Rank correlation methods	M. G. KENDALL
Exercises in theoretical statistics	M. G. KENDALL
Rapid statistical calculations	M. H. QUENOUILLE
The design and analysis of experiment	M. H. QUENOUILLE
Sampling methods for censuses and surveys	F. YATES
Biomathematics	C. A. B. SMITH
Statistical method in biological assay	D. J. FINNEY
The mathematical theory of epidemics	N. T. J. BAILEY
Probability and the weighing of evidence	I. J. GOOD

GRIFFIN'S STATISTICAL MONOGRAPHS AND COURSES :

No. 1 : *The analysis of multiple time-series*	M. H. QUENOUILLE
No. 2 : *A course in multivariate analysis*	M. G. KENDALL
No. 3 : *The fundamentals of statistical reasoning*	M. H. QUENOUILLE
No. 4 : *Basic ideas of scientific sampling*	A. STUART
No. 5 : *Characteristic functions*	E. LUKACS
No. 6 : *An introduction to infinitely many variates*	E. A. ROBINSON
No. 7 : *Mathematical methods in the theory of queueing*	A. Y. KHINTCHINE
No. 8 : *A course in the geometry of* n *dimensions*	M. G. KENDALL
No. 9 : *Random wavelets and cybernetic systems*	E. A. ROBINSON
No. 10 : *Geometrical probability*	M. G. KENDALL and P. A. P. MORAN
No. 11 : *An introduction to symbolic programming*	P. WEGNER
No. 12 : *The method of paired comparisons*	H. A. DAVID

THE ADVANCED THEORY OF STATISTICS

MAURICE G. KENDALL, M.A., Sc.D.

Formerly Professor of Statistics in the University of London
President of the Royal Statistical Society, 1960–2

and

ALAN STUART, B.Sc. (ECON.)

Reader in Statistics in the University of London

IN THREE VOLUMES

VOLUME 1

DISTRIBUTION THEORY

SECOND EDITION



CHARLES GRIFFIN & COMPANY LIMITED
LONDON

Copyright © 1963
CHARLES GRIFFIN & COMPANY LIMITED,
42, DRURY LANE, LONDON, W.C.2

THE ADVANCED THEORY OF STATISTICS

Volume 1		Volume 2	
First published ..	1943	First published ..	1946
Second edition ..	1945	Second edition ..	1947
Third ,, ..	1947	Third ,, ..	1951
Fourth ,, ..	1948	Second impression	1955
Fifth ,, ..	1952	Third ,,	1960

Three-volume edition

Volume 1 : *Distribution Theory*
First published .. 1958
Second edition .. 1963

Volume 2 : *Inference and Relationship*
First published .. 1961

Volume 3 : *Planning and Analysis, and Time-Series*

Made and printed in Great Britain by
Butler & Tanner Ltd., Frome and London

" Let us sit on this log at the roadside," says I, " and forget the inhumanity and ribaldry of the poets. It is in the glorious columns of ascertained facts and legalized measures that beauty is to be found. In this very log we sit upon, Mrs. Sampson," says I, " is statistics more wonderful than any poem. The rings show it was sixty years old. At the depth of two thousand feet it would become coal in three thousand years. The deepest coal mine in the world is at Killingworth, near Newcastle. A box four feet long, three feet wide, and two feet eight inches deep will hold one ton of coal. If an artery is cut, compress it above the wound. A man's leg contains thirty bones. The Tower of London was burned in 1841."

" Go on, Mr. Pratt," says Mrs. Sampson. " Them ideas is so original and soothing. I think statistics are just as lovely as they can be."

O. HENRY, *The Handbook of Hymen*

PREFACE TO FIRST EDITION

This book, in its original form, was planned before the outbreak of the Second World War and written mostly during the early years of that troubled period, the first volume being published in 1943 and the second in 1946. Since that time there have been many advances in statistical theory and an expansion in the use of statistical methods which is striking even by comparison with the development of scientific work generally. Five editions of the original Volume 1, and three of Volume 2, have appeared with little more than the correction of errors and the addition of references. But the time has clearly arrived for a complete revision, of which this is the first volume. Owing to the growth of the subject, two further volumes are projected. I regret the size of the new version, but nothing smaller would have covered the field after the manner of my original design.

The revised work falls into three main divisions. The theory of distributions, including sampling distributions, is covered by the present new Volume 1. The second volume will deal with estimation, the general theory of statistical inference, and with statistical relationship. Volume 3 will cover the theory of the design and analysis of sample surveys and experiments, multivariate analysis, and time-series.

A good deal of the material in the original Volume 1 has survived in this volume. Substantial additions, however, have been necessary, notably on standard distributions, on the theory of approximations to sampling distributions, on order-statistics, on multivariate normal distributions, and on distributions associated with the normal distribution. Three of the chapters are new and much new material has been incorporated in the rest of the text. Very considerable additions and modifications will be necessary for the new Volumes 2 and 3, but as it may be several years before the complete work appears I hope that the original Volume 2 will be useful in the meantime.

Many new examples and exercises have been added. There are now over 300 exercises in this volume. Any student who remains unsated after working through them may like to proceed to my *Exercises in Theoretical Statistics*, after which he should be in a position to compose his own.

The references at the end of this volume cover only those articles which are specifically mentioned in the text. Further references will be given in Volumes 2 and 3.

The heavy labour of revision was something I felt unable to face alone, and I have been very fortunate in enlisting the co-operation of my colleague, Mr. Alan Stuart, who has thrown himself into the work with characteristic energy and ability. This revision is to be regarded as of joint authorship ; for all of it we both accept responsibility.

For permission to quote some of the tables at the end of the book we are indebted to Dr. F. N. David, Professor Sir Ronald Fisher, the late Dr. John Wishart, Dr. Frank Yates, Messrs. Oliver & Boyd, and the editors of *Biometrika*. To these and to all the

authors on whose work we have drawn we express our grateful thanks. Our thanks are also due to Mr. James Durbin, who read the galley proofs with great care and made some valuable suggestions for improvement. We are equally indebted to Mr. E. V. Burke, of Charles Griffin and Company, Limited, who took very great pains with the production of the book in all its stages.

Some readers of previous editions went to considerable trouble to draw attention to misprints, imperfections in the presentation, and points where amplification was desirable. We shall continue to be grateful to any reader who will perform the same service for this reset edition.

M. G. K.

LONDON,
May, 1958

PREFACE TO SECOND EDITION

This edition is substantially the same as the first. Minor misprints and errors have been removed, a few improvements have been made in the text, some new exercises have been added and the references brought up to date. We are grateful to readers who have helped us to improve the book in this way.

M. G. K.
A. S.

LONDON,
June, 1962

CONTENTS

Chapter		Page
	Introductory Note	xi
1.	Frequency-distributions	1
2.	Measures of Location and Dispersion	32
3.	Moments and Cumulants	55
4.	Characteristic Functions	94
5.	Standard Distributions—(1)	120
6.	Standard Distributions—(2)	148
7.	The Calculus of Probabilities	180
8.	Probability and Statistical Inference	198
9.	Random Sampling	206
10.	Standard Errors	228
11.	Exact Sampling Distributions	246
12.	Approximations to Sampling Distributions—(1)	274
13.	Approximations to Sampling Distributions—(2)	308
14.	Order-statistics	325
15.	The Multivariate Normal Distribution and Quadratic Forms	347
16.	Distributions Associated with the Normal	369
	Appendix Tables	397
	References	410
	Index	419

INTRODUCTORY NOTE

0.1 The chapter-sections in this work are numbered serially. The serial numbers are prefixed by the number of the chapter in which they occur and are separated therefrom by a period, e.g. **14.13** refers to the thirteenth section of Chapter 14. A similar procedure is followed for tables, equations and exercises, e.g. (7.15) refers to the fifteenth equation of Chapter 7. In cross-references, chapter-sections are denoted by bold-face type, the others by ordinary type.

0.2 References to printed work are given by author's name and date of publication. In the list of references at the end of the book authors are arranged alphabetically. Where articles from publications are referred to, the number of the volume is given in bold-face type and the number of the first page of the article in ordinary type, e.g. *Ann. Math. Statist.*, **10**, 275, refers to the article beginning on page 275 of volume 10 of the *Annals of Mathematical Statistics*. Where an exercise is followed by an author's name and a date, the result given in the exercise appears in the article listed in the references under these particulars.

0.3 The mathematical notation is that in current use, but a few symbols may be explained.

(1) The exclamation mark ! written after an integer denotes the factorial of that integer. Some writers give the symbol a more extended use for non-integral numbers by writing

$$x! = \Gamma(x+1) = \int_0^\infty e^{-t} t^x \, dt.$$

This, of course, accords with the factorial notation, but will not be used in this book.

(2) The use of rules over the tops of expressions to denote square root signs or bracketed terms has been avoided; instead, parentheses or brackets are used where necessary.

(3) The combinatorial sign $\binom{n}{r} = \dfrac{n!}{r!\,(n-r)!}$ is used in place of the older ${}^nC_r$.

(4) The summation sign is written as Σ, e.g. $\sum_{j=1}^{j=n} x_j = x_1 + x_2 + \ldots + x_n$.

The symbol $\sum_{j=1}^{j=n}$ can as a rule be shortened to $\sum_{j=1}^{n}$ and in many cases to $\sum_{j}$ or merely to Σ, the extent of the summation being clear from the context.

(5) The ordinary notation for the Γ-function (given above), the B-function, and the hypergeometric function is used, i.e.

$$B(p,q) = \int_0^1 x^{p-1}(1-x)^{q-1}\,dx = \frac{\Gamma(p)\,\Gamma(q)}{\Gamma(p+q)}$$

and

$$F(\alpha,\beta,\gamma,x) = 1+\frac{\alpha.\beta}{1.\gamma}x+\frac{\alpha(\alpha+1).\beta(\beta+1)}{1.2.\gamma(\gamma+1)}x^2+\frac{\alpha(\alpha+1)(\alpha+2)\,\beta(\beta+1)(\beta+2)}{1.2.3.\gamma(\gamma+1)(\gamma+2)}x^3+\ldots$$

(6) Where the exponent is concise, the exponential function is written as a power of e, for example $e^{\frac{1}{2}x^2}$. But where it is lengthy we use the notation exemplified by $\exp\{-\frac{1}{2}(x^2-2\rho xy+y^2)\}$ instead of $e^{-\frac{1}{2}(x^2-2\rho xy+y^2)}$.

(7) The notation $O(f(n))$ signifies that an expression is of order not exceeding that of $f(n)$, i.e. that the limit of the ratio of the expression to $f(n)$ is a bounded constant C. If $C = 1$, we say that the expression is asymptotically equal to $f(n)$ and use the symbol $\sim$ to express this. If $C = 0$, we write $o(f(n))$ instead of $O(f(n))$ to signify that the order is less than that of $f(n)$.

0.4 In some fields it is useful to preserve a distinction between a quantity in a population and the estimate of that quantity from a sample. Where possible, the former will be denoted by a Greek letter and the latter by a Roman letter, e.g. the product-moment correlation coefficient of a population is denoted by ρ and that of a sample by r. It is not, however, always possible to preserve this distinction, as for instance with the multiple correlation coefficient R, in which case a Greek capital would be confused with the Roman P. Complete notational consistency can only be achieved at the expense of jettisoning a great deal of accepted statistical usage, and even then would probably result in some cumbrous symbols.

0.5 In order to enable the reader to follow the worked examples and illustrative material, a few tables of functions commonly required are given at the end of this volume. These tables are in no way a substitute for the comprehensive sets which have been published and which are a necessary adjunct to most practical and a good deal of theoretical work. Frequent reference will be made to the following :—

Biometrika Tables for Statisticians, edited by E. S. Pearson and H. O. Hartley, vol. 1. Cambridge University Press.

Statistical Tables for use in Biological, Agricultural and Medical Research, by Sir Ronald Fisher and F. Yates. Oliver and Boyd, Edinburgh.

Many more specialized tables are now available. A reader who requires tabular help in a particular field would do well to enlist expert advice. References in the text to the scope of tables are given in what is now fairly standard form ; e.g. $x = 0(0{\cdot}1)10(1)50(2)100$ means that the function is tabulated for an argument x which proceeds by steps of 0·1 from 0 to 10, thence by steps of 1 from 10 to 50 and thence by steps of 2 from 50 to 100.

CHAPTER 1

FREQUENCY-DISTRIBUTIONS

1.1 The fundamental notion in statistical theory is that of the group or aggregate, a concept for which statisticians use a special word—" population." This term will be generally employed to denote any collection of objects under consideration, whether animate or inanimate ; for example, we shall consider populations of men, of plants, of mistakes in reading a scale, of barometric heights on different days, and even populations of ideas, such as that of the possible ways in which a hand of cards might be dealt. The notion common to all these things is that of aggregation.

The science of Statistics deals with the properties of populations. In considering a population of men we are not interested, statistically speaking, in whether some particular individual has brown eyes or is a forger, but rather in how many of the individuals have brown eyes or are forgers, and whether the possession of brown eyes goes with a propensity to forgery in the population. We are, so to speak, concerned with the properties of the population itself. Such a standpoint can occur in physics as well as in demographic sciences. For instance, in discussing the properties of a gas we are, as a rule, not so much interested in the behaviour of particular molecules, as in that of the aggregate of molecules which go to compose the gas. The statistician, like Nature, is mainly concerned with the species and is careless of the individual.

1.2 We may therefore begin an approach to a definition of our subject by the following : Statistics is the branch of scientific method which deals with the properties of populations. This, however, is rather too general. Statistics deals only with the *numerical* properties. A dictionary, for example, sets out a population of words, and among the properties of that population which are a suitable subject for scientific inquiry is that of word-derivation. It is not of statistical concern, however, to know that some words are derived from Latin, some from Anglo-Saxon and some from Hindustani. The subject would only assume a statistical aspect if we were to inquire *how many* words were derived from the different sources.

1.3 As a second approximation to our definition we may then try the following : Statistics is the branch of scientific method which deals with the data obtained by counting or measuring the properties of populations.

This again is a little too general. A set of logarithm tables is a population of numerals, but it is hardly a subject for statistical inquiry, for every numeral is determined according to mathematical laws. The statistician is rather concerned with populations which occur in Nature and are thus subject to the multitudinous influences at work in the world at large. His populations rarely, if ever, conform exactly to simple mathematical rules, and in fact it is in the departure from such rules that he

often finds topics of the greatest statistical interest. To allow for this factor we may then formulate our definition as follows :—

Statistics is the branch of scientific method which deals with the data obtained by counting or measuring the properties of populations of natural phenomena. In this definition " natural phenomena " includes all the happenings of the external world, whether human or not.

1.4 For the avoidance of misunderstandings in the interpretation of this definition it may be as well to point out that " Statistics," the name of the scientific method, is a collective noun and takes the singular. The same word " statistics " is also applied to the numerical material with which the method operates, and in such a case takes the plural. Later in this book we shall meet the singular form " statistic," which is defined as a function of the observations in a sample from some population. " Statistic " in this sense takes the plural " statistics."

Frequency-distributions

1.5 Consider a population of members each of which bears some numerical value of a variable, e.g. of men measured according to height or of flowers classified according to numbers of petals. This variable we shall call a variate. If it can assume only a number of isolated values it will be called discontinuous, and if it can assume any value of a continuous range, continuous. The population of members will then correspond to a population of variate-values, and it is the properties of this latter population which we have to consider.

If the population consists of only a few members we can without much difficulty consider the population of variate-values exhibited by them ; but if, as usually happens, the aggregate is large (or, in a sense defined later, infinite), the set of variate-values has to be reduced in some way before the mind can grasp their significance. This is done by classification of the individuals into ranges of the variate. So far as possible the ranges should be equal, so that the numbers falling into different ranges are comparable. The interval is called the class-interval (or simply the interval) and the number of members bearing a variate-value falling into a given class-interval is the class-frequency (or simply the frequency). The manner in which the class-frequencies are distributed over the class-intervals is called the frequency-distribution (or simply the distribution).

1.6 Tables 1.1 and 1.2 give two frequency-distributions of observed populations classified according to a single variate. Table 1.1 shows the distribution of 1475 local districts in England and Wales for 1953, classified according to birth-rate (number of births per 1000 of the population). The distribution is shown in this table in a way which would be quite impossible if each of the 1475 districts were shown separately. The greatest number of districts fall within the range 15·5–16·5 per thousand and the frequencies tail off on either side of this value. Table 1.2 shows the number of persons subject to sur-tax in the United Kingdom in 1950 classified according to the variate " income." The class-intervals here are unequal—a typical feature of official figures—and in the last column of the table is a reduction of the class-frequencies to comparability, namely, to frequency per £500 within the class-interval concerned.

Table 1.1—Showing local government areas in England and Wales, 1953, with specified birth-rates per 1000 of population

(Data from the Registrar-General's Statistical Review of England and Wales for 1953)

Birth-rate	Number of districts	Birth-rate	Number of districts
2·5–	1	15·5–	250
3·5–	0	16·5–	178
4·5–	0	17·5–	129
5·5–	1	18·5–	64
6·5–	0	19·5–	39
7·5–	3	20·5–	20
8·5–	7	21·5–	12
9·5–	19	22·5–	7
10·5–	39	23·5–	1
11·5–	96	24·5–	2
12·5–	151	25·5–	3
13·5–	231	26·5–	0
14·5–	221	27·5–	1
		TOTAL	1475

Note—2·5– means "2·5 and less than 3·5" etc.

Table 1.2—Number of persons in the United Kingdom liable to sur-tax in the year beginning 6 April, 1950, classified according to the magnitude of their annual income (*from Annual Abstract of Statistics*, 1950)

Annual income (£000)			Number of persons	Estimated frequency per £500 interval
2 and not exceeding		2·5	60,336	60,336
2·5	,, ,,	3	41,033	41,033
3	,, ,,	4	45,532	22,766
4	,, ,,	5	23,263	11,632
5	,, ,,	6	13,475	6,737
6	,, ,,	8	13,456	3,364
8	,, ,,	10	6,419	1,605
10	,, ,,	12	3,551	888
12	,, ,,	15	2,926	488
15	,, ,,	20	2,007	201
20	,, ,,	25	820	82
25	,, ,,	30	399	40
30	,, ,,	40	376	19
40	,, ,,	50	134	6
50	,, ,,	75	128	2
75	,, ,,	100	45	1
100 and over			38	?
TOTAL NUMBER OF PERSONS			213,938	—

Looking at this column we see that the maximum frequency per £500 in this case is at the beginning of the frequency-distribution.

1.7 The frequency-distribution may be represented graphically. Measuring the variate-value along the x-axis and frequency per class-interval along the y-axis, we

erect at the abscissa corresponding to the centre of each class-interval an ordinate equal to the frequency per unit interval in that interval. The ends of these ordinates

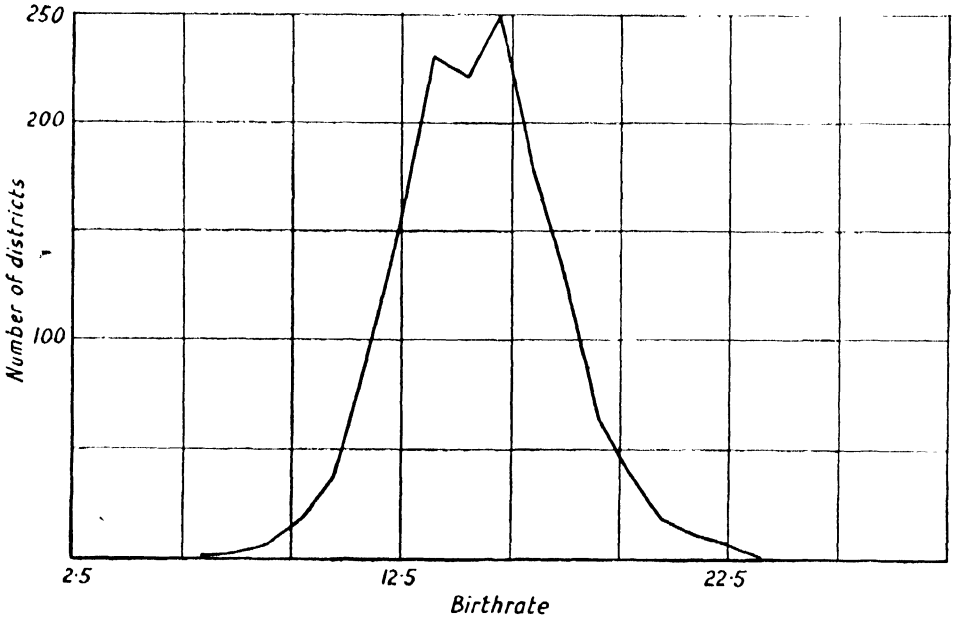


Fig. 1.1—Frequency polygon of the data of Table 1.1

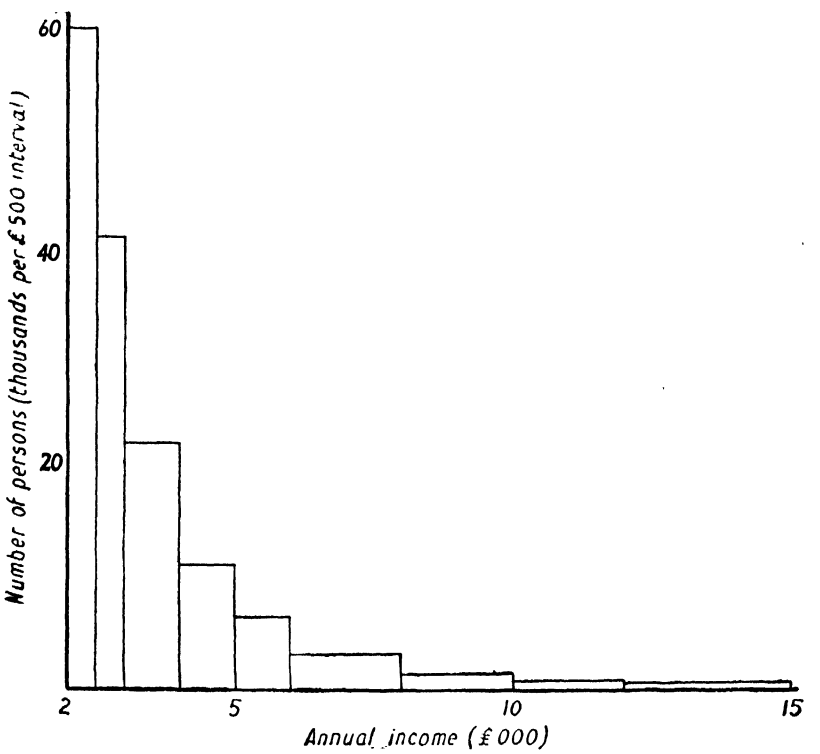


Fig. 1.2—Histogram of the data of Table 1.2

are joined by straight lines, one to the next. The diagram so obtained is called a Frequency Polygon. Fig. 1.1 shows the frequency polygon for the data of Table 1.1.

As a variant of this procedure we may erect on the abscissa range corresponding to each class-interval a rectangle whose area is proportional to the frequency in that interval. A diagram constructed in this way is called a Histogram. Fig. 1.2 shows such a histogram for the data of Table 1.2. It is evident that the histogram is a more suitable form of representation when the class-intervals are unequal.

1.8 A few practical points in the tabulation of observed frequency-distributions may be noted.

(1) It has been remarked that wherever possible the class-intervals should be equal. The importance of this will be more appreciated in subsequent chapters; but it is already evident that comparability is difficult to carry out by inspection when there exist inequalities in class-intervals. On running the eye down the second column of Table 1.2, for example, we note that the frequency in the interval 3–4 is greater than in the immediately preceding interval; but this is merely due to a change in the width of the intervals at that point and, as is seen from the third column, the frequency per unit interval decreases steadily.

(2) It is important to specify the class-interval with precision. We not infrequently meet with such classifications as "0–10, 10–20, 20–30," etc. To which interval is a member with variate-value 10 assigned? Obviously the classification is ambiguous if such values can in fact arise. We must either take the intervals "greater than or equal to 0 and less than 10, greater than or equal to 10 and less than 20," or make it clear what convention we use to allot a variate-value falling on the border between two neighbouring intervals, e.g. it might be decided to allot one-half of the member to each. There are various ways of indicating the class-interval in practical tables, e.g. "10–, 20–, 30–" means "greater than or equal to 10 and less than 20," and so forth. Sometimes, where a continuous variate is concerned, there is an element of imprecision in the specification of the fineness to which the measurements are made; for example, if we are measuring lengths in centimetres to the nearest centimetre, an interval shown as "greater than 15 and less than 18" means an interval of "greater than 15·5 and less than 17·5." When the precision of the measurements is known we can specify an interval by its middle point, for example, in this case, 16·5.

(3) Remark (1) about the importance of equality of class-intervals should not be held to preclude the specification of frequencies in finer intervals where the frequency is changing very rapidly. Table 1.3, for instance, shows the number of deaths from meningitis in England and Wales in 1953 according to the variate "age at death." If the frequencies in the interval "0 and less than 5" were not subdivided and were thus shown as a total 208 for the interval, we might be doubtful whether the greatest number of deaths occurred in the first year of life. This is so; the individual frequencies in the first five years bring out the relatively heavy mortality in the first year.

(4) Perhaps it is hardly necessary to add that the histogram is not a suitable method of representing data classified according to discontinuous variates. It shows the class-frequency uniformly dispersed over the whole interval, whereas if the variate is discontinuous, frequencies must necessarily be concentrated at certain points.

Table 1.3—Deaths from meningitis (except meningo-coccal and tuberculous) in England and Wales in 1953

(From the Registrar-General's Statistical Review for England and Wales for 1953)

Age, years	Number of deaths		Age, years	Number of deaths
0–	162		35–	5
1–	29		40–	11
2–	11	208	45–	17
3–	6		50–	21
4–	0		55–	32
5–	10		60–	23
10–	8		65–	22
15–	6		70–	15
20–	4		75–	11
25–	3		80–	10
30–	5		85 and over	3
			TOTAL	414

Frequency-distributions : discontinuous variates

1.9 It will be useful at this stage to give some examples of the frequency-distributions which occur in practice.

Table 1.4 shows the distribution of digits in numbers taken from a four-figure telephone directory. The numbers were chosen by opening the directory haphazardly

Table 1.4—Showing numbers of digits from the London Telephone Directory

(M. G. Kendall and B. Babington Smith, *Jour. Roy. Statist. Soc.*, **101**, 147)

Digit	0	1	2	3	4	5	6	7	8	9	TOTAL
Frequency	1026	1107	997	966	1075	933	1107	972	964	853	10,000

and taking the last two digits of all the numbers on the page except those in heavy type. The distribution is irregular, but from a cursory inspection of the table we are inclined to suppose that the digits occur approximately equally frequently in the larger population from which these 10,000 members were chosen. We shall see later that the divergences from the average frequency per digit, 1000, are not accidental sampling effects ; but at this stage it is sufficient to note that the data suggest for consideration a population of equally frequent members.

Table 1.5 shows a number of unit samples of foliage on orange trees infested by black-scale insect. The greatest frequency is at zero and the frequencies for higher values of the variate tail off more or less regularly.

Table 1.5—Number of black-scale insects on 821 unit samples (ten adjacent leaves and their stem) of orange trees

(W. M. Upholt and R. Craig, 1940, *Jour. Ecol. Entom.*, **33**, 113)

Number of scales	0	1	2	3	4	5	6	7	8	over 8	TOTAL
Frequency	199	124	106	65	42	46	36	30	14	159	821

In Table 1.6, on the other hand, showing suicides among women in some German states in certain years according to the variate " number of suicides per year," the distribution reaches its maximum frequency in the region 1–3 suicides and then tails off rather slowly.

Table 1.6—Showing suicides of women in eight German States in fourteen years
(Von Bortkiewicz, *Das Gesetz der kleinen Zahlen*, 1898)

Number of suicides per year	0	1	2	3	4	5	6	7	8	9	10 and over	TOTAL
Frequency . . .	9	19	17	20	15	11	8	2	3	5	3	112

Frequency-distributions : continuous variates

1.10 Table 1.7 shows a number of adult males in the United Kingdom (including, at the time of the collection of the data, the whole of Ireland), distributed according to the variate " height in inches." The frequency polygon is shown in Fig. 1.3. It

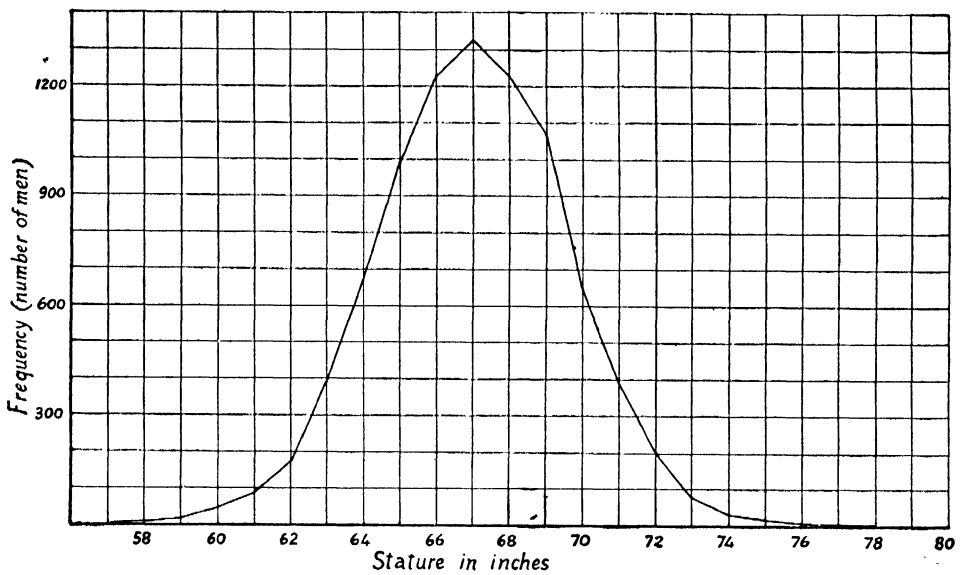


Fig. 1.3—Frequency-distribution of the data of Table 1.7. Values of the abscissa correspond to the beginning of class intervals

will be seen that the distribution is almost symmetrical, there being a maximum ordinate at 67– inches and a steady decrease in frequency on either side of the maximum.

This more-or-less uniform " tailing off " of frequencies is very common in observed distributions, but the symmetrical property is comparatively rare. Table 1.1 is roughly symmetrical, but Tables 1.8 and 1.9, showing respectively a number of Australian marriages distributed according to bridegroom's age, and a number of dairy herds distributed according to costs of production of milk, illustrate that various degrees of asymmetry can occur. An extreme form is shown in Table 1.3. In this connection Table 1.10, showing a number of men distributed according to weight, is of

interest for comparison with the height data of Table 1.7. The latter is symmetrical but the former is not.

Table 1.7—Showing the frequency-distribution of statures for adult males born in the United Kingdom (including the whole of Ireland)

(Final Report of the Anthropometric Committee to the British Association, 1883, p. 256)

As measurements are stated to have been taken to the nearest $\frac{1}{8}$th of an inch, the class-intervals are here presumably $56\frac{15}{16}$–$57\frac{15}{16}$, $57\frac{15}{16}$–$58\frac{15}{16}$, and so on

Height without shoes (inches)	Number of men within said limits of height	Height without shoes (inches)	Number of men within said limits of height
57–	2	68–	1230
58–	4	69–	1063
59–	14	70–	646
60–	41	71–	392
61–	83	72–	202
62–	169	73–	79
63–	394	74–	32
64–	669	75–	16
65–	990	76–	5
66–	1223	77–	2
67–	1329		
		TOTAL	8585

Table 1.8—Showing numbers of marriages contracted in Australia, 1907–14, arranged according to the age of bridegroom in 3-year groups

(From S. J. Pretorius (1930), *Biometrika*, **22**, 210)

Age of bridegroom (central value of 3-year range, in years)	Number of marriages	Age of bridegroom (central value of 3-year range, in years)	Number of marriages
16·5	294	55·5	1,655
19·5	10,995	58·5	1,100
22·5	61,001	61·5	810
25·5	73,054	64·5	649
28·5	56,501	67·5	487
31·5	33,478	70·5	326
34·5	20,569	73·5	211
37·5	14,281	76·5	119
40·5	9,320	79·5	73
43·5	6,236	82·5	27
46·5	4,770	85·5	14
49·5	3,620	88·5	5
52·5	2,190		
		TOTAL	301,785

Table 1.9—Number of dairy herds in a sample of herds from England and Wales, 1948/9 according to cost of production of milk

(From National Investigation into the Economics of Milk Production—Milk Marketing Board, 1950)

Cost in pence per gallon	Number of herds	Cost in pence per gallon	Number of herds
less than 12	3	24–	70
12–	19	26–	49
14–	52	28–	31
16–	96	30–	16
18–	121	32–	6
20–	115	34–	8
22–	86	36 and over	7
		TOTAL	679

Table 1.10—Frequency-distribution of weights for adult males born in the United Kingdom

(Source: as in Table 1.7. Weights were taken to the nearest pound, consequently the true class-intervals are 89·5–99·5, 99·5–109·5, etc.)

Weight in lbs	Frequency	Weight in lbs	Frequency
90–	2	190–	263
100–	34	200–	107
110–	152	210–	85
120–	390	220–	41
130–	867	230–	16
140–	1623	240–	11
150–	1559	250–	8
160–	1326	260–	1
170–	787	270–	—
180–	476	280–	1
		TOTAL	7749

1.11 When the asymmetry of a distribution such as that of Table 1.3 becomes extreme we may be unable to determine whether, near the maximum ordinate, there is a fall on either side, or whether the maximum occurs right at the start of the distribution. This would have been the case in Table 1.3 if we had not the finer grouping for the first five years of life; and it is the case in Table 1.2, in which the maximum frequency apparently occurs at or very close to an income of £2000 per annum. Asymmetrical distributions are sometimes called " skew "; and those such as Table 1.2 are called " J-shaped."

1.12 In rare cases the distribution may have maxima at both ends, as in Table 1.11

Table 1.11—Showing the frequencies of estimated intensities of cloudiness at Greenwich during the years 1890–1904 (excluding 1901) for the month of July

(Data from Gertrude E. Pearse (1928), *Biometrika*, 20A, 336)

Degrees of cloudiness	Frequency	Degrees of cloudiness	Frequency
10	676	4	45
9	148	3	68
8	90	2	74
7	65	1	129
6	55	0	320
5	45		
		TOTAL	1715

showing a number of days distributed according to degree of cloudiness. This is known as a U-shaped distribution.

1.13 Distributions also occur which in general appearance resemble sections of the types already mentioned. A J-shaped distribution, for example, resembles the " tail " of the symmetrical distribution of Table 1.7. The suicide data of Table 1.6 may be regarded as a symmetrical distribution truncated just below the maximum ordinate by the impossibility of the occurrence of negative values of the variate. This sort of conception is sometimes useful in fitting curves to observed data—a given analytical curve may fit the data quite well in a certain variate-range, but may also extend into regions where the data cannot, so to speak, follow it.

1.14 The distributions considered up to this point have one thing in common—they have only one maximum or, in the case of the U-shaped curve, only one minimum. Distributions also occur showing several maxima, Tables 1.12 and 1.13 being instances in point. The first, showing a number of deaths according to age at death, is typical of death distributions. Near the start of the distribution there is a maximum and a rapid fall in the frequency, which then rises to a pronounced maximum about the age 70–75, the frequencies beyond that point tailing off to zero. It is natural to wonder whether such a distribution can be usefully considered as two or more superposed distributions, e.g. a J-shaped distribution indicative of infantile mortality, and a skew distribution with a maximum at 70–75, the ordinary death curve of senescence.

A similar dissection of a complex distribution could be undertaken for the data of Table 1.13, showing a number of trypanosomes from the tsetse fly, *Glossina morsitans*, classified according to length. We are led to suspect here that the distribution is composed of the addition of several others (and this, by the way, has led to a suggestion that the trypanosomes are a mixture of distinct types).

Table 1.12—Showing the number of male deaths in England and Wales for 1953, classified by ages at death (*from reference to Table* 1.1)

Age at death	Number of deaths	Age at death	Number of deaths
0–	12,244	45–	9,016
5–	1,043	50–	14,507
10–	665	55–	19,204
15–	1,104	60–	26,802
20–	1,640	65–	34,292
25–	1,932	70–	39,644
30–	2,449	75–	40,162
35–	3,068	80–	29,061
40–	5,104	80 and over	17,553
		TOTAL	259,490

Table 1.13—Showing number of Trypanosomes from *Glossina morsitans* classified according to length in microns

(From K. Pearson (1914–15), *Biometrika*, **10**, 112. Length presumably to nearest micron)

Length (microns)	Frequency	Length (microns)	Frequency
15	7	26	110
16	31	27	127
17	148	28	133
18	230	29	113
19	326	30	96
20	252	31	54
21	237	32	44
22	184	33	11
23	143	34	7
24	115	35	2
25	130		
		TOTAL	2500

Frequency functions and distribution functions

1.15 The examples given above illustrate the remarkable fact that the majority of the frequency-distributions encountered in practice possess a high degree of regularity. The form of the frequency polygons and histograms above suggests, almost inevitably, that our data are approximations to distributions which can be specified by smooth curves and simple mathematical expressions. This approach to the concept of the frequency *function*, however, requires some care, particularly for continuous distributions.

Consider in the first place a discontinuous distribution such as that of Table 1.4. Let us represent our variate by ξ. Then we may say that ξ can take any of the ten values 0, 1, . . . 9 and that the frequency of a value of ξ equal to x, say $f(x)$, is given by the table, that is to say, $f(0) = 1026$, $f(1) = 1107$, $f(2) = 997$, and so on. The frequency table, in fact, defines the frequency function. Furthermore, most of the frequencies in the table are approximately 1000, and we may then consider the observed

distribution as approximating to that defined by

$$f(x) = 1000, \qquad x = 0, 1, \ldots 9, \tag{1.1}$$

or, more generally, to the distribution

$$f(x) = k, \qquad x = 0, 1, \ldots 9. \tag{1.2}$$

This is perhaps the simplest case of discontinuous frequency function, $f(x)$ being a constant for all permissible values of x.

In Table 1.5 we have a discontinuous variate which can, theoretically, take an infinite number of values, namely, any one of the positive integers. In practice, of course, there must be a limit to the number, but since we do not know that limit we may imagine our variate as infinite in range. The frequency function for the table itself is again simply defined by the frequencies therein; but if we wish to proceed to a conceptual generalization of such a table we must admit a discontinuous function $f(x)$ defined for all positive integral values of x. This occasions no difficulty provided that we are able to attach some meaning to the total frequency, i.e. that $\sum_{j=1}^{\infty} f(x_j)$ converges.

1.16 Consider now the case of a continuous variate. In the ordinary data of experience our distributions are invariably discontinuous, because our measurements can only attain a certain degree of accuracy. For instance, we are accustomed to suppose that the height of a man may in reality be any real number of inches in a certain range, say 50 to 80, such as 20π. In fact, we can measure heights only to a certain accuracy, say to the nearest thousandth of an inch. Our measurements thus consist of whole numbers (of thousandths) from 50,000 to 80,000, and such a number as 62,831·85 (= $20{,}000\pi$ approximately) cannot appear. All physical measurements are subject to this limitation, but we accept it and nevertheless speak of our variables as " continuous," the underlying supposition being that the measurements are approximations to numbers which can fall anywhere in the arithmetic continuum.

1.17 With this understanding we can consider the distribution of grouped frequencies as leading to the concept of a frequency function for a continuous variate. If, in one of the distributions above, say that of Table 1.7, we were to subdivide the intervals, we should probably find that *up to a point* the resulting frequencies were smoother and smoother. The reader can verify the appearance of this effect for himself by grouping the data of Table 1.7 in intervals of 8, 4, and 2 inches. We cannot, however, take the process too far, because, with a finite population, continued subdivision of the interval would sooner or later result in irregular frequencies, there being only a few members in each interval. But we may suppose that for ranges Δx, not too small, the distribution may be specified by a function $f(x)\Delta x$, expressing that in the range $\pm \frac{1}{2}\Delta x$ centred at x the frequency is $f(x)\Delta x$, *wherever x may be in the permissible range of the variate.* We may suppose further that as Δx tends to zero the population is perpetually replenished so as to prevent the occurrence of small and irregular frequencies; and in this way we arrive at the concept of the frequency function for a continuous variable. We write

$$dF = f(x)\,dx \tag{1.3}$$

expressing that the element of frequency dF between $x-\frac{1}{2}dx$ and $x+\frac{1}{2}dx$ is $f(x)\,dx$, for all x and for dx, however small.

1.18 This admittedly somewhat intuitive approach to the concept of the continuous frequency-distribution appears to be the best for statistical purposes, and is certainly the way in which the concept was originally reached. In formulating the axioms and postulates of a rigorous mathematical theory, however, the mathematician considers a rather more general function.

Let $F(x)$ be a function of x which is defined at every point in a continuous range and is continuous except perhaps at a denumerable number of points. We require that F shall be zero at the lower point of the range (which may be $-\infty$) and a constant N at the upper point (which may be $+\infty$) and that it shall not decrease at any point. Such a function is called a Distribution Function. It corresponds to the cumulated frequency of a frequency-distribution, N being the total frequency; for example, in Table 1.4, $F(x) = 0$ for $x<0$, $F(x) = 1026$ for $0\leqslant x<1$, $F(x) = 2133$ $(= 1026+1107)$ for $1\leqslant x<2$, and so on. Here there are ten points of discontinuity for $F(x)$. These points are called "saltuses" (jumps) and $F(x)$ in this case is called a Step Function.

If there is no saltus in the range, $F(x)$ is everywhere continuous and monotonically increasing. If it possesses a derivative we have the equation in differentials

$$\begin{aligned} dF &= F'(x)\,dx \\ &= f(x)\,dx \end{aligned} \tag{1.4}$$

corresponding to (1.3). $f(x)$ is called the Frequency Function. The mathematics of this branch of the subject is then that of the study of functions of the class $F(x)$ and $f(x)$.

1.19 The functions as thus defined are more general than those arrived at from the statistical approach in two ways: (i) $F(x)$ can increase monotonically in part of the range and then possess a saltus, i.e. the frequency may be continuous for a time and then suddenly discontinuous—in statistical practice a variate is either continuous or discontinuous throughout its range; (ii) where no saltus exists $F(x)$ can exist without there existing a frequency function, just as a continuous function need not necessarily possess a derivative. In all the cases we shall consider, the existence of a continuous variate will be accompanied by the existence of a frequency function.

The function $F(x)$ is sometimes called a Probability Function, for reasons which will become evident in Chapter 7 when we consider the theory of probability. Essentially, however, it has nothing to do with probability and we shall use the term "distribution function" only.

Fig. 1.4 shows the distribution function of the data of Table 1.2 in graphical form, $F(x)$ being given as ordinate against x as abscissa. In Fig. 1.5 we show the distribution function of the data of Table 1.7. In both cases we have "smoothed" the functions to some extent by joining the ends of the ordinates instead of erecting narrow rectangles. For a continuous variate such a process is clearly very often justified; it approximates to the continuous distribution which we imagine to underlie the data.

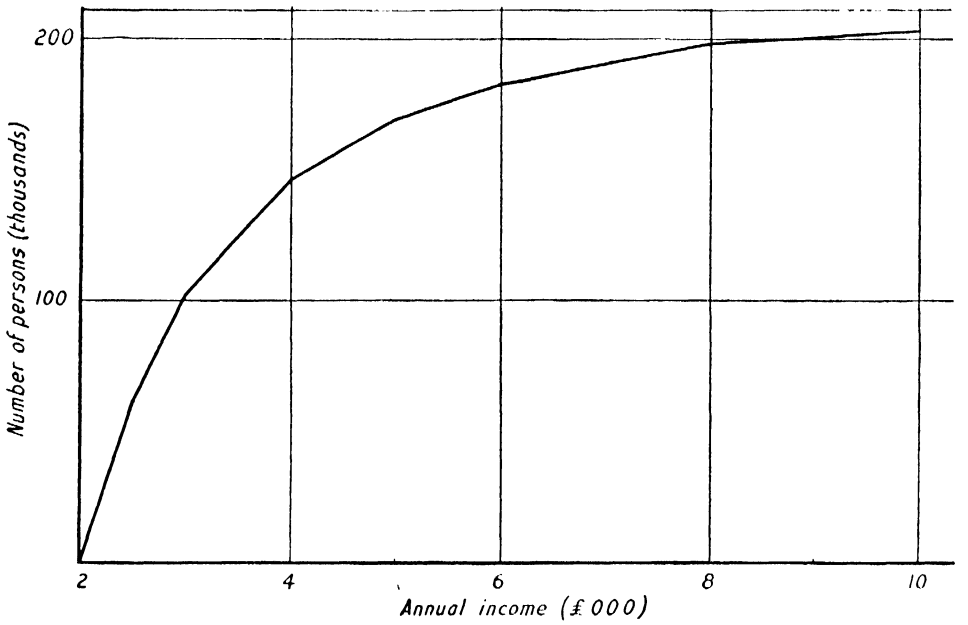


Fig. 1.4—Distribution function of the data of Table 1.2

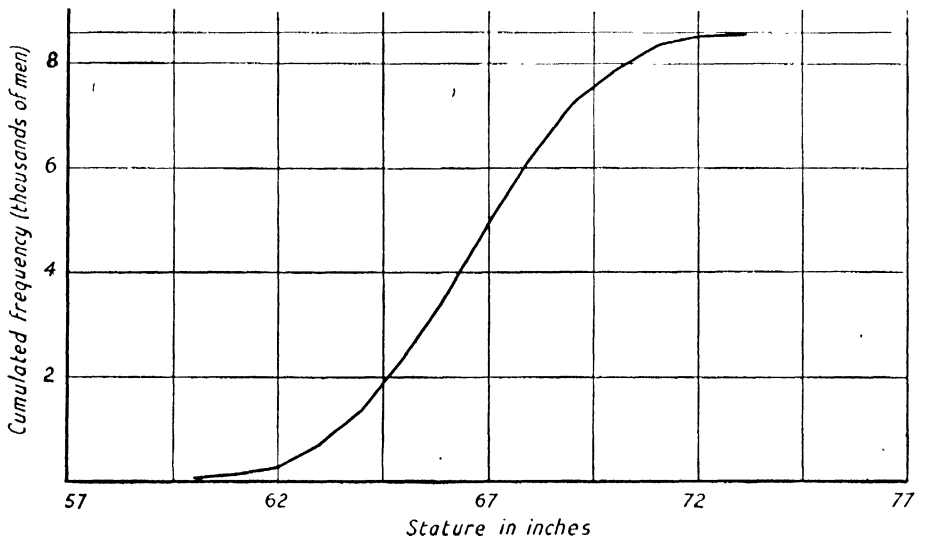


Fig. 1.5—Distribution function of the data of Table 1.7

1.20 If the discontinuous frequency function is $f(x)$, and $F(x)$ is taken to be the total frequency less than *or equal to* x, we have

$$F(x_r) = \sum_{j=1}^{r} f(x_s). \tag{1.5}$$

In the continuous case

$$F(x) = \int_a^x dF$$

$$= \int_a^x f(u)\,du \tag{1.6}$$

where the range is a to b. We now introduce two conventions which simplify these expressions to some extent. We shall suppose, unless the contrary is specified, that in these mathematical expressions our frequencies are always expressed as proportions of the total frequency, so that the total frequency is unity and the sum or integral over the whole range of the frequency function is also unity, i.e. $F(b) = 1$. Secondly, to avoid the constant specification of the limits a and b we may, without loss of generality, suppose that $F(x)$ and $f(x)$ are zero for any x less than a, and that $F(x) = 1$ and $f(x) = 0$ for any x greater than b. With this convention we may write

$$\left.\begin{aligned} F(x_r) &= \sum_{j=-\infty}^{r} f(x_j) \\ F(x) &= \int_{-\infty}^{x} f(u)\,du \end{aligned}\right\} \tag{1.7}$$

and

$$\left.\begin{aligned} \sum_{j=-\infty}^{\infty} f(x_j) &= F(\infty)-F(-\infty) = \\ \int_{-\infty}^{\infty} f(u)\,du &= F(\infty)-F(-\infty) = 1 \end{aligned}\right\} \tag{1.8}$$

Where it is necessary to take account of the total frequency N we may do so by multiplying (by N) frequencies given by the frequency function. In our convention $F(x)$ is always continuous on the right.

1.21 We shall often find it useful to abbreviate the expression "frequency function" to f.f., and similarly "distribution function" to d.f. (In Chapter 4 we shall encounter the "characteristic function," abbreviated to c.f.)

One other convention will occasionally help us to avoid cumbrous notation. Strictly speaking we should use different symbols, say ξ, to denote the variate and x to denote the range of values which it may take. Thus $F(x)$ denotes, as a function of x, the totality of values for which $\xi \leqslant x$. It is often convenient, however, to use the same symbol for both variate ξ and variable x. We may then speak of a *variate* x with f.f. $f(x)$ and d.f. $F(x)$. Wherever this is likely to lead to confusion we shall revert to the stricter notation.

Excursus on integrals

1.22 We have noted above that the functions which arise for consideration in statistical practice are less complicated than the more general type of distribution function which the mathematician wishes to discuss. This illustrates a point which causes constant difficulty in developing a systematic course on theoretical statistics. Modern mathematics has devised rigorous theories of great generality, but largely at the expense of added difficulty and abstractness. For most statistical purposes this elaborate apparatus is not required, and statistical functions can be adequately investigated by simpler methods without loss of rigour. In this book we shall, as a rule, prefer these simpler methods. But this course will not overcome all our difficulties.

Some results which we require are not obtainable by simple methods; for others, the so-called elementary proofs are so long and tedious as to be worse than the more sophisticated versions; and for others again the simpler treatment may obscure the essential statistical point.

1.23 The modern theory of integration is a case in point. For many purposes it would be enough to consider the integral as defined in introductory treatments, or in the manner of Cauchy. Thus if $f(x)$ is defined in the interval (a, b) we divide the range at points $x_1, x_2, \ldots x_n$ and consider the sum

$$s_n = f(a)(x_1-a)+f(x_1)(x_2-x_1)+ \ldots +f(x_n)(b-x_n). \tag{1.9}$$

It may be shown that under certain conditions (such as the continuity of $f(x)$ in the range) this sum tends to a limit as the length of the intervals tends to zero, independently of where the dividing-points are drawn or the way in which the tendency to zero proceeds. This limit is the Cauchy integral: $\int_a^b f(x)\,dx$.

A more general type of integration is due to Riemann and Stieltjes. Let $\psi(x)$ be a continuous function of x. Choose ξ_1 in the range a to x_1, ξ_2 in the range x_1 to x_2 and so on. Consider the sum

$$s'_n = \psi(\xi_1)\{F(x_1)-F(a)\}+\psi(\xi_2)\{F(x_2)-F(x_1)\}+ \ldots +\psi(\xi_n)\{F(b)-F(x_n)\} \tag{1.10}$$

It may be shown that as the length of the intervals tends to zero uniformly s tends to a limit which is independent of the location of the points ξ or the boundary points of the intervals. We then write this limit as

$$\int_a^b \psi(x)\,dF. \tag{1.11}$$

It is known as the Stieltjes integral of $\psi(x)$ with respect to $F(x)$. Riemann considered the case when $F(x) = x$.

1.24 The advantage of the Stieltjes integral is that it reduces to the Cauchy integral if $f(x)$ is continuous and to an ordinary summation if $f(x)$ is discontinuous. To save writing all our formulae down twice, once for the continuous f.f. and once for the discontinuous f.f., we shall usually employ an integral of this type, replacing it in special cases by an ordinary integral or a sum as the circumstances require.

Many of the theorems of ordinary integration are true of the Stieltjes integral. We shall frequently require the following:

$$\left|\int_a^b \psi\,dF\right| \leqslant \int_a^b |\psi|\,dF \tag{1.12}$$

$$\leqslant M\int_a^b dF$$

$$\leqslant M, \tag{1.13}$$

where M is the upper bound of $\psi(x)$ in the range (a, b).

$$\int_a^b \psi\,dF = \psi(\xi)\int_a^b dF, \tag{1.14}$$

where ξ is a value of x in the range (a, b).

If a and b are finite,

$$\int_a^b \sum_{j=1}^{\infty} f_j(x)\,dF = \sum_{j=1}^{\infty} \int_a^b f_j(x)\,dF, \tag{1.15}$$

provided that $\Sigma f_j(x)$ converges uniformly in the range. The theorem is not necessarily true if a or b is infinite.

The ordinary rules of partial integration are also applicable to Stieltjes integrals.

1.25 A more complicated integral is due to Lebesgue. It is equal to the Riemann-Stieltjes integral when the latter exists, but also exists in cases where the Riemann-Stieltjes integral does not. We do not normally require it in the ordinary theory of distributions, but it seems essential for a rigorous treatment of Fourier transforms which we need for certain parts of the theory of time-series.(*)

Variate-transformations

1.26 Suppose that we have a variable y related functionally to a variable x by the relation

$$x = x(y), \qquad y = y(x), \tag{1.16}$$

y being continuous and differentiable in x and vice versa. We then have

$$dx = \frac{dx}{dy}dy. \tag{1.17}$$

Consequently, for a continuous distribution with d.f. $F(x)$,

$$F(x) = \int_{-\infty}^{x} dF = \int_{-\infty}^{x} f(x)\,dx$$
$$= \int_{-\infty}^{x(y)} f(x)\frac{dx}{dy}\,dy.$$

We may thus write the distribution as

$$dF = f\{x(y)\}\frac{dx}{dy}\,dy, \tag{1.18}$$

expressing the fact that an element of frequency between $y-\frac{1}{2}dy$ and $y+\frac{1}{2}dy$ is $f\{x(y)\}\,dx/dy$. If ξ is the variate corresponding to x we may regard this expression as defining a variate η corresponding to y. The equation determining the frequency function is then transformable as if it were an equation in differentials. Such variate-transformations are important in the theory of continuous distributions. By their means many mathematically specified distributions may be reduced to a known form, either exactly or approximately.

(*) The reader who feels that his mathematical equipment needs strengthening or revision may care to refer to C. A. B. Smith's *Biomathematics*, which covers most of the mathematics required for statistical work at the intermediate level. More advanced theory is mostly covered by Sir Harold and Lady Jeffreys' *Methods of Mathematical Physics*.

For example, a distribution which we shall have to study in the theory of sampling is

$$dF = \frac{1}{2^{\frac{1}{2}\nu-1}\Gamma(\frac{1}{2}\nu)} e^{-\frac{1}{2}\chi^2} \chi^{\nu-1}\, d\chi, \qquad 0 \leqslant \chi \leqslant \infty. \tag{1.19}$$

It is readily verified that $F(\infty) = 1$.

By the transformation $t = \frac{1}{2}\chi^2$ we reduce this to

$$dF = \frac{1}{\Gamma(\frac{1}{2}\nu)} e^{-t} t^{\frac{1}{2}\nu-1}\, dt, \qquad 0 \leqslant t \leqslant \infty. \tag{1.20}$$

This is a well-known form in analysis, the distribution function being the incomplete Γ-function :

$$F(t) = \Gamma_t(\tfrac{1}{2}\nu)/\Gamma(\tfrac{1}{2}\nu). \tag{1.21}$$

Again, the distribution

$$dF = \frac{y_0}{(1+t^2/\nu)^{\frac{1}{2}(\nu+1)}}\, dt, \qquad -\infty \leqslant t \leqslant \infty \tag{1.22}$$

(y_0 being chosen so that $F(\infty) = 1$), a symmetrical peaked distribution of infinite range rather like that of Fig. 1.3, may, by the substitution of $t = \sqrt{\nu}\tan\theta$, be transformed into

$$\begin{aligned} dF &= \frac{y_0\sqrt{\nu}\sec^2\theta\, d\theta}{\sec^{\nu+1}\theta}, \qquad -\tfrac{1}{2}\pi \leqslant \theta \leqslant \tfrac{1}{2}\pi, \\ &= y_0\sqrt{\nu}\cos^{\nu-1}\theta\, d\theta, \end{aligned} \tag{1.23}$$

a distribution of finite range $-\frac{1}{2}\pi$ to $+\frac{1}{2}\pi$, but still symmetrical. Putting now $\sin\theta = u$, we have

$$dF = y_0\sqrt{\nu}(1-u^2)^{\frac{1}{2}\nu-1}\, du, \qquad -1 \leqslant u \leqslant 1, \tag{1.24}$$

and again $u^2 = x$,

$$dF = y_0\sqrt{\nu}\, x^{-\frac{1}{2}}(1-x)^{\frac{1}{2}\nu-1}\, dx, \qquad 0 \leqslant x \leqslant 1. \tag{1.25}$$

The effect on the range of this last substitution is to be noted. u ranges from -1 to $+1$, and as it does so x ranges from $+1$ to 0 and back to $+1$. The distribution function of the x-distribution, $F(x)$ from 0 to x, is thus that of the u-distribution from $-u^2$ to u^2. Whenever substitutions are made under which there is not a (1, 1) continuous relation between the variates, points such as this require some watching.

1.27 There is one variate-transformation which is worth special attention. In the distribution

$$dF = f(x)\, dx$$

put

$$y = \int_{-\infty}^{x} f(u)\, du.$$

Then

$$\begin{aligned} dF &= f(x)\frac{dx}{dy}\, dy \\ &= \frac{f(x)}{f(x)}\, dy \\ &= dy, \qquad 0 \leqslant y \leqslant 1, \end{aligned} \tag{1.26}$$

so that the distribution is transformed into the very simple " rectangular " form in which all values of the variate from 0 to 1 are equally frequent. Any continuous distribution can be transformed into the rectangular form ; and it follows that there exists at least one transformation which will transform any continuous frequency-distribution into any other continuous frequency-distribution, viz. the transformation which transforms one into the rectangular form coupled with the reverse of that which transforms the other into the rectangular form.

The genesis of frequency-distributions

1.28 Up to this point we have not inquired into the origin of the various observed frequency-distributions which have been adduced in illustration. Certain of them may be considered apart from any question of origination from a larger population. The death distribution of Table 1.12 is an example ; if we are interested only in the distribution of male deaths in England and Wales in 1953 the whole of the population under consideration is before us.

But in the great majority of cases the population which we are able to examine is only part of a larger population on which our main interest is centred. The height distribution of Table 1.7 is only a part of the population of men in the United Kingdom living at the time of the inquiry, and it is mainly of importance in the light of the information which it gives us about that population. Similarly the distribution of dairy herds of Table 1.9 is mainly of interest in the information it gives about costs of milk production for the whole country.

1.29 In the two cases just mentioned, height and costs of milk production, we have information about a certain sample of individuals chosen from an existing population. Only lack of time and opportunity prevents us from examining the whole population. It sometimes happens, however, that we have data which do not emanate from a finite existent population in this way. Table 1.14 is an example. It shows the distribution of throws with dice.

Table 1.14—Showing the number of successes (throws of 4, 5 or 6) with throws of 12 dice

(Weldon's data, cited by F. Y. Edgeworth, *Encyclopædia Britannica*, 11th ed., **22**, 39)

Number of successes	Frequency	Number of successes	Frequency
0	0	7	847
1	7	8	536
2	60	9	257
3	198	10	71
4	430	11	11
5	731	12	0
6	948		
		TOTAL	4096

Now it is clear that, in a sense, we have not in these data got a complete population, for we can add to them by further casting of the dice. But these further throws do not exist in the sense that the unexamined men of the United Kingdom or the

unexamined dairy herds of England and Wales exist. They have a kind of hypothetical existence conferred on them by our notion of the throwing of the dice.

Even distributions which appear at first sight to be existent may be considered in this light. The trypanosome distribution of Table 1.13, for instance, was obtained from certain tsetse flies. We may consider it as a sample of all the tsetse flies in existence, whether harbouring trypanosomes or not—an existent population; but we may also consider it as a sample of what the distribution would be if all the tsetse flies were infected with trypanosomes—a hypothetical population.

The population conceived of as parental to an observed distribution is fundamental to statistical inference. We shall take up this matter again in later chapters when we consider the sampling problem. The point is mentioned here because it will occasionally arise before we reach those chapters. It must be emphasized that the distinction between existent and hypothetical universes is not merely a matter of ontological speculation—if it were we could safely ignore it—but one of practical importance when inferences are drawn about a population from a sample generated from it.

Multivariate distributions

1.30 In the foregoing sections we have considered the members of a population according to a single variate, and the frequency-distributions may thus be called univariate. The work may be readily generalized to include populations of members considered according to two or more variates, yielding bivariate, trivariate . . . multivariate frequency-distributions. Table 1.15, for example, shows the distribution of a number of beans according to both length and breadth. The border frequencies show the univariate distributions of the beans according to length and breadth separately, and the body of the table shows how the two qualities vary together.

Table 1.15—Showing frequencies of beans with specified lengths and breadths

(Johannsen's data, cited by S. J. Pretorius (1930), *Biometrika*, **22**, 110)

Breadth in millimetres (central values) \ Lengths in millimetres (central values)	17	16·5	16	15·5	15	14·5	14	13·5	13	12·5	12	11·5	11	10·5	10	9·5	Totals
9·125	—	2	—	—	3	—	—	—	—	—	—	—	—	—	—	—	5
8·875	4	8	17	19	—	—	—	—	—	—	—	—	—	—	—	—	48
8·625	2	23	101	156	93	23	2	—	—	—	—	—	—	—	—	—	400
8·375	—	18	105	494	574	227	56	9	—	—	—	—	—	—	—	—	1483
8·125	—	4	44	375	956	913	362	73	12	3	—	—	—	—	—	—	2742
7·875	—	—	7	81	385	871	794	330	89	19	3	—	—	—	—	—	2579
7·625	—	—	1	4	65	236	469	361	175	55	27	4	—	—	—	—	1397
7·375	—	—	—	—	6	23	91	137	124	78	37	22	11	—	1	—	530
7·125	—	—	—	—	—	1	13	18	28	35	25	32	11	6	1	—	170
6·875	—	—	—	—	—	—	—	1	9	8	21	12	13	7	1	—	72
6·625	—	—	—	—	—	—	—	—	—	—	2	—	1	4	3	—	10
6·375	—	—	—	—	—	—	—	—	—	1	—	—	—	1	1	1	4
Totals	6	55	275	1129	2082	2294	1787	929	437	199	115	70	36	18	7	1	9440

As for the univariate case, the variates may be discontinuous or continuous and we sometimes meet cases in which one variate is of one kind and one of the other.

1.31 In generalization of the frequency polygon and the histogram we may construct three-dimensional figures to represent the bivariate distribution. Imagine a horizontal plane containing a pair of perpendicular axes and ruled like a chessboard into cells, the ruled lines being drawn at points corresponding to the terminal points of class-intervals. At the centre of each interval we erect a vertical line proportional in length to the frequency in that interval. The summits of these verticals are joined, each to the four summits of verticals in the neighbouring cells possessing the same values of one or the other variate. The resulting figure is the bivariate frequency polygon or stereogram.

Similarly we may erect on each cell a pillar proportional in volume to the frequency in that cell and thus obtain a bivariate histogram. Fig. 1.6 shows such a figure for the bean data of Table 1.15.

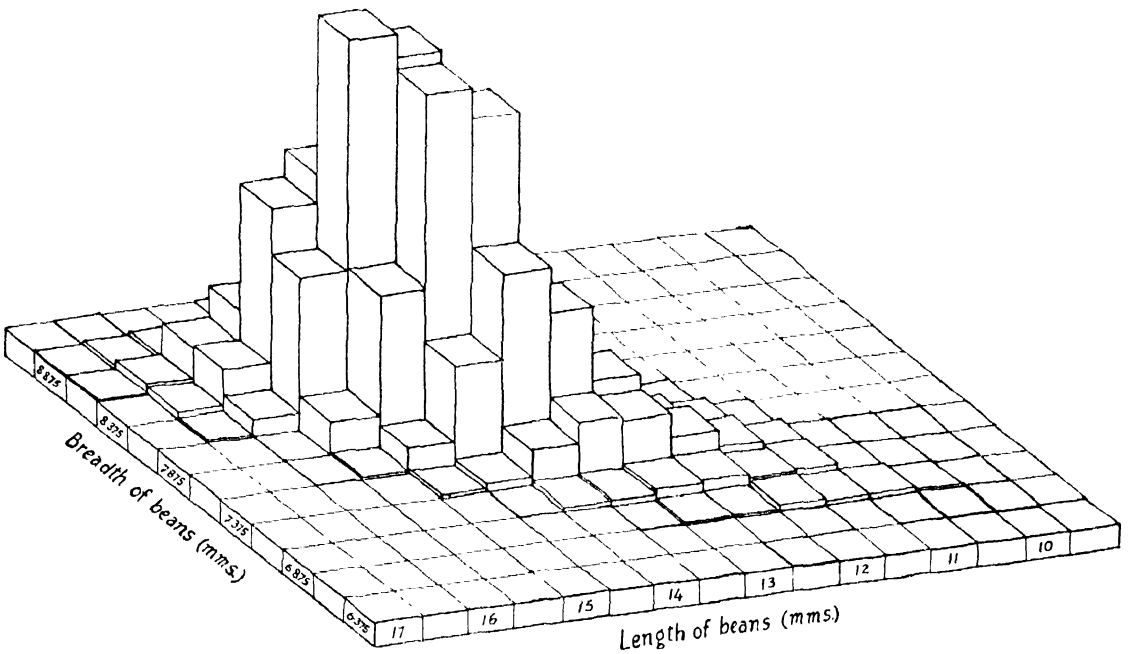


Fig. 1.6—Bivariate histogram of the data of Table 1.15

1.32 We may write the bivariate distribution with variates x_1, x_2, as

$$dF = f(x_1, x_2)\,dx_1\,dx_2. \tag{1.27}$$

With the usual conventions we shall then have for the bivariate distribution function

$$F(x_1, x_2) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2} f(u_1, u_2)\,du_1\,du_2 \tag{1.28}$$

this integral also being understood in the Stieltjes sense, reducing to ordinary integration if $f(x_1, x_2)$ is continuous and to ordinary summation if it is discontinuous.

For any fixed value of one variate, say $x_2 = k$, the other variate will, of course, have a univariate distribution. If we wish to observe the convention that the total

frequency should be unity this will be given by

$$dF_0(x_1) = \frac{f(x_1, k)\,dx_1}{\int_{-\infty}^{\infty} f(x_1, k)\,dx_1}. \tag{1.29}$$

(1.29) is called the conditional distribution of x_1 for $x_2 = k$.

Marginal distributions

1.33 The sums in the margins of tables such as Table 1.15 give us the frequency-distributions of each variate by itself. For any value of x_2 the total frequency is obtained by summing over all values of x_1 to give the frequency for that value of x_2. Thus, if the d.f. is $F(x_1, x_2)$ we may express the univariate d.f. of x_1 as $F(x_1, \infty)$; and similarly the univariate d.f. of x_2 is $F(\infty, x_2)$. In terms of frequency functions

$$dF_1(x_1) = \left\{\int_{-\infty}^{\infty} f(x_1, x_2)\,dx_2\right\} dx_1, \tag{1.30}$$

$$dF_2(x_2) = \left\{\int_{-\infty}^{\infty} f(x_1, x_2)\,dx_1\right\} dx_2. \tag{1.31}$$

These univariate distributions are sometimes called " marginal," for obvious reasons. In economic contexts, where marginality is likely to have a different connotation, the term is, perhaps, better avoided ; but it is convenient and rarely gives rise to difficulty in practice.

Independence

1.34 If

$$\begin{aligned} F(x_1, x_2) &= F(x_1, \infty)F(\infty, x_2) \\ &= F_1(x_1)F_2(x_2), \quad \text{say,} \end{aligned} \tag{1.32}$$

x_1 and x_2 are said to be independent. Where f.f.'s exist we have

$$\int_{-\infty}^{x_1}\int_{-\infty}^{x_2} f(x_1, x_2)\,dx_1\,dx_2 = \int_{-\infty}^{x_1} f_1(x_1)\,dx_1 \int_{-\infty}^{x_2} f_2(x_2)\,dx_2$$

and hence

$$f(x_1, x_2) = f_1(x_1)f_2(x_2). \tag{1.33}$$

Thus the d.f. and the f.f. factorize into two parts, one for each variate alone. It is readily seen that this definition of statistical independence conforms to the colloquial use of the word and also to its mathematical use. The conditional distribution of x_2 for any fixed x_1 (e.g. the distribution in a row or column of the bivariate frequency table) is the same whatever the fixed value of x_1, that is to say, the distribution of x_2 is independent of x_1.

Two variates which are not independent are said to be dependent. Evidently those of Table 1.15 are dependent, for the distributions in rows or in columns are far from similar.

Example 1.1

Consider the distribution

$$dF = z_0 \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{x_1^2}{\sigma_1^2}-\frac{2\rho x_1 x_2}{\sigma_1\sigma_2}+\frac{x_2^2}{\sigma_2^2}\right)\right\} dx_1\,dx_2, \qquad -\infty \leqslant x_1, x_2 \leqslant \infty. \quad (1.34)$$

Here z_0 is to be chosen so that the total frequency is unity. Put

$$y_1 = \frac{1}{(1-\rho^2)^{\frac{1}{2}}}\left(\frac{x_1}{\sigma_1}-\frac{\rho x_2}{\sigma_2}\right)$$

$$y_2 = \frac{x_2}{\sigma_2}.$$

We have

$$\frac{\partial(y_1,y_2)}{\partial(x_1,x_2)} = \begin{vmatrix} \dfrac{1}{\sigma_1(1-\rho^2)^{\frac{1}{2}}} & -\dfrac{\rho}{\sigma_2(1-\rho^2)^{\frac{1}{2}}} \\ 0 & \dfrac{1}{\sigma_2} \end{vmatrix}$$

$$= \frac{1}{\sigma_1\sigma_2(1-\rho^2)^{\frac{1}{2}}}.$$

Also

$$y_1^2+y_2^2 = \frac{1}{1-\rho^2}\left\{\frac{x_1^2}{\sigma_1^2}-\frac{2\rho x_1 x_2}{\sigma_1\sigma_2}+\frac{x_2^2}{\sigma_2^2}\right\}.$$

The distribution then becomes

$$dF = z_0\sigma_1\sigma_2(1-\rho^2)^{\frac{1}{2}} \exp\{-\tfrac{1}{2}(y_1^2+y_2^2)\}\,dy_1\,dy_2$$

$$= z_0\sigma_1\sigma_2(1-\rho^2)^{\frac{1}{2}} e^{-\frac{1}{2}y_1^2}\,dy_1\, e^{-\frac{1}{2}y_2^2}\,dy_2. \quad (1.35)$$

The original variables are clearly dependent, for (1.34) cannot be factorized into a product of a function of x_1 times a function of x_2. From (1.35) we see that the transformed variables are independent.

Incidentally, this gives us a method of evaluating z_0. For

$$\int_{-\infty}^{\infty} e^{-\frac{1}{2}u^2}\,du = \sqrt{(2\pi)}$$

and on integrating the right-hand side of (1.35) with respect to y_1 and y_2 and equating the result to unity we find

$$z_0 = 1/\{2\pi\sigma_1\sigma_2(1-\rho^2)^{\frac{1}{2}}\}.$$

1.35 In this example we have used a Jacobian to transform the integral from one pair of variables to another. The process is evidently quite general. If

$$dF = f(x_1,x_2)\,dx_1\,dx_2$$

and

$$x_1 = x_1(y_1,y_2), \qquad x_2 = x_2(y_1,y_2),$$

the frequency function for y_1 and y_2 will be given by

$$f\{x_1(y_1,y_2),\ x_2(y_1,y_2)\}J, \quad (1.36)$$

where

$$J = \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_2}{\partial y_1} \\ \frac{\partial x_1}{\partial y_2} & \frac{\partial x_2}{\partial y_2} \end{vmatrix}$$

and is to be taken with positive sign in (1.36). If J changes sign in the domain of integration the variate-transformation may develop ambiguities which require special attention. We shall discuss this point when it arises.

1.36 An extension of the bivariate results to n-variate results is immediate. For instance, if a d.f. $F(x_1, x_2, \ldots x_n)$ can be expressed in the form

$$F(x_1, x_2, .., x_n) = F(x_1, \infty, .., \infty)F(\infty, x_2, .., \infty)\ldots F(\infty, \infty, .., x_n) \qquad (1.37)$$

the corresponding variates are independent. Only one point needs special comment: for n variates to be independent it is not enough that each should be independent of each of the others. That is to say, if

$$F(x_1, x_2, \infty, .., \infty) = F(x_1, \infty, \infty, .., \infty)F(\infty, x_2, \infty, .., \infty) \qquad (1.38)$$

and similarly for every pair of x's, equation (1.37) does not necessarily follow. When speaking of independence among sets of variates greater than two in number it is therefore of some importance to specify exactly what is meant. Variates obeying (1.37) we shall describe as *completely* independent. A state of affairs represented by (1.38) might be described as one of *pairwise* independence.

Frequency-generating functions

1.37 For distributions which are mathematically specified it is often convenient, especially for discontinuous functions, to summarize the specification in a generating function. Suppose that the frequency of a variate x, which can take values 0, 1, 2,... is f_r for $x = r$. This frequency is then the coefficient of t^r in

$$P(t) \equiv \sum_{j=0}^{\infty} f_r t^r \qquad (1.39)$$

which is thus a frequency-generating function. For example, the so-called binomial distribution, which we shall study in some detail at a later stage, is such that the frequency at $x = r$ $(r = 0, 1, .., n)$ is given by

$$f_r = \binom{n}{r} p^r (1-p)^{n-r}, \qquad 0 \leqslant p \leqslant 1. \qquad (1.40)$$

Hence a generating function of type (1.39) is given by

$$\begin{aligned} P(t) &= \sum_{j=0}^{n} \binom{n}{j} p^j (1-p)^{n-r} t^j \\ &= \{pt + (1-p)\}^n. \end{aligned} \qquad (1.41)$$

EXERCISES

1.1 Draw frequency polygons or histograms of the following distributions :—

Table 1.16—Frequency-distribution of successes in twelve dice thrown 4096 times, a throw of 6 points reckoned as a success (*Weldon's data ; source as in Table* 1.14)

Number of successes . . .	0	1	2	3	4	5	6	7 and over	TOTAL
Number of throws . . .	447	1145	1181	796	380	115	24	8	4096

Table 1.17—Showing numbers of sentences of given lengths in passages from Macaulay's essays on Bacon and on Chatham

(From G. Udny Yule (1939), *Biometrika*, **30**, 363)

Length of sentence in words	Number of sentences	Length of sentence in words	Number of sentences
1–5	46	66–	2
6–	204	71–	4
11–	252	76–	8
16–	200	81–	2
21–	186	86–	2
26–	108	91–	1
31–	61	96–	2
36–	68	101–	1
41–	38	106–	—
46–	24	111–	1
51–	20	116–	—
56–	12	121–	1
61–	8		
		TOTAL	1251

Table 1.18—Showing the numbers of old Egyptian skulls with specified lengths of the left occipital bone in millimetres

(From T. L. Woo (1930), *Biometrika*, **22**, 324)

Length (central values)	Frequency	Length (central values)	Frequency
84·5	12	102·5	74
86·5	12	104·5	68
88·5	32	106·5	36
90·5	48	108·5	18
92·5	79	110·5	7
94·5	116	112·5	4
96·5	104	114·5	4
98·5	126	116·5	—
100·5	123	118·5	1
		TOTAL	864

Table 1.19—Showing the frequency-distribution of fecundity, i.e. the ratio of the number of yearling foals produced to the number of coverings, for brood-mares (racehorses) covered eight times at least

(Pearson, Lee and Moore (1899), *Phil. Trans.*, A, **192**, 303. Where a case fell on the border between two intervals, one-half was assigned to each)

Fecundity	Number of mares with fecundity between the given limits	Fecundity	Number of mares with fecundity between the given limits
1/30– 3/30	2·0	17/30–19/30	315·0
3/30– 5/30	7·5	19/30–21/30	337·0
5/30– 7/30	11·5	21/30–23/30	293·5
7/30– 9/30	21·5	23/30–25/30	204·0
9/30–11/30	55·0	25/30–27/30	127·0
11/30–13/30	104·5	27/30–29/30	49·0
13/30–15/30	182·0	29/30–1	19·0
15/30–17/30	271·5	TOTAL	2000·0

Table 1.20—Frequency-distribution of plots according to yield of grain in pounds from plots of $\frac{1}{600}$th acre in a wheat-field

(Mercer and Hall (1911), *Jour. Agr. Science*, **4**, 107)

Yield of grain in pounds per $\frac{1}{600}$th acre (central value of range)	2·8	3·0	3·2	3·4	3·6	3·8	4·0	4·2	4·4	4·6	4·8	5·0	5·2	TOTAL
Number of plots . . .	4	15	20	47	63	78	88	69	59	35	10	8	4	500

Table 1.21—Frequency-distribution of 166 London bus-drivers showing the number who met with 0, 1, 2, . . . accidents in an exposure of (*a*) one year, (*b*) five years (*Farmer, E. and Chambers, E. G.* (1939))

Industrial Health Research Board Report 84. London, H.M. Stationery Office)

Number of accidents	Number of drivers meeting with said number of accidents	
	In one year	In five years
0	45	1
1	36	2
2	40	3
3	19	14
4	12	17
5	8	21
6	3	17
7	2	14
8	1	14
9	—	12
10	—	13
11	—	9
12	—	6
13	—	2
14	—	6
15 or more	—	15
TOTAL	166	166

Table 1.22—Showing the number of women aborting at specified term in weeks

(From T. V. Pearce (1930), *Biometrika*, **22**, 250)

Term (weeks)	Frequency	Term (weeks)	Frequency
4	3	17	13
5	7	18	14
6	10	19	8
7	13	20	4
8	14	21	2
9	29	22	10
10	22	23	4
11	21	24	4
12	18	25	3
13	28	26	4
14	16	27	6
15	19	28	1
16	10		
		TOTAL	283

1.2 Sketch the following curves and compare their shapes with those of the distributions in the previous exercise :—

$$y = y_0 e^{-\frac{1}{2}x^2}, \qquad -\infty \leqslant x \leqslant \infty.$$

$$y = y_0 e^{-x}, \qquad 1 \leqslant x \leqslant \infty.$$

$$y = y_0 x^{-\gamma}, \qquad \gamma > 1, 0 \leqslant x \leqslant \infty.$$

$$y = \frac{y_0}{(1+x^2)^n}, \qquad n > 0, -\infty \leqslant x \leqslant \infty.$$

$$y = y_0 (1-x)^a x^b, \qquad a, b > 1, 0 \leqslant x \leqslant 1.$$

$$y = y_0 e^{-\alpha x} x^\gamma, \qquad \gamma > 1, 0 \leqslant x \leqslant \infty.$$

$$y = y_0 (1-x^2)^a, \qquad a < 0, -1 \leqslant x \leqslant 1.$$

1.3 Show that the following distributions can all be transformed into the type

$$dF = y_0 (1-x)^{p-1} x^{q-1}\, dx, \qquad 0 \leqslant x \leqslant 1,$$

and find the transformations :

$$dF = r_0 (1-r^2)^{\frac{1}{2}(n-4)}\, dr, \qquad -1 \leqslant r \leqslant 1.$$

$$dF = \frac{t_0}{(1+t^2/\nu)^{\frac{1}{2}(\nu+1)}}\, dt, \qquad -\infty \leqslant t \leqslant \infty.$$

$$dF = \frac{z_0 e^{\nu_1 z}}{(\nu_1 e^{2z} + \nu_2)^{\frac{1}{2}(\nu_1+\nu_2)}}\, dz, \qquad -\infty \leqslant z \leqslant \infty.$$

(All these distributions are important in statistical theory. The distribution to which they are reduced is called the Type I or B-distribution.)

1.4 Sketch the stereograms or bivariate histograms of the following distributions :—

Table 1.23—Sample of students in the University of London, 1955, classified by (1) the number of newspapers read and (2) the number of newspapers looked at only, on a particular day. (*H. S. Booker, unpublished data*)

(1) Number looked at only

(2) Number read	0	1	2	3	4	5 or more	Totals
0	77	75	19	10	1	2	184
1	179	136	65	20	15	2	417
2	86	70	45	18	1	3	223
3	17	21	13	3	4	—	58
4	4	2	2	2	—	—	10
5 or more	2	2	—	2	—	—	6
Totals	365	306	144	55	21	7	898

Table 1.24—Number of cows distributed according to (1) age in years and (2) yield of milk per week in 4912 Ayrshire cows

(Data from J. F. Tocher (1928), *Biometrika*, **20**B, 106)

(1) Age in years

(2) Yield of milk per week (gallons) (central value of intervals)	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Totals
8	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	1
9	—	2	2	—	1	—	—	—	—	—	—	—	—	—	—	—	5
10	3	5	1	1	3	—	—	—	—	—	—	—	—	—	—	—	13
11	2	10	8	7	1	—	1	—	2	1	—	1	—	—	—	—	33
12	2	25	17	9	5	4	4	2	1	1	—	—	—	1	—	—	71
13	9	76	29	18	9	2	4	1	1	1	—	1	—	—	—	—	151
14	11	76	57	38	23	9	7	6	4	2	3	—	—	—	—	—	236
15	11	115	79	43	34	24	11	8	4	5	1	2	1	—	1	—	339
16	15	149	119	74	59	23	23	16	9	7	4	—	—	—	1	—	499
17	16	148	131	94	58	34	32	15	12	6	5	—	1	—	—	—	552
18	11	146	132	83	73	49	39	22	17	6	5	1	1	—	—	—	585
19	10	117	112	113	87	51	35	33	11	10	2	3	1	—	—	1	586
20	8	97	107	79	69	51	25	30	13	10	3	3	—	—	1	—	496
21	3	63	93	88	70	49	31	29	9	7	4	—	1	—	1	—	448
22	5	42	63	49	45	32	14	18	10	3	1	2	—	—	—	—	284
23	1	19	33	38	38	27	17	17	12	7	1	2	2	—	—	—	214
24	2	20	23	34	27	19	13	9	3	2	1	—	—	—	—	—	153
25	3	10	15	22	17	20	8	10	3	4	—	—	—	—	—	—	112
26	—	7	13	7	4	15	2	4	2	3	1	—	—	—	—	—	58
27	—	2	7	9	5	5	4	2	—	—	—	—	—	1	—	—	35
28	—	—	2	1	4	2	1	1	2	—	—	—	—	—	—	—	13
29	—	—	2	2	4	1	3	—	3	—	—	—	—	—	—	—	15
30	—	—	—	—	—	2	—	—	2	—	—	—	—	—	—	—	4
31	—	—	2	1	—	—	2	—	—	—	—	—	—	—	—	—	5
32	—	—	—	2	—	—	—	—	—	—	—	—	—	—	—	—	2
33	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	1
34	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	1
Totals	112	1129	1047	812	636	419	276	223	122	75	32	15	7	2	4	1	4912

1.5 Show that the conditions that the function

$$f(x_1, x_2) = z_0 \exp\{Ax_1^2 + 2Hx_1x_2 + Bx_2^2\}, \qquad -\infty \leqslant x_1, x_2 \leqslant \infty,$$

may represent a frequency function are

(*a*) $A \leqslant 0$,
(*b*) $B \leqslant 0$,
(*c*) $AB - H^2 \geqslant 0$.

Show further that if these conditions are satisfied and the integral of $f(x_1, x_2)$ between $-\infty$ and ∞ for both variates is unity, then

$$z_0 = \frac{1}{\pi}\begin{vmatrix} -A & H \\ H & -B \end{vmatrix}^{\frac{1}{2}}.$$

1.6 For the bivariate distribution

$$dF = \frac{ky}{(1+x)^4}\exp\left(-\frac{y}{1+x}\right)dx\,dy, \qquad 0 \leqslant x, y \leqslant \infty,$$

show by a transformation $u = y/(1+x)$, $v = 1/(1+x)$ that the distribution of u and v is

$$dF = k\,u\,e^{-u}\,du\,dv, \qquad 0 \leqslant u \leqslant \infty, 0 \leqslant v \leqslant 1,$$

and hence that $k = 1$.

1.7 For the bivariate distribution

$$dF = \frac{k(1+x+y)}{(1+x)^4(1+y)^4}\,dx\,dy, \qquad 0 \leqslant x, y < \infty,$$

show that $k = 9/2$. Show also that the marginal distribution of x is

$$dF = \frac{3}{4}\frac{(2x+3)}{(1+x)^4}\,dx, \qquad 0 \leqslant x \leqslant \infty.$$

1.8 A distribution is given by

$$dF = k\,\mathrm{sech}^n x\,dx, \qquad -\infty \leqslant x \leqslant \infty.$$

Sketch the form of the frequency curve and show that

$$k = \frac{\Gamma\{\frac{1}{2}(n+1)\}}{\Gamma(\frac{1}{2}n)\sqrt{\pi}}.$$

1.9 Three variates are independent and each is distributed in the form

$$dF = \frac{1}{\sqrt{(2\pi)}}\,e^{-\frac{1}{2}x^2}\,dx, \qquad -\infty \leqslant x \leqslant \infty.$$

Transforming to new variates by the equations

$$\begin{aligned} y_1 &= (x_1 - x_2)/\sqrt{2} \\ y_2 &= (x_1 + x_2 - 2x_3)/\sqrt{6} \\ y_3 &= (x_1 + x_2 + x_3)/\sqrt{3}, \end{aligned}$$

show that the y's are also completely independent and are distributed in the same form.

1.10 In the previous exercise, make a transformation of the polar type

$$\begin{aligned} x_1 &= r\cos\theta_1\cos\theta_2 \\ x_2 &= r\cos\theta_1\sin\theta_2 \\ x_3 &= r\sin\theta_1. \end{aligned}$$

Hence show that the variates corresponding to r, θ_1 and θ_2 are completely independent and that the distribution of the first is given by

$$dF = \sqrt{\left(\frac{2}{\pi}\right)}e^{-\frac{1}{2}r^2}r^2\,dr, \qquad 0 \leqslant r \leqslant \infty.$$

1.11 In the distribution

$$dF = \frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{\frac{1}{2}}}\exp\left\{\frac{-1}{2(1-\rho^2)}\left(\frac{x^2}{\sigma_1^2} - \frac{2\rho\,xy}{\sigma_1\sigma_2} + \frac{y^2}{\sigma_2^2}\right)\right\}dx\,dy, \qquad -\infty \leqslant x, y \leqslant \infty,$$

show that the distribution of x is

$$dF = \frac{1}{\sigma_1\sqrt{(2\pi)}}\exp\left(-\frac{x^2}{2\sigma_1^2}\right)dx, \qquad -\infty \leqslant x \leqslant \infty.$$

Sketch the distribution of x and y and consider the limiting case as $\rho \longrightarrow 1$.

1.12 In the distribution

$$dF = \frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{\frac{1}{2}}}\frac{n-1}{n-2}\left[1+\frac{1}{2(n-2)(1-\rho^2)}\left(\frac{x^2}{\sigma_1^2}-\frac{2\rho xy}{\sigma_1\sigma_2}+\frac{y^2}{\sigma_2^2}\right)\right]^{-n}dx\,dy \qquad -\infty \leqslant x, y \leqslant \infty,$$

show that the distribution of x is given by

$$dF = \frac{1}{\sigma_1\sqrt{(2\pi)}}\frac{\sqrt{(n-2)}}{\Gamma(n-1)}\Gamma(n-\tfrac{1}{2})\left(1+\frac{x^2}{2(n-2)\sigma_1^2}\right)^{-n+\frac{1}{2}}dx, \qquad -\infty \leqslant x \leqslant \infty.$$

(K. Pearson, 1923, *Biometrika*, **15**, 231)

1.13 The so-called Poisson distribution is such that the frequency at $x = r$ ($r = 0, 1, 2, \ldots$) is $e^{-\lambda}\lambda^r/r!$ Show that a frequency-generating function is

$$P(t) = \exp\{\lambda(t-1)\}$$

where the frequency at $x = r$ is the coefficient of t^r in $P(t)$.

1.14 A discontinuous distribution (the so-called negative binomial) has frequency a $x = r$ ($r = 0, 1, 2, \ldots$) given by

$$f_r = \left(1+\frac{m}{k}\right)^{-k}\frac{\Gamma(k+r)}{r!\,\Gamma(k)}\left(\quad\ldots\quad\right)^r$$

where k is an integer. Show that a frequency-generating function is given by

$$P(t) = \left\{1+\frac{m}{k}(1-t)\right\}^{-k}.$$

1.15 A variate is distributed over the interval 0 to ∞ in the form $dF = f(x)\,dx$. Write down the distribution of the variate in the interval $x_0 \leqslant x \leqslant \infty$ ($x_0 > 0$). If this is of the same form as the original variate show that $f(x) = ke^{-kx}$. It may be assumed that $f(x)$ is differentiable.

1.16 A n-variate distribution has the form

$$dF = k\exp\left\{-\sum_{i,j=1}^{n} a_{ij}x_ix_j\right\}dx_1\ldots dx_n, \qquad -\infty \leqslant x_1, \ldots, x_n \leqslant \infty,$$

with $a_{ij} = a_{ji}$. By considering x_n show that the exponent can be put in the form

$$-a_{nn}\left(x_n+\frac{1}{a_{nn}}\sum_{i=1}^{n-1}a_{ni}x_i\right)^2-\sum_{i,j=1}^{n-1}b_{ij}x_ix_j$$

where $b_{ij} = a_{ij}-\dfrac{a_{in}a_{jn}}{a_{nn}}$, $\quad i,j = 1,2,\ldots,n-1.$

Hence show how to transform the original variates by linear transformation to a completely independent set each of which is distributed in a form of the type

$$dF = \frac{1}{\sigma\sqrt{(2\pi)}}e^{-u^2/2\sigma^2}du, \qquad -\infty \leqslant u \leqslant \infty.$$

1.17 For a bivariate distribution with a differentiable frequency function $f(x, y)$, show that there exists a linear transformation to new variates which are independent if and only if

$$\left(A\frac{\partial^2}{\partial x^2}+2H\frac{\partial^2}{\partial x\,\partial y}+B\frac{\partial^2}{\partial y^2}\right)\log f = 0$$

where A, H and B are constants. Hence show that for the distribution of Exercise 1.11 there exists an infinite number of linear transformations to independent variates.

1.18 Three variates represented by x, y and z have the distribution

$$dF = k x^{l-1} y^{m-1} z^{n-1} dx\, dy\, dz.$$

They do not assume negative values but are such that $x+y+z \leqslant 1$. Show that

$$k = \frac{\Gamma(l+m+n+1)}{\Gamma(l)\,\Gamma(m)\,\Gamma(n)}.$$

1.19 n variates have the distribution represented by

$$dF = k x_1^{l_1-1} x_2^{l_2-1} \ldots x_n^{l_n-1} g(x_1+x_2+\ldots+x_n)\, dx_1 \ldots dx_n.$$

They do not assume negative values but are such that $\sum_{i=1}^{n} x_i \leqslant 1$. Show that

$$1/k = \frac{\Gamma(l_1)\Gamma(l_2)\ldots\Gamma(l_n)}{\Gamma(l_1+l_2+\ldots+l_n)} \int_0^1 v^{L-1} g(v)\, dv$$

where

$$L = \sum_{i=1}^{n} l_i.$$

1.20 Show that if $g(x)$ can be expanded in a uniformly convergent series of powers in x

$$\int_{-\infty}^{\infty} \frac{\sin^2 x}{x^2} g(\sin^2 x)\, dx = \int_0^{\pi} g(\sin^2 x)\, dx.$$

Sketch the distribution

$$dF = \frac{k \sin^2 x \cos^2 x}{x^2(1+\sin^2 x)} dx, \qquad -\infty \leqslant x \leqslant \infty,$$

and show that $1/k = \pi(\sqrt{2}-1)$.

1.21 Two independent variates x_1, x_2 each have the distribution

$$dF = \frac{1}{\sqrt{(2\pi)}} \exp(-\tfrac{1}{2}x^2)\, dx, \qquad -\infty \leqslant x \leqslant \infty.$$

Show that the variates

$$y_1 = \exp\{-\tfrac{1}{2}(x_1^2+x_2^2)\},$$

$$y_2 = \frac{1}{2\pi} \arctan(x_1/x_2)$$

are independently rectangularly distributed on the range (0, 1).

(Box and Muller, 1958, *Ann. Math. Statist.*, **29**, 610)

1.22 If $F_1(x_1)$, $F_2(x_2)$ are univariate d.f.'s, show that

$$F(x_1, x_2) = F_1(x_1) F_2(x_2)[1+\theta\{1-F_1(x_1)\}\{1-F_2(x_2)\}]$$

(where θ is a constant, $-1 \leqslant \theta \leqslant 1$) is a bivariate d.f.

(cf. E. J. Gumbel, 1960, *J. Amer. Statist. Ass.*, **55**, 698)

CHAPTER 2

MEASURES OF LOCATION AND DISPERSION

2.1 It has been seen in Chapter 1 that the frequency-distributions occurring in statistical practice vary considerably in general nature. Some are finite in range and some are not. Some are symmetrical and some markedly skew. Some present only a single maximum and others present several. Amid this variety we may, however, discern four general types: (*a*) the symmetrical distribution with a single maximum, such as that of Table 1.7; (*b*) the asymmetrical distribution, or skew distribution, with a single maximum, such as those of Tables 1.8 and 1.9; (*c*) the extremely skew, or J-shaped, distribution, such as that of Table 1.2; and (*d*) the U-shaped distribution, such as that of Table 1.11. To make this classification comprehensive we should have to add a fifth class comprising the miscellaneous distributions not falling into the other four.

The distributions with a single maximum will hereafter be called "unimodal."

2.2 It frequently happens in statistical work that we have to compare two distributions. If one is unimodal and the other J-shaped or multimodal a concise comparison is clearly difficult to make, and in such a case it would probably be necessary to specify both distributions completely. But if both are of the same type (and it is in such cases that comparisons most frequently arise) we may be able to make a satisfactory comparison merely by examining their principal characteristics; e.g. if both are unimodal it might be sufficient to compare (*a*) the whereabouts of some central value, such as the maximum—this, as it were, locates the distributions; (*b*) the degree of scatter about this value—the dispersion; and (*c*) the extent to which the distributions deviate from the symmetrical—the skewness.

The same point emerges when our distributions are specified by some mathematical function. If, for example, we have two distributions of the type

$$dF = k\exp\left\{-\frac{(x-m)^2}{2\sigma^2}\right\}dx, \qquad -\infty \leqslant x \leqslant \infty,$$

symmetrical about $x = m$, a complete comparison can be made by comparing the value of the constants m and σ in the distributions. Such constants are called *parameters* of the distribution.(*)

Measures of location: the arithmetic mean

2.3 There are three groups of measures of location in common use: the means

(*) There has, in the past, been some confusion as to what is meant by a parameter of a distribution. We shall use the term only to denote quantities, such as m and σ above, which appear explicitly in the specification of the distribution. Thus the mean and variance are not parameters in our sense; although it so happens that in the frequency function given above m is equal to the mean and σ^2 to the variance.

(arithmetic, geometric and harmonic), the median and the mode. We consider them in turn.

The arithmetic mean is perhaps the most generally used statistical measure, and in fact is far older than the science of statistics itself. If the proportional frequency of the values x of a distribution is $f(x)$, the arithmetic mean μ_1' about the point $x = a$ is defined by

$$\begin{aligned}\mu_1' &= \int_{-\infty}^{\infty} (x-a) f(x)\, dx \\ &= \int_{-\infty}^{\infty} (x-a)\, dF. \end{aligned} \tag{2.1}$$

This integral is to be understood in the Stieltjes sense and hence includes summation in the discontinuous case; e.g. the arithmetic mean of a set of discrete values x is their sum divided by the number of values. In formula (2.1) the frequency, in accordance with our usual convention, is expressed as a proportion of the total frequency. If the *actual* frequencies are $g(x)$, totalling N, we have

$$\mu_1' = \frac{1}{N}\int_{-\infty}^{\infty} (x-a) g(x)\, dx$$

in the continuous case, and

$$\mu_1' = \frac{1}{N}\sum_{j=-\infty}^{\infty} (x_j - a) g(x_j)$$

in the discontinuous case. The value of the arithmetic mean thus depends on the value of a, the point from which it is measured. For a mathematically specified distribution the integral (2.1) need not necessarily converge, in which case no arithmetic mean exists.

2.4 The calculation of the arithmetic mean of a numerically specified distribution (i.e. one whose frequency-distribution is given in the form of a numerical table like those of Chapter 1) is a simple process. If there are relatively few values in the population we merely sum them and divide by their total number N. If they are given in the form of a frequency table a more formal procedure is desirable, but the principle is exactly the same. The following example will make the process clear.

Example 2.1

To calculate the arithmetic mean of the population of males distributed according to height of Table 1.7.

Let us note first of all that if b is some other arbitrary variate-value,

$$\begin{aligned}\mu_1' \text{ (about } a) &= \int_{-\infty}^{\infty} (x-a)\, dF \\ &= \int_{-\infty}^{\infty} (x-b)\, dF + \int_{-\infty}^{\infty} (b-a)\, dF \\ &= \mu_1' \text{ (about } b) + b - a. \end{aligned} \tag{2.2}$$

In other words, we can find the mean about any point very simply when we know

the mean about any other. In calculating the arithmetic mean we can then take an arbitrary point as origin and transfer to any other desired point afterwards. It is generally convenient to choose this arbitrary point somewhere near the maximum frequency.

One further point arises in grouped data such as these. We do not know *exactly* the variate-values of the individuals within a certain class range. We therefore assume them concentrated at the centre of the interval. Corrections for any distortion thus introduced will be considered in Chapter 3. In fact, no correction is required for the arithmetic mean in the case when the frequency "tails off" at both ends of the distribution.

In the particular case before us we take an arbitrary origin at the centre of the interval 67– inches, i.e. at the point $67\frac{7}{16}$ inches, and measure $\xi(=x-a)$ from that point. Column 2 in Table 2.1 shows the frequency, column 3 the value of ξ and column 4 the value of ξf. We find, having due regard to sign,

$$\Sigma(\xi f) = 8763 - 8584 = 179.$$

Hence the mean about $x = 0$ is $67\frac{7}{16} + \frac{179}{8585} = 67{\cdot}46$ inches.

Table 2.1—Calculation of the arithmetic mean for the distribution of Table 1.7

(1) Height, inches	(2) Frequency f	(3) Deviation from arbitrary value ξ	(4) Product ξf
57–	2	− 10	20
58–	4	− 9	36
59–	14	− 8	112
60–	41	− 7	287
61–	83	− 6	498
62–	169	− 5	845
63–	394	− 4	1576
64–	669	− 3	2007
65–	990	− 2	1980
66–	1223	− 1	1223
67–	1329	0	− 8584
68–	1230	+ 1	1230
69–	1063	+ 2	2126
70–	646	+ 3	1938
71–	392	+ 4	1568
72–	202	+ 5	1010
73–	79	+ 6	474
74–	32	+ 7	224
75–	16	+ 8	128
76–	5	+ 9	45
77–	2	+ 10	20
TOTALS	8585	—	+ 8763

Example 2.2

For a distribution specified by a mathematical function, the determination of the mean is a matter of evaluating the integral (2.1), when it exists. For instance, to find the mean of the distribution

$$dF = \frac{1}{B(p,q)}(1-x)^{p-1}x^{q-1}\,dx, \qquad 0 \leqslant x \leqslant 1\ ;\ p,q>0,$$

we have

$$\begin{aligned}\mu_1' &= \frac{1}{B(p,q)}\int_0^1 (1-x)^{p-1}x^q\,dx \\ &= \frac{B(p,q+1)}{B(p,q)} = \frac{\Gamma(p)\Gamma(q+1)}{\Gamma(p+q+1)}\cdot\frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \\ &= \frac{q}{p+q}.\end{aligned}$$

Example 2.3

If a distribution is specified by a generating function the mean can be evaluated in the following manner.

Let

$$P(t) = \sum_{j=0}^{\infty} f_j t^j. \tag{2.3}$$

Then

$$\begin{aligned}\left[\frac{dP}{dt}\right]_{t=1} &= \left[\Sigma f_j t^{j-1} j\right]_{t=1} \\ &= \Sigma f_j j \\ &= \mu_1' \text{ (about zero).}\end{aligned} \tag{2.4}$$

For example, with the binomial distribution specified by

$$P(t) = \{pt+(1-p)\}^n$$

we have

$$\begin{aligned}\left[\frac{dP}{dt}\right]_{t=1} &= np\,\{p+(1-p)\}^n \\ &= np,\end{aligned}$$

which is therefore the mean about zero.

The geometric mean and the harmonic mean

2.5 Two other types of mean are in use in elementary statistics, though they are of less importance in advanced theory.

The geometric mean of N variate-values is the Nth root of their product and is not used if any of the variate-values is zero or negative. For proportional frequencies $f(x)$ we have

$$\left.\begin{aligned} G &= \prod_{j=-\infty}^{\infty} (x_j^{f_j}) \\ \text{or} \qquad \log G &= \sum_{j=-\infty}^{\infty} f_j \log x_j \end{aligned}\right\} \tag{2.5}$$

and for actual frequencies $g(x)$, totalling N,

$$\left.\begin{aligned} G &= \Pi (x_j^{g_j})^{\frac{1}{N}} \\ \log G &= \frac{1}{N}\Sigma g_j \log x_j \end{aligned}\right\} \tag{2.6}$$

The harmonic mean of N variate-values is the reciprocal of the arithmetic mean of their reciprocals. In the usual notation

$$\frac{1}{H} = \int_{-\infty}^{\infty} \frac{dF}{x} = \int_{-\infty}^{\infty} \frac{f(x)\,dx}{x} \tag{2.7}$$

or, for actual frequencies,

$$\frac{1}{H} = \frac{1}{N}\int_{-\infty}^{\infty} \frac{g(x)\,dx}{x} \tag{2.8}$$

provided, of course, that the integral exists.

Example 2.4

To find the geometric and harmonic means of the distribution

$$dF = \frac{1}{B(p,q)}(1-x)^{p-1}x^{q-1}\,dx, \qquad 0 \leqslant x \leqslant 1\,;\ \ p,q>0,$$

we have

$$\log G = \frac{1}{B(p,q)}\int_0^1 (1-x)^{p-1}x^{q-1}\log x\,dx.$$

Now, since by definition

$$\int_0^1 (1-x)^{p-1}x^{q-1}\,dx = B(p,q)$$

we have, differentiating both sides with respect to q, an operation which is legitimate in virtue of the uniform convergence of the integral and the existence of the resulting expressions,

$$\int_0^1 (1-x)^{p-1}x^{q-1}\log x\,dx = \frac{\partial}{\partial q}B(p,q).$$

Thus

$$\begin{aligned} \log G &= \frac{1}{B(p,q)}\frac{\partial}{\partial q}B(p,q) \\ &= \frac{\partial}{\partial q}\log\frac{\Gamma(p)\,\Gamma(q)}{\Gamma(p+q)} \\ &= \frac{\partial}{\partial q}\{\log\Gamma(q)-\log\Gamma(p+q)\}. \end{aligned}$$

The harmonic mean is given by

$$\frac{1}{H} = \frac{1}{B(p,q)}\int_0^1 (1-x)^{p-1}x^{q-2}\,dx.$$

To keep this expression finite, however, we must require $q>1$. Then

$$\frac{1}{H} = \frac{B(p,q-1)}{B(p,q)} = \frac{\Gamma(q-1)}{\Gamma(p+q-1)}\cdot\frac{\Gamma(p+q)}{\Gamma(q)}$$

$$= \frac{p+q-1}{q-1},$$

so that
$$H = \frac{q-1}{p+q-1}.$$

We may note that the arithmetic mean, $\frac{q}{p+q}$, is greater than the harmonic mean, for

$$\frac{q}{p+q} = 1-\frac{p}{p+q}, \quad \frac{q-1}{p+q-1} = 1-\frac{p}{p+q-1}$$

and therefore
$$\mu_1' > H$$

if
$$\frac{p}{p+q-1} > \frac{p}{p+q},$$

which is clearly so.

2.6 In general it may be shown that for distributions in which the variate-values are not negative

$$H \leqslant G \leqslant \mu_1'. \tag{2.9}$$

Consider in fact the quantity

$$A(t) = \left|\frac{1}{N}(x_1^t + x_2^t + \ldots x_N^t)\right|^{\frac{1}{t}}$$

where the x's are real numbers. We shall show that this is an increasing function of t, i.e. $A(t_1) > A(t_2)$ if $t_1 > t_2$. As a trivial case these inequalities may be replaced by equalities, namely if all the x's are equal.

Note that if we put $t = 1$ in $A(t)$ we have the arithmetic mean; when $t = -1$ we have the harmonic mean; and as t tends to zero we have the geometric mean, for

$$\begin{aligned}\lim_{t\to 0} \log A &= \lim (\log \Sigma x^t/N)/t \\ &= \lim (\Sigma x^t \log x)/N \\ &= \Sigma \log x/N.\end{aligned}$$

Put

$$\left.\begin{aligned} I &= \frac{1}{N}\Sigma x^t \\ F &= t^2 \frac{d}{dt}\log A. \end{aligned}\right\}$$

Then

$$F = t^2 \frac{d}{dt}\left(\frac{\log I}{t}\right) = \frac{t}{I}\frac{dI}{dt} - \log I,$$

and hence
$$\frac{dF}{dt} = \frac{t}{I^2}\left\{I\frac{d^2I}{dt^2} - \left(\frac{dI}{dt}\right)^2\right\}. \tag{2.10}$$

The expression in braces in (2.10) is non-negative, in virtue of the Cauchy-Schwarz

inequality, for it is equal to

$$\left(\frac{1}{N}\Sigma x^t\right)\left(\frac{1}{N}\Sigma x^t \log^2 x\right)-\left(\frac{1}{N}\Sigma x^t \log x\right)^2.$$

Hence dF/dt has the sign of t and F thus has a minimum at $t = 0$. But when $t = 0$, $F = 0$ and hence F is non-negative. Therefore $d\log A/dt$ is non-negative, and since A is non-negative dA/dt is non-negative and thus A is a non-decreasing function. It will, in fact, be increasing unless all the x's are equal.

The inequality (2.9) follows.

For simplicity we have stated these results for the discontinuous variate. The analysis, however, is easily seen to remain true for Stieltjes integrals and hence is generally valid.

Hereafter when the " mean " is mentioned without qualification, the arithmetic mean is to be understood.

2.7 There are many statistical uses of the Cauchy-Schwarz inequalities and it may be useful to interpolate a few comments on them. The inequality given by Cauchy in 1821 states that if a_i, b_i $(i = 1, 2, \ldots n)$ are real numbers $\left(\sum_{i=1}^{n} a_i^2\right)\left(\sum_{i=1}^{n} b_i^2\right) > \left(\sum_{i=1}^{n} a_i b_i\right)^2$ unless the a's and b's are proportional, in which case the inequality becomes an equality. The corresponding result for integrals was given by Buniakowsky in 1859 and Schwarz in 1885. Many other inequalities can be derived from it, e.g. Hölder's of 1889 : if $a_i, b_i, \ldots, l_i$ $(i = 1, \ldots n)$ are non-negative, $\alpha, \beta, \ldots, \lambda$ are positive with $\alpha+\beta+\ldots+\lambda = 1$, then

$$\Sigma a^\alpha b^\beta \ldots l^\nu < \Sigma a^\alpha \Sigma b^\beta \ldots \Sigma l^\nu$$

unless one set of the (a) etc. are all zero or unless one set is proportional to all the others, in which cases the inequality becomes an equality. A very clear and comprehensive account is given in the book *Inequalities* by Hardy, Littlewood and Pólya, 1934, Cambridge University Press.

The median

2.8 The median value is that value of the variate which divides the total frequency into two equal halves, i.e. is the value μ_e such that

$$\int_{-\infty}^{\mu_e} f(x)\,dx = \int_{\mu_e}^{\infty} f(x)\,dx = \tfrac{1}{2}. \tag{2.11}$$

There is some small indeterminacy in this definition when the distribution is discontinuous which may be removed by convention. If there are $(2N+1)$ members of the population, we take the median to be the value of the $(N+1)$th member. If there are $2N$ we take it to be halfway between the values of the Nth and the $(N+1)$th. When the distribution is numerically specified in class-intervals there is the usual indeterminacy due to grouping, which may be dealt with in the manner of the following example.

Example 2.5

To find the median value of the distribution of heights considered in Example 2.1.

Half the total frequency of 8585 observations is 4292·5.
There are, up to and including the interval beginning at $66\frac{15}{16}$ inches, 3589
leaving 703·5.
The frequency in the next interval is 1329.
Hence we take the median to be

$$66\tfrac{15}{16}+\frac{703\cdot5}{1329} = 67\cdot47 \text{ inches.}$$

The mean (Example 2.1) is 67·46 inches, practically the same.

A graphical method of determining the median is given later in this chapter (**2.15**).

2.9 From the mathematical viewpoint the indeterminacy in the median can be removed. For a set of values $x_1, \ldots, x_N$ it may be shown that the sum $\Sigma|\xi - x_i|^p$, considered as a function of ξ, has a unique minimum for some ξ_p if $p > 1$; and further, that as p tends to unity ξ_p tends to some unique value, which may be defined as the median. See Jackson (1921).

The extension of the concept of median to the multivariate case is not entirely straightforward. The median of a bivariate distribution is usually taken to be a pair of values, each being the median of one univariate marginal distribution. See Haldane (1948).

The mode

2.10 The mode or modal value is that value of the variate exhibited by the greatest number of members of the distribution. If the frequency function is continuous and differentiable it is the solution of

$$f'(x) = \frac{d}{dx}f(x) = 0, \qquad f''(x) = \frac{d^2}{dx^2}f(x) < 0. \tag{2.12}$$

If $f'(x)$ vanishes and $f''(x)$ is greater than zero we have a minimum, and such a point is sometimes called an "antimode."

In numerically specified distributions and discontinuous distributions generally the mode is sometimes difficult to determine exactly. It is essentially a concept related to the continuous frequency function. For example, if the distribution merely consists of an isolated number of values, each of which occurs only once, there is no mode in the sense defined above. Where, however, the number is large enough to permit grouping, there will usually be an interval containing a maximum frequency, and we may regard the mode as lying in that interval. More generally there may be several maxima, in which case the distribution is multimodal. In the height distribution of Table 1.7, for instance, the mode may be considered as lying somewhere in the interval 67– inches. To estimate its position more accurately it is necessary to fit a continuous curve to the distribution and determine the mode of the curve. The process of fitting will be considered in Chapter 6.

2.11 In a symmetrical distribution the mean, the median and the mode (or in cases such as the U-shaped distribution, the antimode) coincide. For skew distributions they differ. There is an interesting empirical relationship between the three quantities which appears to hold for unimodal curves of moderate asymmetry, namely

$$\text{mean} - \text{mode} = 3\ (\text{mean} - \text{median}). \tag{2.13}$$

A mathematical explanation of this relationship is given in Exercise 6.20.

It is a useful mnemonic to observe that the mean, median and mode occur in the same order (or the reverse order) as in the dictionary; and that the median is nearer to the mean than to the mode, just as the corresponding words are nearer together in the dictionary.

In elementary theory the median and the mode have considerable claims to use as measures of location. They are readily interpretable in terms of ordinary ideas—the median is the middle value and the mode is the most popular value—and the median is usually more easily determined than the mean in numerically specified distributions. What gives the arithmetic mean the greater importance in advanced theory is its superior mathematical tractability and certain sampling properties; but the median has compensating advantages—it is, for instance, less dependent on the form of the frequency-distribution than the mean.

2.12 Mention should also be made of a little-used measure of location, the *centre*. This is the variate-value half-way between the terminal points of the distribution if it possesses such points. Its dependence on the terminal points and the use of mathematical distributions which extend to infinity in one or both directions make it unsuitable for many purposes but, like the *range* (**2.17**, below), it has some uses in sampling theory and is very easily calculated. The centre of a set of values which form a sample is sometimes called the "mid-range."

Quantiles

2.13 The concept of median value can be easily extended to locate the curve more accurately by the use of several measures. We may, for example, find the three variate-values which divide the total frequency into four equal parts. The middle one of these will be the median itself; the other two are called the lower and upper *quartiles* respectively. Similarly, we may find the nine variate-values which divide the total frequency into ten equal parts—the *deciles*. Generally we may find the $(n-1)$ variate-values which divide the total frequency into n equal parts—the *quantiles*. Evidently the knowledge of the quantiles for some fairly high n, such as 10, gives a very good idea of the general form of the frequency-distribution. Even the quartiles and the median are valuable general guides. In Chapter 14, we shall study statistics of this type systematically.

2.14 The determination of the quantiles of a numerically specified distribution proceeds as for the median, indeterminacies being resolved by similar conventions. That of the quantiles of a mathematically specified distribution, say the jth quantile, is a matter of solving for x the equation

$$i/n = \int_{-\infty}^{x} dF \tag{2.14}$$

which can be done without difficulty by interpolation when the integral of dF has been tabulated.

Example 2.6

To find the quartiles of the height distribution considered in Example 2.1.

One-quarter of the total frequency is 8585/4 =	2146·25	
Up to the interval 65– there are	1376	members
leaving	770·25	members
In the next interval there are	990	members
Thus the lower quartile is $64\frac{15}{16}+\frac{770\cdot25}{990}$ =	65·71	inches
The upper quartile will be found to be	69·21	inches

We have already found (Example 2.5) that the median μ_e is 67·47 inches

Denoting the quartiles by Q_1 and Q_3 we see that

$$Q_1-\mu_e = -1\cdot76 \text{ inches}$$
$$Q_3-\mu_e = 1\cdot74 \text{ inches}$$

so that the median is almost half-way between the quartiles, an indication of the symmetry of the distribution.

2.15 The quantiles can be easily determined graphically from the graph of the distribution function. Fig. 2.1 gives the curve for the data of Table 1.7. To find

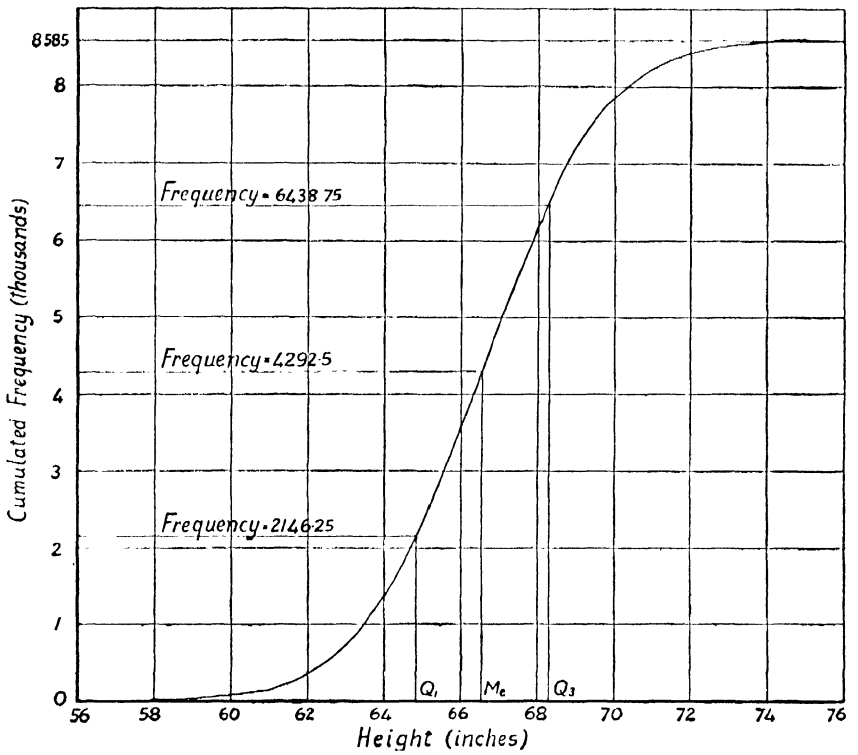


Fig. 2.1—Distribution curve of the data of Table 1.7

(Heights shown to correspond to entries in the Table, e.g. cumulated frequency at 64 inches is the frequency up to and including the range 64– and therefore up to $64\frac{15}{16}$ inches)

the median we determine the ordinate corresponding to the abscissa of $\frac{1}{2}N$ and so on. The positions of the quartiles and the median are shown in Fig. 2.1, and the

reader may care to compare the values obtained by reading the graph by eye with those given in Example 2.6.

Measures of dispersion

2.16 We now proceed to consider the quantities which have been proposed to measure the dispersion of a distribution. They fall into three groups :—

(*a*) Measures of the distance (in terms of the variate) between certain representative values, such as the range, the interdecile range or the interquartile range.

(*b*) Measures compiled from the deviations of every member of the population from some central value, such as the mean deviation from the mean, the mean deviation from the median, and the standard deviation.

(*c*) Measures compiled from the deviations of all the members of the population among themselves, such as the mean difference.

In advanced theory the outstandingly important measure is the standard deviation ; but they all require some mention.

Range and interquantile differences

2.17 The *range* of a distribution is the difference of the greatest and least variate-values borne by its members. As a descriptive measure of a population it has very little use. A knowledge of the whereabouts of the end values obviously tells little about the way the bulk of the distribution is condensed inside the range ; and for distributions of infinite range it is obviously wholly inappropriate.

More useful rough-and-ready measures may be obtained from the quantiles, and there are two such in general use. The interquartile range is the distance between the upper and lower quartiles, and is thus a range which contains one-half the total frequency. The interdecile range (or perhaps, more accurately, the 1–9th interdecile range) is the distance between the first and the ninth decile. Both these measures evidently give some approximate idea of the " spread " of a distribution, and are easily calculable. For this reason they are fairly generally used in elementary descriptive statistics. In advanced theory they suffer from the disadvantage of being relatively difficult to handle mathematically in the theory of sampling.

Mean deviations

2.18 The amount of scatter in a population is evidently measured to some extent by the totality of deviations from the mean. It is easily seen that the sum of these deviations taken with appropriate sign is zero. We may however write

$$\delta_1 = \int_{-\infty}^{\infty} |x-\mu_1'| \, dF \tag{2.15}$$

where the deviations are now taken absolutely, and define δ_1 to be a coefficient of dispersion. We shall call it the mean deviation about the mean.

Similarly for the median μ_e we may write

$$\delta_2 = \int_{-\infty}^{\infty} |x-\mu_e| \, dF \tag{2.16}$$

and call δ_2 the mean deviation about the median.

In future the words " mean deviation " alone will be taken to refer to the mean deviation about the mean.

Both these measures have merits in elementary work, being fairly easily calculable. Once again, however, they are practically excluded from advanced work by their intractability in the theory of sampling.

Standard deviation

2.19 We have seen that the mean about an arbitrary point a is given by

$$\mu_1' = \int_{-\infty}^{\infty} (x-a)\,dF.$$

We may, by analogy with the terminology of Statics, call this the first moment, and define the second moment by

$$\mu_2' = \int_{-\infty}^{\infty} (x-a)^2\,dF. \tag{2.17}$$

The second moment about the mean is written without the prime, thus:

$$\mu_2 = \int_{-\infty}^{\infty} (x-\mu_1')^2\,dF \tag{2.18}$$

and is called the variance. The positive square root of the variance is called the standard deviation, and usually denoted by σ, so that we have

$$\sigma = |\sqrt{\mu_2}|. \tag{2.19}$$

The variance is thus the mean of the *squares* of deviations from the mean. The device of squaring and then taking the square root of the resultant sum in order to obtain the standard deviation may appear a little artificial, but it makes the mathematics of the sampling theory very much simpler than is the case, for example, with the mean deviation.

The calculation of the variance and the standard deviation proceeds by an easy extension of the methods used for the mean. In particular, if b is some arbitrary value,

$$\begin{aligned}\mu_2' \text{ (about } a) &= \int_{-\infty}^{\infty} (x-a)^2\,dF \\ &= \int_{-\infty}^{\infty} \{(x-b)^2 + 2(b-a)(x-b) + (b-a)^2\}\,dF \\ &= \mu_2' \text{ (about } b) + 2(b-a)\mu'_1 \text{ (about } b) + (b-a)^2. \end{aligned} \tag{2.20}$$

If now b is the mean we have

$$\mu_2' = \mu_2 + (\mu_1' - a)^2 \quad \text{or} \quad \mu_2 = \mu_2' - (\mu_1' - a)^2. \tag{2.21}$$

Thus the variance can easily be found from the second moment about an arbitrary point, which can be selected to simplify the calculations.

Example 2.7

To find the mean deviation and the standard deviation for the distribution of men according to height considered in Example 2.1 (Table 1.7).

In the case of the mean deviation for a grouped distribution, the sum of deviations should first be calculated from the centre of the class-interval in which the mean lies

and then reduced to the mean as origin. It so happens that in Table 2.1 the mean fell in the interval taken as origin, so that the preliminary arithmetic already exists in the Table.

The sum of positive deviations is 8763 and that of negative deviations −8584. Hence the sum of deviations regardless of sign is 17,347, the unit being the class-interval and the origin the centre of the interval.

To reduce to the mean as origin, we note that if the number of observations below the mean is N_1 and the number above the mean is N_2, and $d = \mu_1' - a$, we have to add $N_1 d$ to the sum of deviations about the centre of the interval and subtract $N_2 d$. In this case $d = 0{\cdot}02$ (Example 2.1), $N_1 = 4918$, $N_2 = 3667$. Hence we add $(4918-3667)0{\cdot}02 = 25$. Hence the mean deviation

$$\delta_1 = \frac{17{,}347+25}{8585} = 2{\cdot}02 \text{ inches.}^{(*)}$$

For the standard deviation some further calculation is required, as shown in Table 2.2.

Table 2.2—Calculation of the standard deviation for the distribution of Table 1.7 (*some preliminary calculation already carried out in Table* 2.1)

(1) Height, inches	(2) Frequency f	(3) Deviation ξ	(4) $\xi^2 f$
57–	2	− 10	200
58–	4	− 9	324
59–	14	− 8	896
60–	41	− 7	2,009
61–	83	− 6	2,988
62–	169	− 5	4,225
63–	394	− 4	6,304
64–	669	− 3	6,021
65–	990	− 2	3,960
66–	1223	− 1	1,223
67–	1329	0	0
68–	1230	1	1,230
69–	1063	2	4,252
70–	646	3	5,814
71–	392	4	6,272
72–	202	5	5,050
73–	79	6	2,844
74–	32	7	1,568
75–	16	8	1,024
76–	5	9	405
77–	2	10	200
TOTALS	8585	—	56,809

Column (4) shows the sum $\Sigma \xi^2 f$, where f is the actual frequency. We then have, for the second moment about the arbitrary origin,

$$\mu_2' = \frac{56{,}809}{8585} = 6{\cdot}6172.$$

(*) This calculation can be refined slightly—cf. Exercise 2.21.

We have already found in Example 2.1 that

$$\mu_1' - a = \frac{179}{8585} = 0{\cdot}0209.$$

Hence, in virtue of (2.21)

$$\begin{aligned}\mu_2 &= 6{\cdot}6172 - (0{\cdot}0209)^2 \\ &= 6{\cdot}6168 \\ \sigma &= \sqrt{\mu_2} = 2{\cdot}57 \text{ inches.}\end{aligned}$$

It may be noted that the mean deviation is about 80 per cent. of the standard deviation. This relationship often holds approximately for unimodal curves approaching symmetry. The reason will become apparent when we study the so-called " normal " distribution in Chapter 5.(*)

Example 2.8

To find the variance of the distribution

$$dF = \frac{1}{B(p,q)}(1-x)^{p-1}x^{q-1}\,dx, \qquad 0 \leqslant x \leqslant 1\,;\ p,q>0.$$

We have, about the origin,

$$\begin{aligned}\mu_2' &= \frac{1}{B(p,q)}\int_0^1 (1-x)^{p-1}x^{q+1}\,dx \\ &= \frac{B(p,q+2)}{B(p,q)} = \frac{(q+1)q}{(p+q+1)(p+q)}.\end{aligned}$$

We have already found (Example 2.2) that

$$\mu_1' = \frac{q}{p+q}.$$

Thus

$$\begin{aligned}\mu_2 &= \mu_2' - \mu_1'^2 \\ &= \frac{(q+1)q}{(p+q+1)(p+q)} - \frac{q^2}{(p+q)^2} \\ &= \frac{pq}{(p+q+1)(p+q)^2}.\end{aligned}$$

Example 2.9

As for the mean, we may find the variance by differentiating a frequency-generating function. Reverting to the binomial distribution considered in Example 2.3 we have

$$\frac{d^2P}{dt^2} = \Sigma\{j(j-1)t^{j-2}f_j\}$$

and on putting $t = 1$ we have, with moments about zero,

$$\left[\frac{d^2P}{dt^2}\right]_{t=1} = \Sigma\{j(j-1)f_j\} = \mu_2' - \mu_1'. \tag{2.22}$$

(*) M. C. Chakravarty (1948), *Calcutta Statist. Ass. Bull.*, **1**, 187, has shown that, in general, for a set of $n > 2$ values,

$$\begin{aligned}(2/n)^{\frac{1}{2}} \leqslant \delta_1/\sigma &\leqslant (1-n^{-2})^{\frac{1}{2}}, \ n \text{ odd} \\ &\leqslant 1 \quad\quad , \ n \text{ even},\end{aligned}$$

the equalities being attainable.

In this particular case

$$\left[\frac{d^2}{dt^2}\{pt+(1-p)\}^n\right]_{t=1} = n(n-1)p^2.$$

Thus $$\mu_2' = n(n-1)p^2+np,$$

and $$\begin{aligned}\mu_2 &= n(n-1)p^2+np-n^2p^2 \\ &= np(1-p).\end{aligned} \tag{2.23}$$

Sheppard's corrections

2.20 The treatment of the values of a grouped frequency-distribution as if they were concentrated at the mid-points of intervals is an approximation, and in certain circumstances it is possible to make corrections for any distortion introduced thereby. These so-called " Sheppard's corrections " will be discussed at length in the next chapter, but at this stage we may indicate without proof the appropriate correction for the second moment.

If the distribution is continuous and has high order contact with the variate-axis at its extremities, i.e. if it " tails off " smoothly, the crude second moment calculated from grouped frequencies should be corrected by subtracting from it $h^2/12$, where h is the width of the interval. For example, in the height data of Example 2.7, we have $h = 1$, and the corrected second moment is

$$6{\cdot}6168-0{\cdot}0833 = 6{\cdot}5335.$$

The corrected value of σ is $\sqrt{6{\cdot}5335} = 2{\cdot}56$, as against an uncorrected value of 2·57.

Mean difference

2.21 The coefficient of mean difference (not to be confused with mean deviation) is defined by

$$\begin{aligned}\Delta_1 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |x-y|\,dF(x)\,dF(y) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |x-y|f(x)f(y)\,dx\,dy.\end{aligned} \tag{2.24}$$

In the discontinuous case two different formulae arise. We have either

$$\Delta_1 = \frac{1}{N(N-1)}\sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty} |x_j-x_k|f(x_j)f(x_k), \qquad j\neq k, \tag{2.25}$$

the mean difference without repetition, or

$$\Delta_1 = \frac{1}{N^2}\sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty} |x_j-x_k|f(x_j)f(x_k), \tag{2.26}$$

the mean difference with repetition. The difference lies only in the divisor and is unimportant if N is large.

The mean difference is the average of the differences of all the possible pairs of variate-values, taken regardless of sign. In the coefficient with repetition each value is taken with itself, adding of course nothing to the sum of deviations, but resulting in the total number of pairs being N^2. In the coefficient without repetition only different

values are taken, so that the number of pairs is $N(N-1)$. Hence the divisors in (2.25) and (2.26).

2.22 The mean difference, which is due to Gini (1912), has a certain theoretical attraction, being dependent on the spread of the variate-values among themselves and not on the deviations from some central value. It is, however, more difficult to compute than the standard deviation, and the appearance of the absolute values in the defining equations indicates, as for the mean deviation, the appearance of difficulties in the theory of sampling. It might be thought that this inconvenience could be overcome by the definition of a coefficient

$$E^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-y)^2\,dF(x)\,dF(y).$$

This, however, is nothing but twice the variance.

For
$$\begin{aligned} E^2 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\{x^2-2xy+y^2\}\,dF(x)\,dF(y) \\ &= \int_{-\infty}^{\infty}x^2\,dF(x)\int_{-\infty}^{\infty}dF(y)-2\int_{-\infty}^{\infty}x\,dF(x)\int_{-\infty}^{\infty}y\,dF(y) \\ &\qquad\qquad +\int_{-\infty}^{\infty}dF(x)\int_{-\infty}^{\infty}y^2\,dF(y) \\ &= 2\mu_2'-2\mu_1'^2 \\ &= 2\mu_2. \end{aligned} \tag{2.27}$$

This interesting relation shows that the variance may in fact be defined as half the mean square of all possible variate-differences, that is to say, without reference to deviations from a central value, the mean.

Coefficients of variation : standard measure

2.23 The foregoing measures of dispersion have all been expressed in terms of units of the variate. It is thus difficult to compare dispersions in different populations unless the units happen to be identical; and this has led to a search for measures which shall be independent of the variate scale, that is to say, shall be pure numbers.

Several coefficients of this kind may be constructed, such as the mean deviation divided by the mean or by the median. Only two have been used at all extensively in practice, Karl Pearson's coefficient of variation, defined by

$$V = 100\frac{\sigma}{\mu_1'}, \tag{2.28}$$

and Gini's coefficient of concentration, defined by

$$G = \frac{\Delta_1}{2\mu_1'}. \tag{2.29}$$

These coefficients both suffer from the disadvantage of being affected very much by μ_1', the value of the mean measured from some arbitrary origin, and are not usually employed unless there is a natural origin of measurement or comparisons are being made between distributions with similar origins.

2.24 For our purposes, comparability may be attained in a somewhat different way. Let us take σ itself as a new unit and express the frequency function in terms of a new variable y related to x by

$$y = \frac{x - \mu_1'}{\sigma}. \tag{2.30}$$

Any distribution expressed in this way has zero mean and unit variance. It is then said to be expressed in standard measure, or standardized. Two distributions in standard measure can be readily compared in regard to form, skewness, and other qualities, though not of course in regard to mean and variance.

Concentration

2.25 Gini's coefficient of concentration arises in a natural way from the following approach :—

Writing, as usual

$$F(x) = \int_{-\infty}^{x} f(x)\,dx$$

let us define

$$\Phi(x) = \frac{1}{\mu_1'}\int_{-\infty}^{x} x f(x)\,dx \tag{2.31}$$

$\Phi(x)$ exists, of course, only if μ_1' exists. Just as $F(x)$ varies from 0 to 1, $\Phi(x)$ varies from 0 to 1 provided that the origin is taken to the left of the start of the frequency-distribution, assumed to be finite. $\Phi(x)$ may be called the *incomplete* first moment.

Now (2.31) may be regarded as defining a relationship between the variables F and Φ in terms of parametric equations in x. The curve whose ordinate and abscissa are Φ and F is called the curve of concentration. Such a curve is shown in Fig. 2.2.

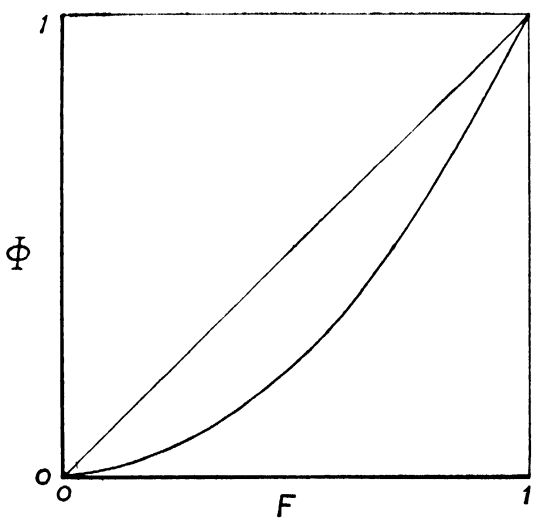


Fig. 2.2—Curve of concentration

The curve of concentration must be convex to the F-axis, for we have

$$\mu_1' \frac{d\Phi}{dF} = \frac{xf(x)}{f(x)} = x,$$

which is positive since our origin is taken to the left of the start of the distribution. Also

$$\mu_1' \frac{d^2\Phi}{dF^2} = \frac{dx}{dF} = \frac{1}{f(x)} > 0.$$

Thus the tangent to the curve makes a positive acute angle with the F-axis, and the angle increases as F increases; in other words, the curve is convex to the F-axis.

The area between the concentration curve and the line $F = \Phi$ is called the area of concentration. We proceed to show that it is equal to one-half the coefficient of concentration.

In fact, we have from Fig. 2.2

$$2 \text{ (area of concentration)} = \int_0^1 F\,d\Phi - \int_0^1 \Phi\,dF$$

and thus

$$\begin{aligned} 2\mu_1' \text{ (area)} &= \int_{-\infty}^{\infty} F(x)\,x\,dF(x) - \mu_1' \int_{-\infty}^{\infty} \Phi(x)\,dF(x) \\ &= \int_{-\infty}^{\infty} x\,dF(x) \int_{-\infty}^{x} dF(y) - \int_{-\infty}^{\infty} dF(x) \int_{-\infty}^{x} y\,dF(y) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{x} (x-y)\,dF(x)\,dF(y). \end{aligned}$$

Now $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x-y)\,dF(x)\,dF(y) = 0$, and hence

$$\begin{aligned} 2\mu_1' \text{ (area)} &= \tfrac{1}{2}\left[\int_{-\infty}^{\infty} \int_{-\infty}^{x} (x-y)\,dF(x)\,dF(y) + \int_{-\infty}^{\infty} \int_{x}^{\infty} (y-x)\,dF(x)\,dF(y)\right] \\ &= \tfrac{1}{2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |x-y|\,dF(x)\,dF(y) \\ &= \tfrac{1}{2}\Delta_1. \end{aligned}$$

Thus the area of concentration is equal to $\frac{1}{4}\Delta_1/\mu_1'$, namely one-half of the coefficient of concentration G at (2.29).

2.26 Various methods have been given for calculating the mean difference. The following is probably the simplest, particularly for distributions specified in equal group-intervals.

Let us, without loss of generality, take an origin at the start of the distribution. We may then write

$$\sum_{j=1}^{N} \sum_{k=1}^{N} |x_j - x_k| = 2\Sigma'(x_j - x_k),$$

the summation Σ' being taken over values such that $j \geqslant k$. We have also

$$x_j - x_k = (x_j - x_{j-1}) + (x_{j-1} - x_{j-2}) + \ldots + (x_{k+1} - x_k).$$

Thus

$$\Sigma'(x_j - x_k) = \sum_{h=1}^{N-1} C_h (x_{h+1} - x_h),$$

where C_h is the number of terms of type $(x - x_k)$ in Σ' containing $x_{h+1} - x_h$. Since h is the number of values of j less than or equal to h (the origin being at the start of the distribution) and $N-h$ the number greater than or equal to $h+1$, we have $C_h = h(N-h)$, and thus

$$\begin{aligned}\Delta_1 &= \frac{2}{N^2}\Sigma'(x_j - x_k) \\ &= \frac{2}{N^2}\sum_{h=1}^{N-1} h(N-h)(x_{h+1} - x_h). \end{aligned} \tag{2.32}$$

This form is particularly useful if all the intervals are equal. F_h being the distribution function of x_h we then have

$$\begin{aligned}\Delta_1 &= \frac{2}{N^2}\sum_{h=1}^{N-1}(NF_h)(N - NF_h) \\ &= 2\sum_{h=1}^{N-1} F_h(1 - F_h). \end{aligned} \tag{2.33}$$

If the actual cumulated frequency for x_h is G_h we have

$$\Delta_1 = \frac{2}{N^2}\sum_{h=1}^{N-1} G_h(N - G_h), \tag{2.34}$$

the most convenient form in practice.

Example 2.10

Returning once more to the height distribution considered in previous examples, we may calculate $\Sigma G_h(N - G_h)$ as in Table 2.3.

Table 2.3—Calculation of the mean difference for the height distribution of Table 1.7

Height, inches	Frequency	G_h	$N - G_h$	$G_h(N - G_h)$
57–	2	2	8583	17,166
58–	4	6	8579	51,474
59–	14	20	8565	171,300
60–	41	61	8524	519,964
61–	83	144	8441	1,215,504
62–	169	313	8272	2,589,136
63–	394	707	7878	5,569,746
64–	669	1376	7209	9,919,584
65–	990	2366	6219	14,714,154
66–	1223	3589	4996	17,930,644
67–	1329	4918	3667	18,034,306
68–	1230	6148	2437	14,982,676
69–	1063	7211	1374	9,907,914
70–	646	7857	728	5,719,896
71–	392	8249	336	2,771,664
72–	202	8451	134	1,132,434
73–	79	8530	55	469,150
74–	32	8562	23	196,926
75–	16	8578	7	60,046
76–	5	8583	2	17,166
77–	2	8585	—	—
TOTALS	8585	—	—	105,990,850

We have, from (2.34), for the mean difference with repetition,

$$\Delta_1 = \frac{2\times 105{,}990{,}850}{8585^2} = 2{\cdot}88 \text{ inches}$$

as against a mean deviation of 2·02 inches and a standard deviation of 2·57 inches (Example 2.7). There is, of course, nothing inconsistent in the difference between these values. The coefficients are different in nature, and there is no reason why their numerical values in any particular case should approach equality.

Mean values in general

2.27 The concept of the arithmetic mean of a variate can be easily extended to that of the mean of a function, say $h(x)$. In fact we define

$$\mathrm{E}(h) = \int_{-\infty}^{\infty} h(x)\,dF, \tag{2.35}$$

subject, of course, to the existence of the sum or integral on the right. Mean values such as this play a fundamental role in the theory of statistics and of probability. The notation E derives from the fact that in probability theory such values are often known as " expected values " or " expectations." It is a remarkable and fruitful fact that in the theory of sampling we are often able to determine certain mean values, especially those of powers of the variate, more easily than the corresponding frequency functions.

Generally, for a function $h(x_1, x_2, .., x_n)$ we have

$$\mathrm{E}(h) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(x_1, .., x_n)\,dF(x_1, .., x_n) \tag{2.36}$$

In particular if h is a function of x_1 only, we have, for the multivariate distribution, on integrating with respect to $x_2, .., x_n$,

$$\mathrm{E}(h) = \int_{-\infty}^{\infty} h(x_1)\,dF(x_1) \tag{2.37}$$

where $F(x_1)$ is the (marginal) distribution function of x_1. This is required for consistence with (2.35) and shows us that the mean value of a variate (or some function of it) is the same whether the variate forms part of a multivariate complex or not.

2.28 Two simple but important properties of mean values are to be noted:

(*a*) The mean of a sum is the sum of the means. If h_1 and h_2 are two functions,

$$\mathrm{E}(h_1+h_2) = \mathrm{E}(h_1)+\mathrm{E}(h_2). \tag{2.38}$$

This follows at once from the definitions.

(*b*) The mean value of a product of two functions is the product of their mean values if each function depends on a set of variates independent of the set on which the other depends. For example, let x_1 and x_2 represent independent variates. Their

distribution function then factorizes into two components and we have

$$\begin{aligned} \mathrm{E}(h_1 h_2) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_1(x_1)\,h_2(x_2)\,dF(x_1, x_2) \\ &= \int_{-\infty}^{\infty} h_1(x_1)\,dF(x_1, \infty) \int_{-\infty}^{\infty} h_2(x_2)\,dF(\infty, x_2) \\ &= \mathrm{E}(h_1)\,E(h_2). \end{aligned} \tag{2.39}$$

Equation (2.38) is true even if the variates are dependent; (2.39) will not be true in general for dependent variates.

EXERCISES

2.1 Show that the mean deviation about an arbitrary point is least when that point is the median.

2.2 Show that the mean (about the origin) of the discontinuous distribution whose frequencies at 0, 1, 2,.., r,... are

$$e^{-\lambda}\left(1, \lambda, \frac{\lambda^2}{2!}, .., \frac{\lambda^r}{r!}, \ldots\right)$$

is λ, and that the variance is also λ.

2.3 Show that, if deviations are small compared with the value of the mean, we have approximately, for the geometric and harmonic means,

$$G = \mu_1'\left(1 - \frac{1}{2}\frac{\sigma^2}{\mu_1'^2}\right)$$

$$H = \mu_1'\left(1 - \frac{\sigma^2}{\mu_1'^2}\right)$$

and hence that

$$\mu_1' - 2G + H = 0.$$

2.4 Show that the mean deviation about the mean is not greater than the standard deviation; and that the mean difference cannot exceed $\sqrt{2}$ times the standard deviation.

2.5 Show that for the "rectangular" population

$$\begin{aligned} dF &= dx, \qquad 0 \leqslant x \leqslant 1, \\ \mu_1' \text{ (about the origin)} &= \tfrac{1}{2} \\ \mu_2 &= \tfrac{1}{12} \\ \text{mean deviation} &= \tfrac{1}{4} \\ \Delta_1 &= \tfrac{1}{3}. \end{aligned}$$

2.6 Show that for the distribution (the negative exponential)

$$dF = k\,e^{-x/\sigma}\,dx, \qquad 0 \leqslant x \leqslant \infty,$$

the mean, standard deviation and mean difference are all equal to σ; and that the interquartile range is $\sigma \log_e 3$.

2.7 Show that for the distribution

$$\begin{aligned} dF &= k\,e^{-\frac{1}{2}\chi^2}\,\chi^{\nu-1}\,d\chi, \qquad 0 \leqslant \chi \leqslant \infty, \\ \mu_1 \text{ (about the origin)} &= \sqrt{2}\,\Gamma\{\tfrac{1}{2}(\nu+1)\}/\Gamma(\tfrac{1}{2}\nu) \\ \mu_2' &= \nu. \end{aligned}$$

2.8 Show that if a range of six times the standard deviation contains at least 18 class-intervals, Sheppard's correction will make a difference of less than 0·5 per cent. in the uncorrected value of the standard deviation.

2.9 Show that for a continuous distribution whose mean exists

$$\Delta_1 = 2\int_{-\infty}^{\infty} F(x)\{1-F(x)\}dx = 4\int_{-\infty}^{\infty} x\{F(x)-\tfrac{1}{2}\}\,dF.$$

2.10 If the variate-values of a distribution are $x_1, \ldots, x_N$ in ascending order of magnitude and

$$s_r = \sum_{j=1}^{r} x_j \qquad U = \sum_{r=1}^{N} s_r$$

$$t_r = \sum_{j=1}^{r} x_{N-j+1} \qquad V = \sum_{r=1}^{N} t_r$$

show that

$$\Delta_1 = \frac{2}{N^2}(V-U)$$

$$= \frac{2}{N^2}\{N(N+1)\mu_1' - 2U\}.$$

2.11 Referring to Exercise 1.14, show that the mean of the negative binomial distribution is m and that its variance is $m\{m/k+1\}$.

2.12 Show that the variance of the distribution

$$dF = \frac{k}{1+x^4}dx, \qquad -\infty < x \leqslant \infty,$$

is unity.

2.13 With reference to Exercise 1.10 find the mean value of $x_1^2+x_2^2+x_3^2$ (a) directly and (b) from the distribution of r^2.

2.14 For the distribution of Exercise 1.11 show that the mean value of x^2 is σ_1^2 and that of y^2 is σ_2^2. By differentiating $\int dF = 1$ with respect to ρ show that $\mathrm{E}(xy) = \rho\sigma_1\sigma_2$.

2.15 A variate takes only non-negative values and has mean μ_1'. Show that for any positive t

$$F(t) > 1-\mu_1'/t.$$

2.16 Show that if a distribution

$$dF = \frac{1}{\sqrt{(2\pi)}}e^{-\frac{1}{2}x^2}dx, \qquad -\infty \leqslant x \leqslant \infty,$$

is transformed by $x = a\log(y-b)+c$, the distribution of y has

$$\frac{\text{mean}-\text{mode}}{\text{mean}-\text{median}} = \frac{\exp\left(\frac{1}{2a^2}\right)-\exp\left(-\frac{1}{a^2}\right)}{\exp\left(\frac{1}{2a^2}\right)-1},$$

and that this ratio tends to the value 3 as $a \longrightarrow \infty$.

2.17 A non-increasing frequency function $f(x)$ on the range 0 to ∞ has mean μ. Show that $\mu \geqslant 1/\{2\,f(0)\}$. Illustrate on the distribution

$$dF = \frac{4}{\pi a^2}(a^2-x^2)^{\frac{1}{2}}\,dx, \qquad 0 \leqslant x \leqslant a.$$

2.18 In the manner of the previous exercise find an inequality involving the second moment of $f(x)$. Hence show that for a continuous differentiable frequency function with a single mode the ordinate at the mode is not less than $1/(\sqrt{3}\sigma)$ where σ is the square root of the second moment about the mode.

2.19 Show that for a distribution with existent mean the mean difference Δ_1 also exists. Illustrate by reference to the so-called Pareto curve

$$dF = \frac{k}{x^\alpha}dx, \qquad 0 < a \leqslant x \leqslant \infty,\ \alpha > 1,$$

where the mean may exist but the variance does not exist for $\alpha \leqslant 3$.

2.20 For the distribution of Exercise 1.11 show that if $\mathrm{E}(x^2)\mathrm{E}(y^2) = \mathrm{E}(x^2y^2)$ then the variates are independent.

2.21 In the mean deviation calculation for grouped data in Example 2.7, the interval in which the mean fell made contribution $f\,|\,d\,|$ to the sum of deviations. Show that if the frequency f is split into frequencies above and below μ_1' proportionately to the fractions of the interval-width h above and below μ_1', and these frequencies are taken to be concentrated at the centre of their intervals, the sum of deviations is increased by $fh\left(\frac{|d|}{h} - \frac{1}{2}\right)^2$. Confirm that, in Example 2.7, this correction increases δ_1 by 0·04.

CHAPTER 3

MOMENTS AND CUMULANTS

Definition of moments

3.1 In the previous chapter we defined the first moment (arithmetic mean) about an arbitrary point a by the Stieltjes integral

$$\mu_1' = \int_{-\infty}^{\infty} (x-a)\, dF \tag{3.1}$$

and the second moment about the point by

$$\mu_2' = \int_{-\infty}^{\infty} (x-a)^2\, dF. \tag{3.2}$$

In generalization of these equations we may define a series of coefficients μ_r', $r = 1, 2\ldots$, by the relation

$$\mu_r' = \int_{-\infty}^{\infty} (x-a)^r\, dF. \tag{3.3}$$

μ_r' is called the moment of order r about the point a. When a is the mean μ_1' we write the moment without the prime,

$$\mu_r = \int_{-\infty}^{\infty} (x-\mu_1')^r\, dF, \tag{3.4}$$

and call it a central moment or a mean-moment. In particular

$$\mu_1 = 0,$$

and we may also define a moment of zero order

$$\mu_0' = \mu_0 = \int_{-\infty}^{\infty} dF = 1.$$

It is assumed that when reference is made to the rth moment of a particular distribution, the appropriate integral (3.3) converges for that distribution. As will be seen later, some of the theoretical distributions encountered in statistics do not possess moments of all orders ; some, in fact, possess only a few moments of low order, and one or two do not possess any, except of course the moment of order zero. It is obvious that if a moment of order r exists, moments of all lower orders will exist.

3.2 If a and b are two variate-values, let $b-a = c$ and denote the moments about a and b by $\mu'(a)$ and $\mu'(b)$ respectively. Then we have, by the binomial theorem,

$$(x-a)^r = (x-b+b-a)^r = (x-b+c)^r$$

$$= \sum_{j=0}^{r} \binom{r}{j} (x-b)^{r-j} c^j.$$

Hence

$$\begin{aligned}\mu'_r(a) &= \int_{-\infty}^{\infty}(x-a)^r\,dF \\ &= \int_{-\infty}^{\infty}\sum_{j=0}^{r}\binom{r}{j}(x-b)^{r-j}c^j\,dF \\ &= \sum_{j-0}^{r}\binom{r}{j}c^j\int_{-\infty}^{\infty}(x-b)^{r-j}\,dF \\ &= \sum_{j=0}^{r}\binom{r}{j}\mu'_{r-j}(b)\,c^j. \end{aligned} \tag{3.5}$$

This equation gives the rth moment about a in terms of the rth and lower moments about b. It may be written in a symbolic form which will be found to provide a useful mnemonic, namely

$$\mu'_r(a) = \{\mu'(b)+c\}^r$$

with the convention that the expression on the right is to be expanded binomially and the form $\{\mu'(b)\}^j$ replaced by $\mu'_j(b)$.

The equation (3.5) is of particular importance if one of the values a or b is the mean of the distribution. In this case we have

$$\mu'_r = \sum_{j=0}^{r}\binom{r}{j}\mu_{r-j}\mu'^j_1. \tag{3.6}$$

$$\mu_r = \sum_{j=0}^{r}\binom{r}{j}\mu'_{r-j}(-\mu'_1)^j. \tag{3.7}$$

In particular

$$\left.\begin{aligned}\mu'_2 &= \mu_2+\mu'^2_1 \\ \mu'_3 &= \mu_3+3\mu'_1\mu_2+\mu'^3_1 \\ \mu'_4 &= \mu_4+4\mu'_1\mu_3+6\mu'^2_1\mu_2+\mu'^4_1\end{aligned}\right\} \tag{3.8}$$

and

$$\left.\begin{aligned}\mu_2 &= \mu'_2-\mu'^2_1 \\ \mu_3 &= \mu'_3-3\mu'_1\mu'_2+2\mu'^3_1 \\ \mu_4 &= \mu'_4-4\mu'_1\mu'_3+6\mu'^2_1\mu'_2-3\mu'^4_1\end{aligned}\right\} \tag{3.9}$$

Calculation of moments

3.3 For a distribution specified numerically in a frequency table the calculation of moments of third and higher orders is akin to that of the first and second moments. For grouped data (high order moments are hardly ever required for ungrouped data) the observations are regarded as concentrated at the mid-points of intervals; a convenient arbitrary origin a is chosen, the moments about a calculated, and then if necessary the moments about the mean are ascertained from (3.6) or (3.7). The effect of grouping may be corrected for in certain cases.

In practice numerical moments of order higher than the fourth are rarely required, being so sensitive to sampling fluctuations that values computed from moderate numbers of observations are subject to a large margin of error.

There are two methods in general use for arriving at the moments about an arbitrary origin. The first is an immediate generalization of the methods used in Chapter 2 for the first two moments. The second will be considered in **3.10** in connection with factorial moments.

Example 3.1

To find the first four moments about the mean of the distribution of Australian marriages of Table 1.8.

Until the last stage we work in units of three years, the variate interval. A working mean is taken at 28·5 years. To check the arithmetic we use an identity of type

$$(x+1)^3 = x^3+3x^2+3x+1,$$
$$(x+1)^4 = x^4+4x^3+6x^2+4x+1.$$

Thus, for instance, the value of $g(x)(x+1)^r$ is found in addition to that of $g(x)x^r$ and the two checked by identities such as

$$\Sigma g(x)(x+1)^3 = \Sigma g(x)x^3+3\Sigma g(x)x^2+3\Sigma g(x)x+\Sigma g(x),$$

$g(x)$ being the actual frequencies. The arithmetical work is shown in Table 3.1.

Table 3.1—Calculation of the first four moments of the distribution of marriages of Table 1.8

Mid-value of inter-vals, years	g	x	xg	$(x+1)g$	x^2g	$(x+1)^2g$	x^3g	$(x+1)^3g$	x^4g	$(x+1)^4g$
16·5	294	−4	− 1,176	− 882	4,704	2,646	− 18,816	− 7,938	75,264	23,814
19·5	10,995	−3	− 32,985	−21,990	98,955	43,980	−296,865	− 87 960	890,595	175,920
22·5	61,001	−2	−122,002	−61,001	244,004	61,001	−488,008	− 61,001	976,016	61,001
25·5	73,054	−1	− 73,054	−83,873	73,054	—	− 73,054	−156,899	73,054	—
28·5	56,501	0	−229,217	56,501	—	56,501	−876,743	56,501	—	56,501
31·5	33,478	1	33,478	66,956	33,478	133,912	33,478	267,824	33,478	535,648
34·5	20,569	2	41,138	61,707	82,276	185,121	164,552	555,363	329,104	1,666,089
37·5	14,281	3	42,843	57,124	128,529	228,496	385,587	913,984	1,156,761	3,655,936
40·5	9,320	4	37,280	46,600	149,120	233,000	596,480	1,165,000	2,385,920	5,825,000
43·5	6,236	5	31,180	37,416	155,900	224,496	779,500	1,346,976	3,897,500	8,081,856
46·5	4,770	6	28,620	33,390	171,720	233,730	1,030,320	1,636,110	6,181,920	11,452,770
49·5	3,620	7	25,340	28,960	177,380	231,680	1,241,660	1,853,440	8,691,620	14,827,520
52·5	2,190	8	17,520	19,710	140,160	177,390	1,121,280	1,596,510	8,970,240	14,368,590
55·5	1,655	9	14,895	16,550	134,055	165,500	1,206,495	1,655,000	10,858,455	16,550,000
58·5	1,100	10	11,000	12,100	110,000	133,100	1,100,000	1,464,100	11,000,000	16,105,100
61·5	810	11	8,910	9,720	98,010	116,640	1,078,110	1,399,680	11,859,210	16,796,160
64·5	649	12	7,788	8,437	93,456	109,681	1,121,472	1,425,853	13,457,664	18,536,089
67·5	487	13	6,331	6,818	82,303	95,452	1,069,939	1,336,328	13,909,207	18,708,592
70·5	326	14	4,564	4,890	63,896	73,350	894,544	1,100,250	12,523,616	16,503,750
73·5	211	15	3,165	3,376	47,475	54,016	712,125	864,256	10,681,875	13,828,096
76·5	119	16	1,904	2,023	30,464	34,391	487,424	584,647	7,798,784	9,938,999
79·5	73	17	1,241	1,314	21,097	23,652	358,649	425,736	6,097,033	7,663,248
82·5	27	18	486	513	8,748	9,747	157,464	185,193	2,834,352	3,518,667
85·5	14	19	266	280	5,054	5,600	96,026	112,000	1,824,494	2,240,000
88·5	5	20	100	105	2,000	2,205	40,000	46,305	800,000	972,405
TOTALS OF +VE TERMS	301,785	—	318,049	474,490	2,155,838	2,635,287	13,675,105	19,991,056	137,306,162	202,091,751

From this table we find

$$\begin{aligned}\Sigma(xg) &= 88{,}832\\ \Sigma(x^2g) &= 2{,}155{,}838\\ \Sigma(x^3g) &= 12{,}798{,}362\\ \Sigma(x^4g) &= 137{,}306{,}162.\end{aligned}$$

The values will be found to check and we have, about the working mean, on dividing by the total frequency 301,785,

$$\begin{aligned}\mu_1' &= 0{\cdot}294{,}355{,}253\\ \mu_2' &= 7{\cdot}143{,}622{,}115\\ \mu_3' &= 42{\cdot}408{,}873{,}867\\ \mu_4' &= 454{\cdot}980{,}075{,}219.\end{aligned}$$

For the moments about the mean, substitution in equations (3.9) gives

$$\begin{aligned}\mu_2 &= 7{\cdot}056{,}977\\ \mu_3 &= 36{\cdot}151{,}595\\ \mu_4 &= 408{\cdot}738{,}210.\end{aligned}$$

These are expressed in class-intervals, which are units of three years. To express the results in units of one year we multiply the rth moment by 3^r, e.g.

$$\mu_2 = 7{\cdot}056{,}977 \times 9 = 63{\cdot}512{,}79.$$

3.4 If a distribution is specified mathematically the determination of moments is equivalent to the evaluation of certain sums or integrals. It is usually necessary to consider whether the moments exist. Some examples will illustrate the general principles involved.

Example 3.2

Consider the binomial distribution of Examples 2.3 and 2.9. We will write $q = 1-p$ so that the generating function of the distribution is $(pt+q)^n$. In the manner of those examples we have

$$\begin{aligned}\left[\frac{d^3P}{dt^3}\right]_{t=1} &= \sum_{j=0}^{n} j(j-1)(j-2)f_j\\ &= \Sigma(j^3f_j) - 3\Sigma(j^2f_j) + 2\Sigma(jf_j)\\ &= \mu_3' - 3\mu_2' + 2\mu_1',\end{aligned}$$

the moments being measured from zero origin. This gives us

$$\mu_3' = 3\mu_2' - 2\mu_1' + n(n-1)(n-2)p^3.$$

We have already found that

$$\mu_2' = n(n-1)p^2 + np, \qquad \mu_1' = np,$$

and hence we find, after substitution and reduction,

$$\mu_3' = n(n-1)(n-2)p^3 + 3n(n-1)p^2 - 2np.$$

We are usually more interested in the moments about the mean, and from (3.9) we have

$$\mu_3 = \mu_3' - 3\mu_1'\mu_2' + 2\mu_1'^3$$

which, on substitution on the right, reduces to

$$\mu_3 = np(p-1)(2p-1) \tag{3.10}$$

$$= npq(q-p). \tag{3.11}$$

In a similar manner we find

$$\mu_4 = 3n^2p^2(1-p)^2+np(1-p)(1-6p+6p^2) \tag{3.12}$$

$$= 3n^2p^2q^2+npq(1-6pq). \tag{3.13}$$

The process may evidently be extended to give as many moments as are desired. But we shall see later that there are easier ways of finding the moments of the binomial distribution.

Example 3.3

Consider the distribution

$$dF = \frac{k}{(1+x^2)^m}dx, \qquad -\infty \leqslant x \leqslant \infty\ ;\ m \geqslant 1.$$

This is a unimodal distribution symmetrical about $x = 0$. All existent moments of odd order about the origin therefore vanish. The constant k is given by the equation

$$1 = k\int_{-\infty}^{\infty} \frac{dx}{(1+x^2)^m}$$

$$= k\frac{\Gamma(\frac{1}{2})\Gamma(m-\frac{1}{2})}{\Gamma(m)}.$$

The moment about the mean of order $2r$, if it exists, is given by

$$\mu_{2r} = k\int_{-\infty}^{\infty} \frac{x^{2r}}{(1+x^2)^m}dx,$$

and this integral converges if and only if

$$2m > 2r+1.$$

Thus the moments about the origin of order $< (2m-1)$ exist and those of higher order do not.

If $m = 1$ it may be noted that the integral $\int_{-\infty}^{\infty} \frac{kx\,dx}{1+x^2}$ is not completely convergent, i.e. $\lim_{n\to\infty,\, n'\to\infty} \int_{-n}^{n'} \frac{kx\,dx}{(1+x^2)}$ does not exist, although the principal value

$$\lim_{n\to\infty} \int_{-n}^{n} \frac{kx\,dx}{(1+x^2)}$$

does exist and is equal to zero. It is a matter of convention whether we regard the distribution as possessing a mean in this case.(*) For $m > 1$ the mean exists and is located at the origin.

(*) A disadvantage of the convention that the distribution has a mean is that for variates of such a kind it is no longer necessarily true that the mean of a sum is the sum of the means. Cf. Fréchet (1937), page 45, and Exercise 7.19.

Making the substitution $z = \dfrac{1}{1+x^2}$ in the formula for μ_{2r}, we find

$$\mu_{2r} = k\int_0^1 (1-z)^{r-\frac{1}{2}} z^{m-r-\frac{3}{2}}\,dz$$
$$= k\frac{\Gamma(r+\frac{1}{2})\,\Gamma(m-r-\frac{1}{2})}{\Gamma(m)}$$

and on substituting for k,

$$\mu_{2r} = \frac{\Gamma(r+\frac{1}{2})\,\Gamma(m-r-\frac{1}{2})}{\Gamma(\frac{1}{2})\,\Gamma(m-\frac{1}{2})} \quad \text{if } 2m > 2r+1.$$

Example 3.4

Consider the " normal " distribution

$$dF = \frac{1}{\sigma\sqrt{(2\pi)}} e^{-\frac{x^2}{2\sigma^2}}\,dx, \qquad -\infty \leqslant x \leqslant \infty. \tag{3.14}$$

This is symmetrical about the origin. All moments exist, those of odd order vanishing. Thus

$$\mu_{2r} = \frac{1}{\sigma\sqrt{(2\pi)}}\int_{-\infty}^{\infty} x^{2r} e^{-\frac{x^2}{2\sigma^2}}\,dx.$$

This may be evaluated by partial integration, but a more direct method is as follows.

Consider the integral

$$M(t) = \frac{1}{\sigma\sqrt{(2\pi)}}\int_{-\infty}^{\infty} e^{tx} e^{-\frac{x^2}{2\sigma^2}}\,dx$$
$$= \exp(\tfrac{1}{2}\sigma^2 t^2). \tag{3.15}$$

We have, for all real values of t,

$$e^{tx} e^{-\frac{x^2}{2\sigma^2}} = \sum_{r=0}^{\infty}\left(\frac{t^r}{r!} x^r e^{-\frac{x^2}{2\sigma^2}}\right).$$

The series on the right is uniformly convergent in x and may be integrated term by term if the resulting series is uniformly convergent. We then have

$$M(t) = \sum_{r=0}^{\infty}\left(\frac{t^r}{r!}\mu_r\right).$$

In other words, μ_r is the coefficient of $t^r/r!$ in $\exp(\frac{1}{2}\sigma^2 t^2)$ and hence

$$\mu_{2r} = \frac{\sigma^{2r}}{2^r}\frac{(2r)!}{r!}. \tag{3.16}$$

Moment-generating functions and characteristic functions

3.5 The previous example shows that in some cases we can derive from the distribution- or the frequency-function a function $M(t)$ which, when expanded in powers of t, will yield the moments of the distribution as the coefficients of those powers. Such a function is accordingly called a moment-generating function. If the frequency-generating function is $P(t)$, the moment-generating function is simply $P(e^t)$. The name is conveniently abbreviated to m.g.f.

For many frequency functions the integral $\int_{-\infty}^{\infty} e^{tx}\,dF$ or the sum $\{\Sigma e^{tx_j} f(x_j)\}$ does not exist for real values of t. This is, for example, true of the function $dF = k(1+x^2)^{-m}\,dx$ for finite positive values of m. A more serviceable auxiliary function is

$$\phi(t) = \int_{-\infty}^{\infty} e^{itx}\,dF \quad (t \text{ real}) \tag{3.17}$$

where i is the pure imaginary quantity sometimes written as $\sqrt{(-1)}$. This is known as the characteristic function (often abbreviated to c.f.) and is of great theoretical importance. It will be seen in Chapter 4 that under certain general conditions the characteristic function determines and is completely determined by the distribution function. It also yields many valuable results in the theory of sampling.

Since by the nature of the distribution function the integral $\int_{-\infty}^{\infty} dF$ converges,

$$|\phi(t)| \leqslant \int_{-\infty}^{\infty} |e^{itx}|\,dF = 1$$

and hence the Stieltjes integral (3.17) converges absolutely and uniformly in t. It may therefore be integrated under the summation signs with respect to t, and may be differentiated provided that the resulting expressions exist and are uniformly convergent. We have, for example, writing D_t for $\frac{d}{dt}$,

$$D_t^r \phi(t) = i^r \int_{-\infty}^{\infty} e^{itx} x^r\,dF,$$

and hence, putting $t = 0$,

$$\mu_r' = (-i)^r [D_t^r \phi(t)]_{t=0} \tag{3.18}$$

provided that μ_r exists. If $\phi(t)$ be expanded in powers of t, μ_r' must thus be equal to the coefficient of $(it)^r/r!$ in the expansion. Thus the characteristic function is also a moment-generating function. For many formal purposes it is sufficient to write, say, θ instead of it and to ignore the fact that it is imaginary.

Example 3.5

Consider again the binomial distribution generated by the expansion of $(p+q)^n$. The frequency-generating function is $(pt+q)^n$ and hence a moment-generating function is

$$M(t) = (pe^t + q)^n.$$

Hence

$$\begin{aligned} \mu_1' &= \left[\frac{d}{dt} M(t)\right]_{t=0} \\ &= np. \\ \mu_2' &= \left[\frac{d^2}{dt^2} M(t)\right]_{t=0} \\ &= np + n(n-1)p^2, \end{aligned}$$

and so on.

Example 3.6

Consider the distribution

$$dF = \frac{a^\gamma}{\Gamma(\gamma)} x^{\gamma-1} e^{-ax}\, dx, \qquad 0 \leqslant x \leqslant \infty\ ;\ a>0,\ \gamma>0,$$

which is known as Pearson's Type III (cf. Chapter 6). The distribution may have a variety of shapes, depending on the value of γ, but moments of all orders exist in virtue of the convergence of the integral $\int_0^\infty x^r e^{-ax}\, dx$, the Γ-function integral, for $r > -1$ We have then, for the characteristic function,

$$\phi(t) = \frac{a^\gamma}{\Gamma(\gamma)} \int_0^\infty e^{x(-a+it)} x^{\gamma-1}\, dx.$$

By the substitution $z = x(a-it)$ this becomes

$$\begin{aligned} \phi(t) &= \frac{a^\gamma}{\Gamma(\gamma)(a-it)^\gamma} \int_0^\infty e^{-z} z^{\gamma-1}\, dz \\ &= \frac{1}{\left(1-\frac{it}{a}\right)^\gamma}, \end{aligned} \tag{3.19}$$

since $\int_0^\infty e^{-z} z^{\gamma-1}\, dz = \Gamma(\gamma)$ whether z is real or complex.
Hence

$$\phi(t) = 1 + \gamma\frac{it}{a} + \frac{\gamma(\gamma+1)}{2!}\left(\frac{it}{a}\right)^2 + \ldots$$

and thus

$$\begin{aligned} \mu_1' &= \frac{\gamma}{a}, \\ \mu_2' &= \frac{\gamma(\gamma+1)}{a^2}, \\ \mu_3' &= \frac{\gamma(\gamma+1)(\gamma+2)}{a^3}, \end{aligned}$$

and so on. In particular,

$$\left.\begin{aligned} \mu_2 &= \frac{\gamma}{a^2}, \\ \mu_3 &= \frac{2\gamma}{a^3}. \end{aligned}\right\} \tag{3.20}$$

Absolute moments

3.6 The quantity

$$\nu_r' = \int_{-\infty}^\infty |x-a|^r\, dF \tag{3.21}$$

is called the absolute moment of order r about a. The absolute moments about the

mean are written without primes. Clearly if the moment of order r exists, the absolute moment exists for any order less than or equal to r.

If r is even, the absolute moment is equal to the ordinary moment, and if the range of the distribution is positive the absolute moments about any point to the left of the start of the distribution are equal to the ordinary moments of corresponding order.

There are some interesting inequalities concerning the absolute moments. Referring to the function $A(t)$ of **2.6** and remembering that it is a non-decreasing function of t, we find, on putting $t = 1, 2, \ldots$, that

$$(\nu_1')^{1/1} \leqslant (\nu_2')^{\frac{1}{2}} \leqslant (\nu_3')^{\frac{1}{3}} \ldots \leqslant (\nu_r')^{1/r}.$$

A more general inequality, due to Liapounoff (1901), is

$$(\nu_b')^{a-c} \leqslant (\nu_c')^{a-b}\ (\nu_a')^{b-c}, \qquad a \geqslant b \geqslant c \geqslant 0. \tag{3.22}$$

A proof of this result is given in Exercise 3.15.

Factorial moments

3.7 The factorial expression

$$x(x-h)(x-2h)\ldots\{x-(r-1)h\}$$

may conveniently be written $x^{[r]}$, a notation which brings out an analogy with the power x^r. Taking first differences with respect to x and with unit h, we have

$$\begin{aligned}\Delta x^{[r]} &= (x+h)^{[r]} - x^{[r]} \\ &= (x+h)x(x-h)\ldots\{x-(r-2)h\} - x(x-h)\ldots\{x-(r-1)h\} \\ &= rx^{[r-1]}h,\end{aligned}$$

which may be compared with the equation in differentials

$$dx^r = rx^{r-1}dx.$$

Conversely

$$\sum_{x=0}^{x} x^{[r]} = \frac{1}{(r+1)h}(x+h)^{[r+1]}$$

corresponding to

$$\int_0^x x^r dx = \frac{1}{r+1}x^{r+1}.$$

The rth factorial moment about an arbitrary origin may then be defined by the equation

$$\mu'_{[r]} = \sum_{j=-\infty}^{\infty} (x_j - a)^{[r]} f(x_j) \tag{3.23}$$

where we have chosen the summation sign Σ rather than the Stieltjes integral because it is almost entirely for discontinuous distributions, or continuous distributions grouped in intervals of width h, that the factorial moments are used. In statistical theory they are not very prominent, but they provide very concise formulae for the moments of certain discontinuous distributions of the binomial type.

As usual, when it is necessary to distinguish between factorial moments about the mean and those about an arbitrary point we may write the former without the prime.

3.8 The factorial moments obey laws of transformation similar to those of equation (3.5) governing ordinary moments. In fact we have the expansion (*)

$$(a+b)^{[r]} = \sum_{j=0}^{r} \binom{r}{j} a^{[r-j]} b^{[j]}$$

and hence
$$(x-a)^{[r]} = (x-b+c)^{[r]}, \text{ where } c = b-a,$$
$$= \sum_{j=0}^{r} \binom{r}{j} (x-b)^{[r-j]} c^{[j]}$$

and hence
$$\mu'_{[r]}(a) = \sum_{j=0}^{r} \binom{r}{j} \mu'_{[r-j]}(b)\, c^{[j]} \tag{3.24}$$

which may be written symbolically

$$\mu'_{[r]}(a) = \{\mu'(b)+c\}^{[r]}.$$

3.9 By direct expansion of (3.23) it is seen that

$$\left.\begin{aligned} \mu'_{[1]} &= \mu'_1 \\ \mu'_{[2]} &= \mu'_2 - h\mu'_1 \\ \mu'_{[3]} &= \mu'_3 - 3h\mu'_2 + 2h^2\mu'_1 \\ \mu'_{[4]} &= \mu'_4 - 6h\mu'_3 + 11h^2\mu'_2 - 6h^3\mu'_1 \end{aligned}\right\} \tag{3.25}$$

and conversely that

$$\left.\begin{aligned} \mu'_2 &= \mu'_{[2]} + h\mu'_{[1]} \\ \mu'_3 &= \mu'_{[3]} + 3h\mu'_{[2]} + h^2\mu'_{[1]} \\ \mu'_4 &= \mu'_{[4]} + 6h\mu'_{[3]} + 7h^2\mu'_{[2]} + h^3\mu'_{[1]} \end{aligned}\right\} \tag{3.26}$$

Since the first moments are equal the equations remain true when the primes are dropped and terms in first moments omitted.

Frisch (1926) gives general formulae showing the factorial moments about one point in terms of the ordinary moments about another, and vice versa. In fact

$$\mu'_r(a) = \sum_{j=0}^{r} \left\{ \binom{r}{j} B_{r-j}^{(-j)}(c) h^{r-j} \mu'_{[j]}(b) \right\} \tag{3.27}$$

$$\mu'_{[r]}(a) = \sum_{j=0}^{r} \left\{ \binom{r}{j} B_{r-j}^{(r+1)}(c+h)\, h^{r-j} \mu'_j(b) \right\} \tag{3.28}$$

where $B_r^{(n)}(x)$ is the Bernoulli polynomial of order n and degree r in x (see **3.25** below).

Calculation of factorial moments

3.10 The calculation of factorial moments for grouped data may be effected by a process of progressive summation which is illustrated in Table 3.2.

(*) It is clear that $(a+b)^{[r]}$ will be a polynomial of degree r in a, and may therefore be equated to $\sum_{j=0}^{r} k_j\, a^{[r-j]}$, where the k's are polynomials in b and h but do not contain a. Putting $a = 0$ we obtain $b^{[r]} = k_r$. Taking first differences with respect to a and putting $a = 0$ we obtain $rb^{[r-1]} = k_{r-1}$. Successive differences give the k's and the above result follows.

Table 3.2

(1) Frequency	(2) First summation	(3) Second summation	(4) Third summation
f_1	$f_1+\ldots+f_n$		
f_2	$f_2+\ldots+f_n$	$f_2+2f_3+\ldots+(n-1)f_n$	
f_3	$f_3+\ldots+f_n$	$f_3+2f_4+\ldots+(n-2)f_n$	$f_3+3f_4+\ldots+\frac{(n-1)(n-2)}{2}f_n$
f_4	$f_4+f_5+\ldots+f_n$	$+\ldots+(n-3)f_n$	$f_4+3f_5+\ldots+\frac{(n-2)(n-3)}{2}f_n$
.	.	.	.
.	.	.	.
.	.	.	.
f_{n-2}	$f_{n-2}+f_{n-1}+f_n$	$f_{n-2}+2f_{n-1}+3f_n$	$f_{n-2}+3f_{n-1}+6f_n$
f_{n-1}	$f_{n-1}+f_n$	$f_{n-1}+2f_n$	$f_{n-1}+3f_n$
f_n	f_n	f_n	f_n
TOTALS	$\Sigma(jf_j)=\mu'_{[1]}$	$\Sigma\left\{\frac{j(j-1)}{2}f_j\right\}=\frac{1}{2!}\mu'_{[2]}$	$\Sigma\left\{\frac{j(j-1)(j-2)}{3!}f_j\right\}=\frac{1}{3!}\mu'_{[3]}$

Writing the proportional frequencies in the successive n intervals as $f_1 \ldots f_n$, as shown in the left-hand column, we construct column 2 by adding frequencies from the bottom. In the nth row we write f_n, in the $(n-1)$th row the sum f_n+f_{n-1}, in the $(n-2)$th row the sum $f_n+f_{n-1}+f_{n-2}$, and so on, the first row containing the sum $f_n+f_{n-1}+\ldots+f_1$.

In column 3 the process is repeated with the rows of column 2, stopping at the second row, e.g. the nth row contains f_n, the $(n-1)$th row $(f_n+f_{n-1})+f_n = 2f_n+f_{n-1}$, and so on, the second row containing the sum $(n-1)f_n+(n-2)f_{n-1}+\ldots+2f_3+f_2$.

Column 4 repeats the process with the entries of column 3, but stopping at the third row; and so on.

Consider now the sum of the entries in column 2. In that sum f_1 appears once, f_2 twice, . . , f_n n times. Hence the sum is equal to $\sum_{j=1}^{n} jf_j$

$$= \mu'_{[1]}.$$

In column 3, f_2 appears once, f_3 appears 3 times, . . , f_n appears $\frac{1}{2}n(n-1)$ times. Hence

$$\text{Sum} = \sum_{j=1}^{n}\left(\frac{j(j-1)}{2}f_j\right)$$

$$= \tfrac{1}{2}\mu'_{[2]}.$$

In general, the sum of the $(r+1)$th column will be given by

$$\text{Sum} = \frac{1}{r!}\mu'_{[r]}.$$

If the actual frequencies are used instead of the proportional frequencies the sums have to be divided by the total frequency N.

Thus the process of summation gives the factorial moments directly.

Example 3.7

Consider again the data of Table 1.7, showing the distribution of 8585 men according to height in inches. The columns on the right in Table 3.3 show the successive sums.

At the top of the last three columns there has been placed within brackets the number which would have been obtained if the summation were continued up the column one place further than is required for the sum at the foot. These bracketed figures are useful to have as a check since each must equal the sum at the foot of the preceding column.

Table 3.3—Calculation of the factorial moments of a distribution of men according to height in inches (*Table* 1.7)

Height	Frequency	First sum	Second sum	Third sum	Fourth sum
57–	2	8,585	(94,614)	—	—
58–	4	8,583	86,029	(502,459)	—
59–	14	8,579	77,446	416,430	(1,709,785)
60–	41	8,565	68,867	338,984	1,293,355
61–	83	8,524	60,302	270,117	954,371
62–	169	8,441	51,778	209,815	684,254
63–	394	8,272	43,337	158,037	474,439
64–	669	7,878	35,065	114,700	316,402
65–	990	7,209	27,187	79,635	201,702
66–	1,223	6,219	19,978	52,448	122,067
67–	1,329	4,996	13,759	32,470	69,619
68–	1,230	3,667	8,763	18,711	37,419
69–	1,063	2,437	5,096	9,948	18,438
70–	646	1,374	2,659	4,852	8,490
71–	392	728	1,285	2,193	3,638
72–	202	336	557	908	1,445
73–	79	134	221	351	537
74–	32	55	87	130	186
75–	16	23	32	43	56
76–	5	7	9	11	13
77–	2	2	2	2	2
TOTALS	8,585	94,614	502,459	1,709,785	4,186,163

From this table we find

$$\begin{aligned} \mu'_{[1]} &= 11{\cdot}020,850,320,33 \\ \mu'_{[2]} &= 117{\cdot}055,096,097,84 \\ \mu'_{[3]} &= 1,194{\cdot}957,483,983,69 \\ \mu'_{[4]} &= 11,702{\cdot}727,082,119,98. \end{aligned}$$

From these values we may derive the ordinary moments, using equations (3.26) and find

$$\begin{aligned} \mu'_1 &= 11{\cdot}020,850,320,33 \\ \mu'_2 &= 128{\cdot}075,946,418,2 \\ \mu'_3 &= 1,557{\cdot}143,622,597,5 \\ \mu'_4 &= 19,702{\cdot}878,509,027,3, \end{aligned}$$

from which we find, for the moments about the mean,

$$\begin{aligned} \mu_2 &= 6{\cdot}616,805 \\ \mu_3 &= -0{\cdot}207,840 \\ \mu_4 &= 137{\cdot}689,185, \end{aligned}$$

the units being one inch.

Factorial moment-generating functions

3.11 If the frequency-generating function is given by

$$P(t) = \sum_{j=0}^{\infty} f_j t^j$$

we have, replacing t by $1+t$,

$$\begin{aligned} P(1+t) &= \Sigma f_j(1+t)^j \\ &= \sum_{j=0}^{\infty} f_j \sum_{i=1}^{j} \binom{j}{i} t^i \\ &= \sum_{i=0}^{\infty} \frac{t^i}{i!} \sum_{j=0}^{\infty} [f_j j(j-1)\ldots\{j-i+1\}] \\ &= \sum_{i=0}^{\infty} (t^i \mu'_{[i]}/i!). \end{aligned} \tag{3.29}$$

Hence $P(1+t)$ is a factorial moment-generating function (f.m.g.f.), the factorial moment of order i being the coefficient of $t^i/i!$ The moments here are taken with a unit interval between the frequencies.

Example 3.8

Consider again the binomial distribution whose frequency-generating function is $\{pt+(1-p)\}^n$. We have

$$P(1+t) = (1+pt)^n = \sum_{j=0}^{n} p^j t^j \binom{n}{j}$$

and hence, with origin at zero,

$$\mu'_{[r]} = p^r n^{[r]}, \quad r \leqslant n,$$

the limitation on r arising from the fact that in this case $\mu'_{[r]} \equiv 0$ for $r > n$. We can now convert to the first n ordinary moments, if desired, by equations (3.26). For instance

$$\mu'_1 = \mu_{[1]} = pn,$$
$$\mu'_2 = p^2 n(n-1)+pn, \text{ etc.}$$

Cumulants

3.12 The moments are a set of descriptive constants of a distribution which are useful for measuring its properties and, in certain circumstances, for specifying it. Their use in these connections will be considered in later chapters. They are not, however, the only set of constants for the purpose, or even the best set. Another series of constants, the so-called cumulants, have properties which are more useful from the theoretical standpoint.

Formally, the cumulants $\kappa_1, \kappa_2, \ldots, \kappa_r$ are defined by the identity in t

$$\exp\left\{\kappa_1 t+\frac{\kappa_2 t^2}{2!}+\ldots+\frac{\kappa_r t^r}{r!}+\ldots\right\} = 1+\mu'_1 t+\frac{\mu'_2 t^2}{2!}+\ldots+\frac{\mu'_r t^r}{r!}+\ldots \tag{3.30}$$

It is sometimes more convenient to write the same equation with it for t, thus:

$$\begin{aligned} \exp\left\{\kappa_1(it)+\kappa_2\frac{(it)^2}{2!}+\ldots+\kappa_r\frac{(it)^r}{r!}+\ldots\right\} &= 1+\mu'_1\frac{(it)}{1!}+\ldots+\mu'_r\frac{(it)^r}{r!}+\ldots \\ &= \int_{-\infty}^{\infty} e^{itx}\,dF \\ &= \phi(t). \end{aligned} \tag{3.31}$$

Thus, whereas μ'_r is the coefficient of $(it)^r/r!$ in $\phi(t)$, the characteristic function, κ_r is the coefficient of $(it)^r/r!$ in $\log \phi(t)$, if an expansion in power series exists. Log $\phi(t)$

may then be called a cumulant-generating function and denoted by c.g.f. It is also known as the cumulative function.

3.13 If in equation (3.31) the origin is changed from a to b, where as usual $b-a=c$, the effect on $\phi(t)$ is to multiply it by e^{-itc}, for $\int e^{itx}\,dF$ becomes $\int e^{it(x-c)}e^{itc}\,dF$. Hence the effect on $\log\phi(t)$ is merely to add the term $-itc$, and consequently the coefficients in $\log\phi(t)$ are unchanged, except the first, which is decreased by c.

Hence the cumulants are invariant under change of origin, except the first. In this they stand in sharp contrast to the moments about an arbitrary point.

Both cumulants and moments have another property of an invariantive kind, namely, that if the variate-values are multiplied by a constant a, μ'_r and κ_r are multiplied by a^r. This is at once evident from their definitions. Thus any linear transformation of the kind

$$\xi = lx+m \tag{3.32}$$

leaves the cumulants unchanged so far as the constant m is concerned and multiplies κ_r by l^r. The sole exception is the first cumulant, which is equal to the mean. In particular, if we transform a distribution to standard measure, the only effect is to multiply κ_r by σ^{-r}, σ being the standard deviation and, as we shall see in a moment, being equal to $\kappa_2^{\frac{1}{2}}$.

The invariantive properties of the cumulants were the origin of their original name of semi-invariants, seminvariants or half-invariants (Thiele, 1903). In accordance with the theory of algebraic invariants, however, it seems best to reserve the word "seminvariant" for any constant λ_r which, under the transformation (3.32), is multiplied by l^r. The cumulants and the moments *about the mean* are thus particular cases of seminvariants.

Relations between moments and cumulants

3.14 Subject to conditions of existence, we have, from (3.30),

$$\begin{aligned}1+\mu'_1\frac{t}{1!}+\ldots+\mu'_r\frac{t^r}{r!}+\ldots &= \exp\left\{\kappa_1\frac{t}{1!}+\kappa_2\frac{t^2}{2!}+\ldots\kappa_r\frac{t^r}{r!}\cdots\right\}\\ &= \exp\left(\frac{\kappa_1 t}{1!}\right)\exp\left(\frac{\kappa_2 t^2}{2!}\right)\ldots\exp\left(\frac{\kappa_r t^r}{r!}\right)\ldots\\ &= \left\{1+\frac{\kappa_1 t}{1!}+\frac{\kappa_1^2 t^2}{2!}+\ldots\right\}\left\{1+\frac{\kappa_2 t^2}{2!}+\frac{1}{2!}\left(\frac{\kappa_2^2 t^4}{2!^2}\right)+\ldots\right\}\\ &\quad\ldots\left\{1+\left(\frac{\kappa_r t^r}{r!}\right)+\frac{1}{2!}\left(\frac{\kappa_r t^r}{r!}\right)^2+\ldots\right\}\ldots\end{aligned}$$

Picking out the terms in the exponential expansions which, when multiplied together, give a power of t^r, we have

$$\mu'_r = \sum_{m=0}^{r}\Sigma\left(\frac{\kappa_{p_1}}{p_1!}\right)^{\pi_1}\left(\frac{\kappa_{p_2}}{p_2!}\right)^{\pi_2}\cdots\left(\frac{\kappa_{p_m}}{p_m!}\right)^{\pi_m}\frac{r!}{\pi_1!\pi_2!\ldots\pi_m!}, \tag{3.33}$$

where the second summation extends over all non-negative values of the π's such that

$$p_1\pi_1+p_2\pi_2+\ldots p_m\pi_m = r. \tag{3.34}$$

It is worth noting that the rather tedious process of writing down the explicit relations for particular values of r may be shortened considerably. In fact, differentiating (3.30) with respect to κ_j we have

$$\frac{t^j}{j!}\left(1+\mu_1' t+\ldots+\frac{\mu_r' t^r}{r!}+\ldots\right)=\frac{\partial \mu_1'}{\partial \kappa_j}t+\ldots+\frac{t^r}{r!}\frac{\partial \mu_r'}{\partial \kappa_j}+\ldots$$

and hence, identifying powers of t,

$$\frac{\partial \mu_r'}{\partial \kappa_j}=\binom{r}{j}\mu_{r-j}'. \tag{3.35}$$

In particular

$$\frac{\partial \mu_r'}{\partial \kappa_1}=r\mu_{r-1}' \tag{3.36}$$

and thus, given any μ_r' in terms of the κ's, we can write down successively those of lower orders by a differentiation.(*)

The first ten of these expressions are, for moments about an arbitrary point :—

$$\left.\begin{aligned}
\mu_1' &= \kappa_1,\\
\mu_2' &= \kappa_2+\kappa_1^2,\\
\mu_3' &= \kappa_3+3\kappa_2\kappa_1+\kappa_1^3,\\
\mu_4' &= \kappa_4+4\kappa_3\kappa_1+3\kappa_2^2+6\kappa_2\kappa_1^2+\kappa_1^4,\\
\mu_5' &= \kappa_5+5\kappa_4\kappa_1+10\kappa_3\kappa_2+10\kappa_3\kappa_1^2+15\kappa_2^2\kappa_1+10\kappa_2\kappa_1^3+\kappa_1^5,\\
\mu_6' &= \kappa_6+6\kappa_5\kappa_1+15\kappa_4\kappa_2+15\kappa_4\kappa_1^2+10\kappa_3^2+60\kappa_3\kappa_2\kappa_1+20\kappa_3\kappa_1^3\\
&\quad +15\kappa_2^3+45\kappa_2^2\kappa_1^2+15\kappa_2\kappa_1^4+\kappa_1^6,\\
\mu_7' &= \kappa_7+7\kappa_6\kappa_1+21\kappa_5\kappa_2+21\kappa_5\kappa_1^2+35\kappa_4\kappa_3+105\kappa_4\kappa_2\kappa_1\\
&\quad +35\kappa_4\kappa_1^3+70\kappa_3^2\kappa_1+105\kappa_3\kappa_2^2+210\kappa_3\kappa_2\kappa_1^2+35\kappa_3\kappa_1^4\\
&\quad +105\kappa_2^3\kappa_1+105\kappa_2^2\kappa_1^3+21\kappa_2\kappa_1^5+\kappa_1^7,\\
\mu_8' &= \kappa_8+8\kappa_7\kappa_1+28\kappa_6\kappa_2+28\kappa_6\kappa_1^2+56\kappa_5\kappa_3+168\kappa_5\kappa_2\kappa_1+56\kappa_5\kappa_1^3\\
&\quad +35\kappa_4^2+280\kappa_4\kappa_3\kappa_1+210\kappa_4\kappa_2^2+420\kappa_4\kappa_2\kappa_1^2+70\kappa_4\kappa_1^4\\
&\quad +280\kappa_3^2\kappa_2+280\kappa_3^2\kappa_1^2+840\kappa_3\kappa_2^2\kappa_1+560\kappa_3\kappa_2\kappa_1^3+56\kappa_3\kappa_1^5\\
&\quad +105\kappa_2^4+420\kappa_2^3\kappa_1^2+210\kappa_2^2\kappa_1^4+28\kappa_2\kappa_1^6+\kappa_1^8,\\
\mu_9' &= \kappa_9+9\kappa_8\kappa_1+36\kappa_7\kappa_2+36\kappa_7\kappa_1^2+84\kappa_6\kappa_3+252\kappa_6\kappa_2\kappa_1\\
&\quad +84\kappa_6\kappa_1^3+126\kappa_5\kappa_4+504\kappa_5\kappa_3\kappa_1+378\kappa_5\kappa_2^2+756\kappa_5\kappa_2\kappa_1^2\\
&\quad +126\kappa_5\kappa_1^4+315\kappa_4^2\kappa_1+1260\kappa_4\kappa_3\kappa_2+1260\kappa_4\kappa_3\kappa_1^2+1890\kappa_4\kappa_2^2\kappa_1\\
&\quad +1260\kappa_4\kappa_2\kappa_1^3+126\kappa_4\kappa_1^5+280\kappa_3^3+2520\kappa_3^2\kappa_2\kappa_1+840\kappa_3^2\kappa_1^3\\
&\quad +1260\kappa_3\kappa_2^3+3780\kappa_3\kappa_2^2\kappa_1^2+1260\kappa_3\kappa_2\kappa_1^4+84\kappa_3\kappa_1^6+945\kappa_2^4\kappa_1\\
&\quad +1260\kappa_2^3\kappa_1^3+378\kappa_2^2\kappa_1^5+36\kappa_2\kappa_1^7+\kappa_1^9,\\
\mu_{10}' &= \kappa_{10}+10\kappa_9\kappa_1+45\kappa_8\kappa_2+45\kappa_8\kappa_1^2+120\kappa_7\kappa_3+360\kappa_7\kappa_2\kappa_1\\
&\quad +120\kappa_7\kappa_1^3+210\kappa_6\kappa_4+840\kappa_6\kappa_3\kappa_1+630\kappa_6\kappa_2^2+1260\kappa_6\kappa_2\kappa_1^2\\
&\quad +210\kappa_6\kappa_1^4+126\kappa_5^2+1260\kappa_5\kappa_4\kappa_1+2520\kappa_5\kappa_3\kappa_2+2520\kappa_5\kappa_3\kappa_1^2\\
&\quad +3780\kappa_5\kappa_2^2\kappa_1+2520\kappa_5\kappa_2\kappa_1^3+252\kappa_5\kappa_1^5+1575\kappa_4^2\kappa_2+1575\kappa_4^2\kappa_1^2\\
&\quad +2100\kappa_4\kappa_3^2+12600\kappa_4\kappa_3\kappa_2\kappa_1+4200\kappa_4\kappa_3\kappa_1^3+3150\kappa_4\kappa_2^3\\
&\quad +9450\kappa_4\kappa_2^2\kappa_1^2+3150\kappa_4\kappa_2\kappa_1^4+210\kappa_4\kappa_1^6+2800\kappa_3^3\kappa_1\\
&\quad +6300\kappa_3^2\kappa_2^2+12600\kappa_3^2\kappa_2\kappa_1^2+2100\kappa_3^2\kappa_1^4+12600\kappa_3\kappa_2^3\kappa_1\\
&\quad +12600\kappa_3\kappa_2^2\kappa_1^3+2520\kappa_3\kappa_2\kappa_1^5+120\kappa_3\kappa_1^7+945\kappa_2^5+4725\kappa_2^4\kappa_1^2\\
&\quad +3150\kappa_2^3\kappa_1^4+630\kappa_2^2\kappa_1^6+45\kappa_2\kappa_1^8+\kappa_1^{10};
\end{aligned}\right\} \tag{3.37}$$

(*) The coefficients of μ's in terms of κ's are those of the unitary symmetric functions (1^r) in terms of the augmented symmetric functions and can be read from the tables of David and Kendall (1949) as far as order 12. The inverse relations may also be obtained from the tables on multiplication by appropriate factorials.

or, for moments about the mean ($\kappa_1 = 0$),

$$\left.\begin{aligned}
\mu_2 &= \kappa_2,\\
\mu_3 &= \kappa_3,\\
\mu_4 &= \kappa_4 + 3\kappa_2^2,\\
\mu_5 &= \kappa_5 + 10\kappa_3\kappa_2,\\
\mu_6 &= \kappa_6 + 15\kappa_4\kappa_2 + 10\kappa_3^2 + 15\kappa_2^3,\\
\mu_7 &= \kappa_7 + 21\kappa_5\kappa_2 + 35\kappa_4\kappa_3 + 105\kappa_3\kappa_2^2,\\
\mu_8 &= \kappa_8 + 28\kappa_6\kappa_2 + 56\kappa_5\kappa_3 + 35\kappa_4^2 + 210\kappa_4\kappa_2^2 + 280\kappa_3^2\kappa_2 + 105\kappa_2^4,\\
\mu_9 &= \kappa_9 + 36\kappa_7\kappa_2 + 84\kappa_6\kappa_3 + 126\kappa_5\kappa_4 + 378\kappa_5\kappa_2^2 + 1260\kappa_4\kappa_3\kappa_2 + 280\kappa_3^3\\
&\quad + 1260\kappa_3\kappa_2^3,\\
\mu_{10} &= \kappa_{10} + 45\kappa_8\kappa_2 + 120\kappa_7\kappa_3 + 210\kappa_6\kappa_4 + 630\kappa_6\kappa_2^2 + 126\kappa_5^2\\
&\quad + 2520\kappa_5\kappa_3\kappa_2 + 1575\kappa_4^2\kappa_2 + 2100\kappa_4\kappa_3^2 + 3150\kappa_4\kappa_2^3\\
&\quad + 6300\kappa_3^2\kappa_2^2 + 945\kappa_2^5.
\end{aligned}\right\} \tag{3.38}$$

Conversely we have

$$\frac{\kappa_1 t}{1!} + \ldots + \frac{\kappa_r t^r}{r!} + \ldots = \log\left(1 + \frac{\mu_1' t}{1!} + \ldots + \frac{\mu_r' t^r}{r!} + \ldots\right).$$

Expanding the logarithm and picking out powers of t^r as before, we have

$$\kappa_r = r! \sum_{m=0}^{r} \Sigma \left(\frac{\mu_{p_1}'}{p_1!}\right)^{\pi_1} \ldots \left(\frac{\mu_{p_m}'}{p_m!}\right)^{\pi_m} \frac{(-1)^{\rho-1}(\rho-1)!}{\pi_1! \ldots \pi_m!}, \tag{3.39}$$

the second summation extending over all non-negative π's and ρ's, subject to (3.34) and the further condition

$$\pi_1 + \pi_2 + \ldots + \pi_m = \rho. \tag{3.40}$$

The first ten formulae are, in terms of moments about an arbitrary point :—

$$\left.\begin{aligned}
\kappa_1 &= \mu_1',\\
\kappa_2 &= \mu_2' - \mu_1'^2,\\
\kappa_3 &= \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3,\\
\kappa_4 &= \mu_4' - 4\mu_3'\mu_1' - 3\mu_2'^2 + 12\mu_2'\mu_1'^2 - 6\mu_1'^4,\\
\kappa_5 &= \mu_5' - 5\mu_4'\mu_1' - 10\mu_3'\mu_2' + 20\mu_3'\mu_1'^2 + 30\mu_2'^2\mu_1' - 60\mu_2'\mu_1'^3 + 24\mu_1'^5,\\
\kappa_6 &= \mu_6' - 6\mu_5'\mu_1' - 15\mu_4'\mu_2' + 30\mu_4'\mu_1'^2 - 10\mu_3'^2 + 120\mu_3'\mu_2'\mu_1' - 120\mu_3'\mu_1'^3\\
&\quad + 30\mu_2'^3 - 270\mu_2'^2\mu_1'^2 + 360\mu_2'\mu_1'^4 - 120\mu_1'^6,\\
\kappa_7 &= \mu_7' - 7\mu_6'\mu_1' - 21\mu_5'\mu_2' + 42\mu_5'\mu_1'^2 - 35\mu_4'\mu_3' + 210\mu_4'\mu_2'\mu_1'\\
&\quad - 210\mu_4'\mu_1'^3 + 140\mu_3'^2\mu_1' + 210\mu_3'\mu_2'^2 - 1260\mu_3'\mu_2'\mu_1'^2 + 840\mu_3'\mu_1'^4\\
&\quad - 630\mu_2'^3\mu_1' + 2520\mu_2'^2\mu_1'^3 - 2520\mu_2'\mu_1'^5 + 720\mu_1'^7,\\
\kappa_8 &= \mu_8' - 8\mu_7'\mu_1' - 28\mu_6'\mu_2' + 56\mu_6'\mu_1'^2 - 56\mu_5'\mu_3' + 336\mu_5'\mu_2'\mu_1'\\
&\quad - 336\mu_5'\mu_1'^3 - 35\mu_4'^2 + 560\mu_4'\mu_3'\mu_1' + 420\mu_4'\mu_2'^2 - 2520\mu_4'\mu_2'\mu_1'^2\\
&\quad + 1680\mu_4'\mu_1'^4 + 560\mu_3'^2\mu_2' - 1680\mu_3'^2\mu_1'^2 - 5040\mu_3'\mu_2'^2\mu_1'\\
&\quad + 13440\mu_3'\mu_2'\mu_1'^3 - 6720\mu_3'\mu_1'^5 - 630\mu_2'^4 + 10080\mu_2'^3\mu_1'^2\\
&\quad - 25200\mu_2'^2\mu_1'^4 + 20160\mu_2'\mu_1'^6 - 5040\mu_1'^8,\\
\kappa_9 &= \mu_9' - 9\mu_8'\mu_1' - 36\mu_7'\mu_2' + 72\mu_7'\mu_1'^2 - 84\mu_6'\mu_3' + 504\mu_6'\mu_2'\mu_1'\\
&\quad - 504\mu_6'\mu_1'^3 - 126\mu_5'\mu_4' + 1008\mu_5'\mu_3'\mu_1' + 756\mu_5'\mu_2'^2 - 4536\mu_5'\mu_2'\mu_1'^2\\
&\quad + 3024\mu_5'\mu_1'^4 + 630\mu_4'^2\mu_1' + 2520\mu_4'\mu_3'\mu_2' - 7560\mu_4'\mu_3'\mu_1'^2\\
&\quad - 11340\mu_4'\mu_2'^2\mu_1' + 30240\mu_4'\mu_2'\mu_1'^3 - 15120\mu_4'\mu_1'^5 + 560\mu_3'^3
\end{aligned}\right\} \tag{3.41}$$

$$
\begin{aligned}
&-15120\mu_3'^2\mu_2'\mu_1'+20160\mu_3'^2\mu_1'^3-7560\mu_3'\mu_2'^3+90720\mu_3'\mu_2'^2\mu_1'^2\\
&-151200\mu_3'\mu_2'\mu_1'^4+60480\mu_3'\mu_1'^6+22680\mu_2'^4\mu_1'-151200\mu_2'^3\mu_1'^3\\
&+272160\mu_2'^2\mu_1'^5-181440\mu_2'\mu_1'^7+40320\mu_1'^9,\\
\kappa_{10} = \mu_{10}' &-10\mu_9'\mu_1'-45\mu_8'\mu_2'+90\mu_8'\mu_1'^2-120\mu_7'\mu_3'+720\mu_7'\mu_2'\mu_1'\\
&-720\mu_7'\mu_1'^3-210\mu_6'\mu_4'+1680\mu_6'\mu_3'\mu_1'+1260\mu_6'\mu_2'^2\\
&-7560\mu_6'\mu_2'\mu_1'^2+5040\mu_6'\mu_1'^4-126\mu_5'^2+2520\mu_5'\mu_4'\mu_1'\\
&+5040\mu_5'\mu_3'\mu_2'-15120\mu_5'\mu_3'\mu_1'^2-22680\mu_5'\mu_2'^2\mu_1'+60480\mu_5'\mu_2'\mu_1'^3\\
&-30240\mu_5'\mu_1'^5+3150\mu_4'^2\mu_2'-9450\mu_4'^2\mu_1'^2+4200\mu_4'\mu_3'^2\\
&-75600\mu_4'\mu_3'\mu_2'\mu_1'+100800\mu_4'\mu_3'\mu_1'^3-18900\mu_4'\mu_2'^3\\
&+226800\mu_4'\mu_2'^2\mu_1'^2-378000\mu_4'\mu_2'\mu_1'^4+151200\mu_4'\mu_1'^6-16800\mu_3'^3\mu_1'\\
&-37800\mu_3'^2\mu_2'^2+302400\mu_3'^2\mu_2'\mu_1'^2-252000\mu_3'^2\mu_1'^4+302400\mu_3'\mu_2'^3\mu_1'\\
&-1512000\mu_3'\mu_2'^2\mu_1'^3+1814400\mu_3'\mu_2'\mu_1'^5-604800\mu_3'\mu_1'^7\\
&+22680\mu_2'^5-567000\mu_2'^4\mu_1'^2+2268000\mu_2'^3\mu_1'^4-3175200\mu_2'^2\mu_1'^6\\
&+1814400\mu_2'\mu_1'^8-362880\mu_1'^{10};
\end{aligned}
\qquad (3.42)
$$

or, for moments about the mean,

$$
\begin{aligned}
\kappa_2 &= \mu_2,\\
\kappa_3 &= \mu_3,\\
\kappa_4 &= \mu_4-3\mu_2^2,\\
\kappa_5 &= \mu_5-10\mu_3\mu_2,\\
\kappa_6 &= \mu_6-15\mu_4\mu_2-10\mu_3^2+30\mu_2^3,\\
\kappa_7 &= \mu_7-21\mu_5\mu_2-35\mu_4\mu_3+210\mu_3\mu_2^2,\\
\kappa_8 &= \mu_8-28\mu_6\mu_2-56\mu_5\mu_3-35\mu_4^2+420\mu_4\mu_2^2+560\mu_3^2\mu_2-630\mu_2^4,\\
\kappa_9 &= \mu_9-36\mu_7\mu_2-84\mu_6\mu_3-126\mu_5\mu_4+756\mu_5\mu_2^2+2520\mu_4\mu_3\mu_2\\
&\quad+560\mu_3^3-7560\mu_3\mu_2^2,\\
\kappa_{10} &= \mu_{10}-45\mu_8\mu_2-120\mu_7\mu_3-210\mu_6\mu_4+1260\mu_6\mu_2^2-126\mu_5^2\\
&\quad+5040\mu_5\mu_3\mu_2+3150\mu_4^2\mu_2+4200\mu_4\mu_3^2-18900\mu_4\mu_2^3\\
&\quad-37800\mu_3^2\mu_2^2+22680\mu_2^5.
\end{aligned}
\qquad (3.43)
$$

Existence of cumulants

3.15 The formal expression (3.30) may be regarded as defining the cumulants in terms of the moments, and it is thus evident that the cumulant of order r exists if the moments of orders r and lower exist. If, however, we look to the equation

$$\exp\left(\Sigma\,\kappa_r\frac{(it)^r}{r!}\right)=\phi(t)$$

as defining the cumulants, it is not quite so easy to show that κ_r exists if μ_r and lower μ's exist.

Since e^{itx} is an analytic function we can expand it in the expression for the characteristic function and obtain

$$
\begin{aligned}
\phi(t) &= \int_{-\infty}^{\infty}\left\{\sum_0^r\frac{(itx)^j}{j!}+O\{|tx|^{r+1}/(r+1)!\}\right\}dF\\
&= \sum_{j=0}^{r}\mu_j'(it)^j/j!+R_r
\end{aligned}
\qquad (3.44)
$$

where R_r is a remainder term not greater than $\rho\nu_{r+1}|t|^{r+1}/(r+1)!$ with $|\rho|\leqslant 1$.

We may write this as

$$\phi(t) = \sum_{j=0}^{r} \mu_j'(it)^j/j! + o(t^r).$$

For some small t we may then take logarithms and expand to obtain

$$\log\phi(t) = \sum_{j=1}^{r} \kappa_j(it)^j/j! + o(t^r), \tag{3.45}$$

the coefficients κ being the cumulants by definition. Thus if ν_{r+1} exists κ_r and lower cumulants exist. It is possible to prove stronger results, but this is sufficient for many purposes.

Calculation of cumulants

3.16 The cumulants are not, like the moments, directly ascertainable by summatory or integrative processes, and to find them it is necessary either to find the moments and employ equations (3.43) or to derive them from the characteristic function. The following examples will illustrate the processes involved.

Example 3.9

In Example 3.7 we found the following values for the moments about the mean of the height data of Table 1.7 :—

$$\begin{aligned} \mu_1' &= \ \ 11{\cdot}020{,}850 \\ \mu_2 &= \ \ \ 6{\cdot}616{,}805 \\ \mu_3 &= -0{\cdot}207{,}840 \\ \mu_4 &= 137{\cdot}689{,}185, \end{aligned}$$

whence, from (3.43), κ_2 and κ_3 have the same values as μ_2 and μ_3 and

$$\begin{aligned} \kappa_4 &= \mu_4 - 3\mu_2^2 \\ &= 6{\cdot}342{,}86. \end{aligned}$$

κ_1 is the same as μ_1', in this case measured from the centre of the interval 56– inches.

The same results would, of course, have been obtained if we had used equations (3.42) and moments about the origin.

Example 3.10

Consider the discontinuous distribution whose frequencies at the values 0, 1, . . . , j, . . . are $e^{-\lambda}\left(1, \frac{\lambda}{1!}, \ldots, \frac{\lambda^j}{j!}, \ldots\right)$. The characteristic function is given by

$$\begin{aligned} \phi(t) &= e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} e^{ijt} \\ &= e^{-\lambda}\exp(\lambda e^{it}) \\ &= \exp\{\lambda(e^{it}-1)\}. \end{aligned}$$

Since the variate is non-negative, for any r the absolute moment is the same as the ordinary moment, and we have

$$\mu_r' = e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j j^r}{j!};$$

and since this converges (*) cumulants of all orders exist. They are therefore given by the expansion of $\log \phi(t)$ as a power series in t. But

$$\log \phi(t) = \lambda(e^{it}-1)$$
$$= \lambda \sum_{j=1}^{\infty} (it)^j/j!$$

and hence $$\kappa_r = \lambda$$

for all r. Thus all cumulants of the distribution are equal to λ.

Example 3.11

In Example 3.4 we found, in effect, for the characteristic function of the distribution

$$dF = \frac{1}{\sigma\sqrt{(2\pi)}} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx, \qquad -\infty \leqslant x \leqslant \infty,$$
$$\phi(t) = \exp(-\tfrac{1}{2}\sigma^2 t^2).$$

Thus $$\log \phi(t) = -\tfrac{1}{2}\sigma^2 t^2.$$

It is easily seen that the absolute moments and hence the cumulants of all orders exist. Thus

$$\kappa_2 = \sigma^2$$

and $$\kappa_r = 0, \qquad r > 2.$$

Example 3.12

In Example 3.6 it was found that for the distribution

$$dF = \frac{a^\gamma}{\Gamma(\gamma)} x^{\gamma-1} e^{-ax}\, dx, \qquad 0 \leqslant x \leqslant \infty\ ;\ a,\ \gamma > 0,$$

the characteristic function is given by

$$\phi(t) = \frac{1}{\left(1-\frac{it}{a}\right)^\gamma}.$$

It is readily verified that cumulants of all orders exist and hence

$$\kappa_r = \text{coeff. of } \frac{(it)^r}{r!} \text{ in } -\gamma \log\left(1-\frac{it}{a}\right)$$
$$= \gamma(r-1)!\, a^{-r}.$$

Example 3.13

Consider again the distribution of Example 3.3,

$$dF = \frac{k}{(1+x^2)^m} dx, \qquad -\infty \leqslant x \leqslant \infty\ ;\ m \geqslant 1.$$

(*) For the ratio of the $(n+1)$th term of the series to the nth is

$$\frac{\lambda^{n+1}(n+1)^r}{(n+1)!} \Big/ \frac{\lambda^n n^r}{n!} = \frac{\lambda}{n+1}\left(1+\frac{1}{n}\right)^r = O\left(\frac{\lambda}{n}\right),$$

and thus the series converges for all finite values of λ.

The characteristic function is given by

$$\phi(t) = k\int_{-\infty}^{\infty} \frac{e^{ixt}}{(1+x^2)^m}\,dx,$$

which, since $\sin xt$ is an odd function, reduces to

$$k\int_{-\infty}^{\infty} \frac{\cos xt}{(1+x^2)^m}\,dx.$$

This integral may be evaluated by complex integration round a contour consisting of the x-axis, the infinite semicircle above the x-axis and the infinitely small circle round the point $x = i$. It is found that

$$\phi(t) = \frac{k\pi}{2^{2m-2}(m-1)!}\, e^{-|t|}\Big\{(2|t|)^{m-1} + m(m-1)(2|t|)^{m-2} + \frac{(m+1)(m)(m-1)(m-2)}{2!}(2|t|)^{m-3} + \ldots + \frac{(2m-2)!}{(m-1)!}\Big\}.$$

If $r < 2m-1$ the absolute moment of order r

$$\nu_r = k\int_{-\infty}^{\infty} \frac{|x|^r\,dx}{(1+x^2)^m}$$

exists and hence so does the cumulant of order r. But in this case we cannot expand $\log\phi(t)$ in an *infinite* series of powers of t, though this might perhaps be thought possible from the form of $\phi(t)$. In fact, we can only expand $\log\phi(t)$ in powers of t up to the point at which the derivatives of $\phi(t)$ exist, for $t = 0$.

To simplify the discussion, consider the case when $m = 2$. We have then, since $k = 2/\pi$ in this case,

$$\phi(t) = e^{-|t|}\{|t|+1\}$$
$$\log\phi(t) = -|t| + \log\{1+|t|\}.$$

If t is positive this equals

$$-\tfrac{1}{2}t^2 + \tfrac{1}{3}t^3 - \ldots$$

but if t is negative it equals

$$-\tfrac{1}{2}t^2 - \tfrac{1}{3}t^3 - \ldots$$

the two expressions differing in the sign of the term in t^3 and every second term thereafter. There is thus no unique expansion of $\log\phi(t)$ in powers of t about the point $t = 0$. There are two forms of the function expressing $\log\phi(t)$ according as t is positive or negative.

However, these expressions coincide as far as their terms in t and t^2, and the first and second derivatives of $\log\phi(t)$ are uniquely defined when $t = 0$. Thus the first and second cumulants exist and are given by

$$\kappa_1 = 0$$
$$\kappa_2 = 1.$$

Cumulants of higher orders do not exist.

Factorial cumulants

3.17 Analogously to the relation between cumulants and moments we may define a factorial cumulant $\kappa_{[r]}$ as the coefficient of $t^r/r!$ in the expansion of the logarithm of the factorial moment-generating function. Thus if $P(t)$ is the frequency-generating function the factorial cumulant-generating function (*f.c.g.f.*) is

$$\omega(t) = \log P(1+t). \tag{3.46}$$

Like the factorial moment-generating function, $\omega(t)$ is mainly of use for certain classes of discontinuous distribution.

The relations between cumulants and factorial cumulants are formally similar to those between moments and factorial moments. Taking a unit interval between frequencies we have, analogously to (3.25) and (3.26),

$$\left.\begin{aligned} \kappa_{[1]} &= \kappa_1 \\ \kappa_{[2]} &= \kappa_2 - \kappa_1 \\ \kappa_{[3]} &= \kappa_3 - 3\kappa_2 + 2\kappa_1 \\ \kappa_{[4]} &= \kappa_4 - 6\kappa_3 + 11\kappa_2 - 6\kappa_1 \end{aligned}\right\} \tag{3.47}$$

Conversely

$$\left.\begin{aligned} \kappa_2 &= \kappa_{[2]} + \kappa_{[1]} \\ \kappa_3 &= \kappa_{[3]} + 3\kappa_{[2]} + \kappa_{[1]} \\ \kappa_4 &= \kappa_{[4]} + 6\kappa_{[3]} + 7\kappa_{[2]} + \kappa_{[1]} \end{aligned}\right\} \tag{3.48}$$

Example 3.14

Reverting once more to the binomial distribution we have

$$\omega(t) = n\log(1+pt)$$

and hence

$$\kappa_{[r]} = (-1)^{r+1} n(r-1)!\, p^r$$

Corrections for grouping

3.18 When moments are calculated from a numerically specified distribution which is grouped, there is present a certain amount of approximation owing to the fact that the frequencies are assumed to be concentrated at the mid-points of intervals. It is possible to correct for this effect under certain conditions.

Let $f(x)$ be continuous and of finite range a to b; and let the range be divided into n intervals of width h. The frequency in the jth interval, centred at $x_j = a+(j-\frac{1}{2})h$, is then given by

$$f_j = \int_{-\frac{1}{2}h}^{\frac{1}{2}h} f(x_j+\xi)\,d\xi. \tag{3.49}$$

The moments calculated from the grouped frequencies, which we will denote by a bar, are then given by

$$\bar{\mu}'_r = \sum_{j=1}^{n} x_j^r \int_{-\frac{1}{2}h}^{\frac{1}{2}h} f(x_j+\xi)\,d\xi. \tag{3.50}$$

Now from one version of the Euler-Maclaurin sum formula (*) we have, for a function $y(x)$ with $2m$ derivatives,

$$\frac{1}{h}\int_a^b y(x)\,dx = [y(a+\tfrac{1}{2}h)+y(a+\tfrac{3}{2}h)+\ldots+y\{a+(n-\tfrac{1}{2})h\}] - \sum_{j=1}^{m} \frac{h^{2j-1}}{(2j)!} B_{2j}(\tfrac{1}{2}) \left[y^{(2j-1)}(x)\right]_a^b - S_{2m} \tag{3.51}$$

(*) See for example, Milne-Thomson, *The Calculus of Finite Differences*, section 7.5.

where $y^{(r)}$ is the rth derivative of y, $B_{2m}(\frac{1}{2})$ is the value of the $2m$'th Bernoulli polynomial for argument $\frac{1}{2}$ and S_{2m} is a remainder term which can be expressed as

$$S_{2m} = \frac{nh^{2m}}{(2m)!} B_{2m}(\tfrac{1}{2}) y^{(2m)}(a+nh\theta), \qquad 0<\theta<1. \tag{3.52}$$

Putting

$$y(x) = x^r \int_{-\frac{1}{2}h}^{\frac{1}{2}h} f(x+\xi)\, d\xi$$

in (3.51) and neglecting S_{2m} we have, if the first $2m-3$ derivatives of y vanish at the terminals,

$$\frac{1}{h}\int_a^b x^r \int_{-\frac{1}{2}h}^{\frac{1}{2}h} f(x+\xi)\, d\xi\, dx = \bar{\mu}'_r. \tag{3.53}$$

Hence

$$\begin{aligned}
\bar{\mu}'_r &= \frac{1}{h}\int_a^b \int_{-\frac{1}{2}h}^{\frac{1}{2}h} (u-\xi)^r f(u)\, du\, d\xi \\
&= \frac{1}{h}\int_a^b \frac{(u+\frac{1}{2}h)^{r+1} - (u-\frac{1}{2}h)^{r+1}}{r+1} f(u)\, du \\
&= \sum_{j=0}^{[\frac{1}{2}r]} \binom{r}{2j} (\tfrac{1}{2}h)^{2j} \frac{1}{2j+1} \mu'_{r-2j}
\end{aligned} \tag{3.54}$$

where $[\frac{1}{2}r]$ is the integral part of $\frac{1}{2}r$. This gives the raw moments in terms of the actual moments. In practice we require the latter in terms of the former, and it is easy to find from (3.54) the following expressions:

$$\left.\begin{aligned}
\mu'_1 &= \bar{\mu}'_1 \\
\mu'_2 &= \bar{\mu}'_2 - \frac{1}{12}h^2 \\
\mu'_3 &= \bar{\mu}'_3 - \frac{1}{4}\bar{\mu}'_1 h^2 \\
\mu'_4 &= \bar{\mu}'_4 - \frac{1}{2}\bar{\mu}'_2 h^2 + \frac{7}{240}h^4 \\
\mu'_5 &= \bar{\mu}'_5 - \frac{5}{6}\bar{\mu}'_3 h^2 + \frac{7}{48}\bar{\mu}'_1 h^4 \\
\mu'_6 &= \bar{\mu}'_6 - \frac{5}{4}\bar{\mu}'_4 h^2 + \frac{7}{16}\bar{\mu}'_2 h^4 - \frac{31}{1344}h_6
\end{aligned}\right\} \tag{3.55}$$

The general expression for these formulae is

$$\mu'_r = \sum_{j=0}^{r} \left\{ \binom{r}{j} (2^{1-j}-1) B_j h^j \bar{\mu}'_{r-j} \right\} \tag{3.56}$$

where B_j is the Bernoulli number of order j. (Cf. Wold, 1934*a*)

3.19 We have made various assumptions in arriving at these formulae: that there is high-order contact at the terminals of the range, that the range itself is finite and that the remainder term may be neglected. In absolute magnitude $B_{2m}(\frac{1}{2})/(2m)!$

is less than $4/(2\pi)^{2m}$ and thus $|S_{2m}|$ in (3.52) is less than $4nh^{2m}/(2\pi)^{2m}$ multiplied by some value of $y^{(2m)}(x)$ in the range a to b. The extent to which the remainder term is negligible then depends on the behaviour of $f^{(2m-1)}(x)$ in the range a to b.

In practice the corrections result in improvements if there is reasonably high-order contact, but may break down if there is not. For distributions of infinite range the frequency tails off near the ends and we may still apply the corrections with some confidence to those distributions (in practice, of course, necessarily of finite range) where the major part of the distribution is concentrated and the frequencies at the extremes are small.

Example 3.15

Consider the distribution

$$dF = \frac{1}{B(12,6)} x^{11}(1-x)^5\,dx, \qquad 0 \leqslant x \leqslant 1,$$

a case of the so-called Type I distribution. The exact frequencies for intervals of 0·1 may be obtained from the Tables of the Incomplete B-Function, and are as follows :—

Centre of interval	Frequency
0·05	0·000,000,0
0·15	0,000,009,2
0·25	0·000,646,8
0·35	0·009,938,2
0·45	0·061,137,4
0·55	0·192,199,6
0·65	0·332,887,7
0·75	0·297,479,9
0·85	0·101,033,7
0·95	0·004,667,5
TOTAL	1·000,000,0

The raw moments about $x = 0$ are shown in the following table :—

Moment	Raw	Exact	Corrected
μ_1'	0·666,662,8	0·666,666,7	0·666,662,8
μ_2'	0·456,965,5	0·456,140,4	0·456,132,2
μ_3'	0·320,952,3	0·319,298,2	0·319,285,7
μ_4'	0·230,335,1	0·228,070,2	0·228,053,2
μ_5'	0·168,512,9	0·165,869,2	0·165,848,0
μ_6'	0·125,433,2	0·122,599,0	0·122,574,0

The exact values are calculable by direct integration. It is to be noted that the corrected second and third moments are more accurate than the order of the terms, $h^2/12$ and $h^2/4$, used in making the corrections. The corrected fourth moment is in error by a term of order 2×10^{-5}, which is of the same magnitude as one of the correcting terms, $7h^4/240$, used in arriving at it. Similarly the errors in the corrected fifth and sixth moments are of the same order as or higher order than some of the correcting terms.

Example 3.16

As an illustration of the way in which Sheppard's corrections break down when the condition for high-order contact is violated, an example is taken from a paper by Pairman and Pearson (1919). The following table shows the frequencies in a certain range of the normal distribution

$$dF = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2} dx,$$

with intervals of width 0·5.

Interval centred at	Frequency
1·5	0·655,91
2·0	0·278,34
2·5	0·092,45
3·0	0·024,02
3·5	0·004,89
4·0	0·000,78
4·5	0·000,10
5·0	0·000,01
TOTAL	1·056,50

The distribution has high-order contact at one end but not at the start of the curve, being in fact J-shaped and very abrupt at that point.

The following table shows the raw moments about the mean up to the fourth order, the moments with Sheppard's corrections, and the true moments calculated from the continuous normal distribution :—

Moment	Raw	Exact	Corrected
μ_2	0·158,524	0·172,222	0·137,691
μ_3	0·104,226	0·098,612	0·104,226
μ_4	0·149,090	0·156,405	0·131,097

It will be noted that in the two cases where the corrections are made they operate in the wrong direction. For the fourth moment they increase the difference between calculated and true values from about 4 per cent. to about 16 per cent. It is clear that, at least for the fairly coarse grouping of this example, Sheppard's corrections may fail completely.

3.20 Equations (3.55) were written in terms of moments about an arbitrary point. This point can, in particular, be the mean of the distribution, and accordingly we may drop the dashes and put μ_1' equal to zero in (3.55), to get the corrections appropriate for moments about the mean.

Average corrections

3.21 There is a distinct type of problem which also leads to the Sheppard corrections. Suppose there is given a distribution of unknown range and the frequencies falling into specified intervals. One may ask what are the corrections to be applied to

the raw moments so as to bring them *on the average* into closer relation with the real moments. In other words, supposing that the interval-mesh is located at random on the distribution, what are the average values of the raw moments ?

Let X_j be a fixed set of values of x, j varying from $-\infty$ to ∞ by integral values. As x_j varies from $X_j-\frac{1}{2}h$ to $X_j+\frac{1}{2}h$, x_k varies from $X_k-\frac{1}{2}h$ to $X_k+\frac{1}{2}h$.

By definition

$$\bar{\mu}'_r = \sum_{j=-\infty}^{\infty} \left\{ x_j^r \int_{-\frac{1}{2}h}^{\frac{1}{2}h} f(x_j+\xi)\, d\xi \right\}.$$

Denoting by $\mathrm{E}(\bar{\mu}'_r)$ the average as x_j varies from $X_j-\frac{1}{2}h$ to $X_j+\frac{1}{2}h$, we have

$$\begin{aligned} \mathrm{E}(\bar{\mu}'_r) &= \frac{1}{h} \int_{X_j-\frac{1}{2}h}^{X_j+\frac{1}{2}h} \Sigma \left\{ x_j^r \int_{-\frac{1}{2}h}^{\frac{1}{2}h} f(x_j+\xi)\, d\xi \right\} dx_j \\ &= \frac{1}{h} \sum_{j=-\infty}^{\infty} \int_{X_j-\frac{1}{2}h}^{X_j+\frac{1}{2}h} x_j^r \int_{-\frac{1}{2}h}^{\frac{1}{2}h} f(x_j+\xi)\, d\xi\, dx_j \\ &= \frac{1}{h} \int_{-\infty}^{\infty} x^r \int_{-\frac{1}{2}h}^{\frac{1}{2}h} f(x+\xi)\, d\xi\, dx, \end{aligned} \tag{3.57}$$

which is the same as equation (3.53) with the substitution of $\mathrm{E}(\bar{\mu}'_r)$ for $\bar{\mu}'_r$. Thus the Sheppard corrections apply for the average group-moments whatever the nature of the terminal contact.

They cannot, however, be applied indiscriminately on that ground. In place of the conditions about terminal contact and bounded derivatives, which ensure the applicability of Sheppard's corrections to any particular distribution, there is the condition that the grouping intervals are located at random on the range, which implies that although the corrections may be wrong in any given instance, the average effect in a large number of cases will be correct. In actual fact the condition about the random location of grouping does not operate very frequently for J- and U-shaped distributions, where the Sheppard corrections would not ordinarily apply ; for instance, in a distribution of incomes or deaths at given ages it is almost inevitable to begin the grouping at zero.

3.22 It is also illegitimate to drop the dashes in order to obtain corrections for moments about the mean. If the mean of the grouped distribution is denoted by m, the average value of the rth moment about the mean is given by

$$\mathrm{E}(\bar{\mu}_r) = \frac{1}{h} \int_{-\infty}^{\infty} (x-m)^r \int_{-\frac{1}{2}h}^{\frac{1}{2}h} f(x+\xi)\, d\xi\, dx,$$

where m is a function of x and the transformation of the integral which has been used earlier in this chapter is no longer legitimate. Explicit expressions for average corrections to moments about the mean have not yet been obtained. From a consideration of some particular distributions, however, Kendall (1938) concluded that for all ordinary purposes it is sufficient to use equations (3.55) as if the mean were a fixed point.

3.23 Average corrections may also be applied to discrete data which have been grouped in wider intervals, but are different from those of the continuous case. Cf. Exercise 3.13 and Craig (1936). For corrections when the Sheppard conditions are violated see Pairman and Pearson (1919), Sandon (1924), Martin (1934) and Elderton (1938*b*).

Sheppard's corrections to factorial moments

3.24 It has been shown by Wold (1934a) that for factorial moments the Sheppard corrections are as follows :—

$$\left.\begin{aligned}
\mu'_{[1]} &= \tilde{\mu}'_{[1]} \\
\mu'_{[2]} &= \tilde{\mu}'_{[2]} - \frac{h^2}{12} \\
\mu'_{[3]} &= \tilde{\mu}'_{[3]} - \frac{h^2}{4}\tilde{\mu}_{[1]} + \frac{h^3}{4} \\
\mu'_{[4]} &= \tilde{\mu}'_{[4]} - \frac{h^2}{2}\tilde{\mu}'_{[2]} + h^3\tilde{\mu}'_{[1]} - \frac{71}{80}h^4 \\
\mu'_{[5]} &= \tilde{\mu}'_{[5]} - \frac{5}{6}\tilde{\mu}'_{[3]}h^2 + \frac{5}{2}\tilde{\mu}'_{[2]}h^3 - \frac{71}{16}\tilde{\mu}'_{[1]}h^4 + \frac{31}{8}h^5 \\
\mu'_{[6]} &= \tilde{\mu}'_{[6]} - \frac{5}{4}\tilde{\mu}'_{[4]}h^2 + 5\,\tilde{\mu}'_{[3]}h^3 - \frac{213}{16}\tilde{\mu}'_{[2]}h^4 \\
&\quad + \frac{93}{4}\tilde{\mu}'_{[1]}h^5 - \frac{9129}{448}h^6
\end{aligned}\right\} \tag{3.58}$$

and in general are given by

$$\mu'_{[r]} = \sum_{j=0}^{r}\binom{r}{j}B_j^{(j+2)}(\tfrac{3}{2})\,h^j\,\tilde{\mu}'_{[r-j]}, \tag{3.59}$$

where the Bernoulli polynomial $B_j^{(j+2)}(\frac{3}{2})$ is equal to

$$(-1)^{j+1}\frac{(2j)!}{2^{2j}(j+1)!}\left(\tfrac{1}{3}+\tfrac{1}{5}+\ldots+\frac{1}{2j-1}\right), \qquad j>1,$$

and

$$B_0^{(2)}(\tfrac{3}{2}) = 1, \qquad B_1^{(3)}(\tfrac{3}{2}) = 0.$$

Note on Bernoulli numbers and polynomials

3.25 The Bernoulli numbers and (to a smaller extent) the Bernoulli polynomials have several important applications in statistics. The number of order j, B_j, is defined as the coefficient of $t^j/j!$ in $t/(e^t-1)$. Thus

$$\frac{t}{e^t-1} = \sum_{j=0}^{\infty}\frac{B_j t^j}{j!}. \tag{3.60}$$

Direct expansion gives us

$B_0 = 1$; $B_1 = -\frac{1}{2}$; odd order B's $= 0$ except B_1 ;
$B_2 = 1/6$; $B_4 = -1/30$; $B_6 = 1/42$; $B_8 = -1/30$; $B_{10} = 5/66$;
$B_{12} = -691/2730$; $B_{14} = 7/6$.

An expression we shall require (one of many expressible in terms of the Bernoulli numbers) is obtained by putting (3.60) in the form

$$\frac{1}{e^t-1} - \frac{1}{t} + \frac{1}{2} = \sum_{j=2}^{\infty}\frac{B_j t^{j-1}}{j!}$$

whence, on integration from 0 to t, we find

$$\log\{(\sinh\tfrac{1}{2}t)/\tfrac{1}{2}t\} = \sum_{j=2}^{\infty}(B_j\,t^j/j!\,j) = \sum_{j=1}^{\infty}B_{2j}\,t^{2j}/(2j)!\,2j. \tag{3.61}$$

An expansion of frequent use is the so-called Stirling's series :—

$$\log\Gamma(x) = (x-\tfrac{1}{2})\log x - x + \tfrac{1}{2}\log(2\pi) + \sum_{j=1}^{\infty} \frac{B_{2j}}{2j(2j-1)x^{2j-1}}. \tag{3.62}$$

This is an asymptotic expansion in the sense that the true value of $\log\Gamma(x)$, for x real and positive, always lies between the sum of n and $n+1$ terms, whatever n may be. In particular

$$\log\Gamma(x) = (x-\tfrac{1}{2})\log x - x + \tfrac{1}{2}\log(2\pi) + \frac{1}{12x} - \frac{1}{360x^3} + \frac{1}{1260x^5} - \frac{1}{1680x^7} + \frac{1}{1188x^9} - \dots \tag{3.63}$$

and

$$\Gamma(x) = e^{-x}x^{x-\frac{1}{2}}\sqrt{(2\pi)}\left\{1 + \frac{1}{12x} + \frac{1}{288x^2} - \frac{139}{51840x^3} - \frac{571}{2488320x^4}\dots\right\}. \tag{3.64}$$

The Bernoulli polynomial of order n and degree j in x, $B_j^{(n)}(x)$, is defined by

$$e^{xt}t^n/(e^t-1)^n = \sum_{j=0}^{\infty} t^j B_j^{(n)}(x)/j! \tag{3.65}$$

Sheppard's corrections for cumulants

3.26 As in **3.18** we have, writing 0 for it,

$$\begin{aligned}\Sigma\left\{e^{\theta x_j}\int_{-\frac{1}{2}h}^{\frac{1}{2}h} f(x_j+\xi)/d\xi\right\} &= \frac{1}{h}\int_{-\infty}^{\infty} e^{\theta x}\,dx \int_{-\frac{1}{2}h}^{\frac{1}{2}h} f(x+\xi)\,d\xi \\ &= \frac{1}{h}\int_{-\frac{1}{2}h}^{\frac{1}{2}h} e^{-\theta\xi}\,d\xi \int_{-\infty}^{\infty} e^{\theta x} f(x)\,dx \\ &= \frac{\sinh\frac{1}{2}\theta h}{\frac{1}{2}\theta h}\int_{-\infty}^{\infty} e^{\theta x} f(x)\,dx. \end{aligned} \tag{3.66}$$

The expression on the left gives the characteristic function for the grouped data and the integral on the right the true c.f. Taking logarithms and recalling (3.61) we have, for the coefficient of $0^r/r!$,

$$\kappa_r = \bar{\kappa}_r - B_r h^r/r, \qquad r>1, \tag{3.67}$$

an attractively simple result for the Sheppard corrections to cumulants. Since all B's of odd order are zero except B_1 and the first cumulant is equal to the mean, no cumulant of odd order needs any correction. For the others we have

$$\left.\begin{aligned}\kappa_2 &= \bar{\kappa}_2 - \frac{h^2}{12} \\ \kappa_4 &= \bar{\kappa}_4 + \frac{h^4}{120} \\ \kappa_6 &= \bar{\kappa}_6 - \frac{h^6}{252}\end{aligned}\right\} \tag{3.68}$$

Multivariate moments and cumulants

3.27 The foregoing results in this chapter may be readily generalized to the multivariate case. To save complicating the algebraic expressions, we shall deal with

two variates x_1 and x_2; but the reader will have little difficulty in carrying out any generalizations for more variates.

The bivariate moment μ'_{rs} about an origin a_1 for x_1 and a_2 for x_2 is defined by

$$\mu'_{rs} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x_1-a_1)^r (x_2-a_2)^s \, dF. \tag{3.69}$$

If one of r, s is zero the moment becomes the ordinary univariate moment of the row or column-border distribution of the bivariate population—the marginal distribution. In the contrary case we meet a new type of moment—the product-moment. The first product-moment μ'_{11} is of particular importance in the theory of correlation. The first product-moment about the variate means μ_{11} is known as the covariance.

As in the univariate case, bivariate moments about certain points can be expressed in terms of those about other points. If the x_1 origin is transferred from a_1 to b_1, where $c_1 = b_1-a_1$, and the x_2 origin from a_2 to b_2, where $c_2 = b_2-a_2$, we have

$$\mu'_{rs}(a_1, a_2) = (\mu'+c_1)^r (\mu'+c_2)^s \tag{3.70}$$

where the product $\mu'^i \mu'^k$ on the right is to be replaced by $\mu'_{jk}(b_1, b_2)$. This corresponds to the symbolic equation

$$\mu^r(a) = \{\mu'(b)+c\}^r$$

for the univariate case.

Methods of calculating the product-moments for numerically specified distributions will be considered in Volume 2 when we discuss the theory of correlation. The determination of bivariate moments for a mathematically specified population is a matter of evaluating double sums or double integrals, and no new statistical points call for comment.

Example 3.17

Consider the bivariate distribution

$$dF = \frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{\frac{1}{2}}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{x_1^2}{\sigma_1^2}-\frac{2\rho x_1 x_2}{\sigma_1\sigma_2}+\frac{x_2^2}{\sigma_2^2}\right)\right\} dx_1\, dx_2, \quad -\infty \leqslant x_1,\ x_2 \leqslant \infty.$$

Let us evaluate

$$M(t_1, t_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{x_1 t_1 + x_2 t_2} \, dF.$$

Making the substitution

$$\begin{aligned} \xi &= x_1-\sigma_1^2 t_1-\rho\sigma_1\sigma_2 t_2 \\ \eta &= x_2-\rho\sigma_1\sigma_2 t_1-\sigma_2^2 t_2 \end{aligned}$$

we find

$$\begin{aligned} M(t_1, t_2) &= \exp\{\tfrac{1}{2}(t_1^2\sigma_1^2+2t_1 t_2\sigma_1\sigma_2\rho+t_2^2\sigma_2^2)\} \times \\ &\qquad \frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{\frac{1}{2}}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{\xi^2}{\sigma_1^2}-\frac{2\rho\xi\eta}{\sigma_1\sigma_2}+\frac{\eta^2}{\sigma_2^2}\right)\right\} d\xi\, d\eta \\ &= \exp\{\tfrac{1}{2}(t_1^2\sigma_1^2+2\rho\sigma_1\sigma_2 t_1 t_2+t_2^2\sigma_2^2)\}. \end{aligned}$$

Now μ_{rs} is the coefficient of $\frac{t_1^r}{r!}\frac{t_2^s}{s!}$ in $M(t_1, t_2)$ and thus we find, for instance,

$$\mu_{20} = \sigma_1^2, \qquad \mu_{11} = \rho\sigma_1\sigma_2, \qquad \mu_{02} = \sigma_2^2;$$
$$\mu_{30} = \mu_{21} = \mu_{12} = \mu_{03} = 0;$$
$$\mu_{40} = 3\sigma_1^4, \qquad \mu_{31} = 3\rho\sigma_1^3\sigma_2, \qquad \mu_{22} = (1+2\rho^2)\sigma_1^2\sigma_2^2,$$
$$\mu_{13} = 3\rho\sigma_1\sigma_2^3, \qquad \mu_{04} = 3\sigma_2^4.$$

3.28 The bivariate analogue of equation (3.30) may be written

$$\exp\left\{\frac{\kappa_{10}}{1!0!}t_1+\frac{\kappa_{01}}{0!1!}t_2+\ldots+\frac{\kappa_{rs}}{r!s!}t_1^r t_2^s+\ldots\right\}$$
$$= 1+\frac{\mu'_{10}}{1!0!}t_1+\frac{\mu'_{01}}{0!1!}t_2+\ldots+\frac{\mu'_{rs}}{r!s!}t_1^r t_2^s+\ldots$$

or symbolically,

$$\exp\left\{\Sigma\frac{1}{p!}\kappa(t_1+t_2)^p\right\} = \Sigma\frac{1}{p!}\mu'(t_1+t_2)^p \tag{3.71}$$

where
$$\kappa(t_1+t_2)^p = p!\left\{\frac{\kappa_{p0}}{p!0!}t_1^p+\frac{\kappa_{p-1,1}}{(p-1)!1!}t_1^{p-1}t_2+\ldots\right\}. \tag{3.72}$$

In terms of characteristic functions we may define

$$\phi(t_1, t_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{ix_1t_1+ix_2t_2}\,dF \tag{3.73}$$

and, as before, write

$$\phi(t_1, t_2) = \sum_{r,s=0}^{\infty}\mu'_{rs}\frac{(it_1)^r}{r!}\frac{(it_2)^s}{s!}$$
$$= \exp\left\{\sum_{r,s=0}^{\infty}\kappa_{rs}\frac{(it_1)^r}{r!}\frac{(it_2)^s}{s!}\right\}, \tag{3.74}$$

subject to conditions of existence, where κ_{00} is defined equal to zero.

From these equations the bivariate moments can be expressed in terms of bivariate cumulants and vice versa. It is also possible to derive bivariate equations from the univariate equations by the following simple symbolic process.

Consider the equation (one of (3.37))

$$\mu'_4 = \kappa_4+4\kappa_3\kappa_1+3\kappa_2^2+6\kappa_2\kappa_1^2+\kappa_1^4. \tag{3.75}$$

Write this formally as

$$\mu'(r^4) = \kappa(r^4)+4\kappa(r^3)\kappa(r)+3\{\kappa(r^2)\}^2+6\kappa(r^2)\{\kappa(r)\}^2+\{\kappa(r)\}^4.$$

Treat this as a function of r and operate by $s\frac{\partial}{\partial r}$, giving

$$4\mu'(r^3s) = 4\kappa(r^3s)+12\kappa(r^2s)\kappa(r)+4\kappa(r^3)\kappa(s)+12\kappa(rs)\kappa(r^2)$$
$$+12\kappa(rs)\{\kappa(r)\}^2+12\kappa(r^2)\kappa(r)\kappa(s)+4\{\kappa(r)\}^3\kappa(s).$$

Divide through by a factor in 4 and replace r and s by suffixes relating to the first

and second variate. We obtain

$$\mu'_{31} = \kappa_{31}+3\kappa_{21}\kappa_{10}+\kappa_{30}\kappa_{01}+3\kappa_{11}\kappa_{20}+3\kappa_{11}\kappa_{10}^2+3\kappa_{20}\kappa_{10}\kappa_{01}+\kappa_{10}^3\kappa_{01}. \quad (3.76)$$

This is the expression of μ'_{31} in terms of cumulants. The process is quite general and can be justified by reference to (3.71). It also applies to the expression of cumulants in terms of moments and to formulae involving mean-moments. For example, from

$$\mu_4 = \kappa_4+3\kappa_2^2 \quad (3.77)$$

we derive, in the same manner,

$$\mu_{31} = \kappa_{31}+3\kappa_{11}\kappa_{20}. \quad (3.78)$$

A further operation of the same kind will yield (*)

$$\mu_{22} = \kappa_{22}+\kappa_{20}\kappa_{02}+2\kappa_{11}^2. \quad (3.79)$$

3.29 The following formulae are sometimes required :—

$$\left.\begin{aligned}
&\mu_{11} = \kappa_{11}; \quad \mu_{21} = \kappa_{21}; \quad \mu_{31} = \kappa_{31}+3\kappa_{20}\kappa_{11};\\
&\mu_{22} = \kappa_{22}+\kappa_{20}\kappa_{02}+2\kappa_{11}^2; \quad \mu_{41} = \kappa_{41}+4\kappa_{30}\kappa_{11}+6\kappa_{21}\kappa_{20};\\
&\mu_{32} = \kappa_{32}+\kappa_{30}\kappa_{02}+6\kappa_{21}\kappa_{11}+3\kappa_{20}\kappa_{12};\\
&\mu_{51} = \kappa_{51}+5\kappa_{40}\kappa_{11}+10\kappa_{31}\kappa_{20}+10\kappa_{30}\kappa_{21}+15\kappa_{20}^2\kappa_{11};\\
&\mu_{42} = \kappa_{42}+\kappa_{40}\kappa_{02}+8\kappa_{31}\kappa_{11}+4\kappa_{30}\kappa_{12}+6\kappa_{22}\kappa_{20}+6\kappa_{21}^2+3\kappa_{20}^2\kappa_{02}+12\kappa_{20}\kappa_{11}^2;\\
&\mu_{33} = \kappa_{33}+3\kappa_{31}\kappa_{02}+\kappa_{30}\kappa_{03}+9\kappa_{22}\kappa_{11}+9\kappa_{21}\kappa_{12}+3\kappa_{20}\kappa_{13}+9\kappa_{20}\kappa_{11}\kappa_{02}+6\kappa_{11}^3.
\end{aligned}\right\} \quad (3.80)$$

$$\left.\begin{aligned}
&\kappa_{31} = \mu_{31}-3\mu_{20}\mu_{11}; \quad \kappa_{22} = \mu_{22}-\mu_{20}\mu_{02}-2\mu_{11}^2;\\
&\kappa_{41} = \mu_{41}-4\mu_{30}\mu_{11}-6\mu_{21}\mu_{20};\\
&\kappa_{32} = \mu_{32}-\mu_{30}\mu_{02}-6\mu_{21}\mu_{11}-3\mu_{20}\mu_{12};\\
&\kappa_{51} = \mu_{51}-5\mu_{40}\mu_{11}-10\mu_{31}\mu_{20}-10\mu_{30}\mu_{21}+30\mu_{20}\mu_{11};\\
&\kappa_{42} = \mu_{42}-\mu_{40}\mu_{02}-8\mu_{31}\mu_{11}-4\mu_{30}\mu_{12}-6\mu_{22}\mu_{20}-6\mu_{21}^2+6\mu_{20}^2\mu_{02}+24\mu_{20}\mu_{11}^2;\\
&\kappa_{33} = \mu_{33}-3\mu_{31}\mu_{02}-\mu_{30}\mu_{03}-9\mu_{22}\mu_{11}-9\mu_{21}\mu_{12}-3\mu_{20}\mu_{13}+18\mu_{20}\mu_{11}\mu_{02}+12\mu_{11}^3
\end{aligned}\right\} \quad (3.81)$$

3.30 Wold (1934*b*) has given the following expressions for Sheppard corrections to bivariate moments and cumulants, the variates being grouped in intervals h_1, h_2.

$$\mu'_{rs} = \sum_{j=0}^{r}\sum_{k=0}^{s} h_1^j h_2^k \binom{r}{j}\binom{s}{k}(2^{1-j}-1)(2^{1-k}-1)B_j B_k \bar{\mu}'_{r-j,\,s-k}. \quad (3.82)$$

In particular

$$\left.\begin{aligned}
&\mu'_{20} = \bar{\mu}'_{20}-\tfrac{1}{12}h_1^2, \quad \mu'_{11} = \bar{\mu}'_{11}, \quad \mu'_{02} = \bar{\mu}'_{02}-\tfrac{1}{12}h_2^2;\\
&\mu'_{30} = \bar{\mu}'_{30}-\mu'_{10}h_1^2/4, \quad \mu'_{21} = \bar{\mu}'_{21}-\bar{\mu}'_{01}h_1^2/12;\\
&\mu'_{40} = \bar{\mu}'_{40}-\bar{\mu}'_{20}h_1^2/2+7h_1^4/240, \quad \mu'_{31} = \bar{\mu}'_{31}-\bar{\mu}'_{11}h_1^2/4;\\
&\mu'_{22} = \bar{\mu}'_{22}-\bar{\mu}'_{20}h_2^2/12-\bar{\mu}'_{02}h_1^2/12+h_1^2h_2^2/144.
\end{aligned}\right\} \quad (3.83)$$

For cumulants we have

$$\left.\begin{aligned}
\kappa_{rs} &= \bar{\kappa}_{rs}, && r,s>0\\
\kappa_{r0} &= \bar{\kappa}_{r0}-B_r h_1^r/r, && r\geqslant 2\\
\kappa_{0s} &= \bar{\kappa}_{0s}-B_s h_2^s/s. && s\geqslant 2
\end{aligned}\right\} \quad (3.84)$$

(*) The formulae have been given for μ'_{rs} in terms of κ_{rs} and vice versa up to $r+s\leqslant 6$ by Cook (1951).

Measures of skewness

3.31 We have considered measures of location and dispersion in Chapter 2. With the aid of the moments we can now proceed to consider measures of other qualities of the population, and in particular its departure from symmetry.

In a symmetrical population, mean, median and mode coincide. It is thus natural to take the deviation mean to mode or mean to median as measuring the skewness of the distribution. K. Pearson proposed the measure

$$Sk = \frac{\text{mean} - \text{mode}}{\sigma}$$

which is subject to the inconvenience of determining the mode. For a wide class of frequency-distributions known as Pearson's (cf. Chapter 6), this measure may, however, be expressed exactly in terms of the first four moments. We define

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} \tag{3.85}$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2}. \tag{3.86}$$

Then it may be shown that for Pearson curves

$$Sk = \frac{\sqrt{\beta_1}(\beta_2+3)}{2(5\beta_2-6\beta_1-9)} \tag{3.87}$$

and this equation may be taken as defining a measure of skewness applicable to all distributions whose moments up to and including the fourth exist.

The coefficient β_1 itself is also a measure of skewness. Clearly if the distribution is symmetrical it vanishes, since μ_3 vanishes, and the size of μ_3 relative to $\mu_2^{3/2}$ (or $\sqrt{\beta_1}$) will indicate the extent of the departure from symmetry.

Generally we may define

$$\left.\begin{aligned} \beta_{2n+1} &= \frac{\mu_3 \mu_{2n+3}}{\mu_2^{n+3}} \\ \beta_{2n} &= \frac{\mu_{2n+2}}{\mu_2^{n+1}} \end{aligned}\right\} \tag{3.88}$$

quantities which are not in general use but will be found to occur occasionally in statistical literature.

More convenient quantities than β_1 and β_2 for certain purposes are

$$\gamma_1 = \frac{\mu_3}{\mu_2^{3/2}} = \frac{\kappa_3}{\kappa_2^{3/2}} \tag{3.89}$$

$$\gamma_2 = \frac{\mu_4}{\mu_2^2} - 3 = \frac{\kappa_4}{\kappa_2^2}. \tag{3.90}$$

If the distribution is expressed in standard measure, γ_1 and γ_2 are its third and fourth cumulants.

Kurtosis

3.32 In the so-called "normal" distribution

$$dF = \frac{1}{\sigma\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2/\sigma^2} dx, \qquad -\infty \leqslant x \leqslant \infty,$$

β_2 attains the value 3 and γ_2 is zero. Curves for which $\gamma_2 = 0$ are called mesokurtic. Those for which $\gamma_2 > 0$ are called leptokurtic; those for which $\gamma_2 < 0$ are called platykurtic.

These names were originally conferred because it was thought that leptokurtic curves were more sharply peaked, and platykurtic curves more flat-topped, than the normal curve. This, however, is not necessarily so and although the terms are useful they are best regarded as describing the sign of γ_2 rather than the shape of the curve. Cf. Exercises 3.20 and 3.21.

Example 3.18

For the distribution of Australian marriages considered in Example 3.1 we found, for the raw moments about the mean in units of three years,

$$\bar{\mu}_2 = 7{\cdot}056{,}977, \quad \bar{\mu}_3 = 36{\cdot}151{,}595, \quad \bar{\mu}_4 = 408{\cdot}738{,}210.$$

With Sheppard's corrections these become

$$\mu_2 = 6{\cdot}973{,}644, \quad \mu_3 = 36{\cdot}151{,}595, \quad \mu_4 = 405{\cdot}238{,}888.$$

From these values we find

$$\beta_1 = 3{\cdot}854, \quad \gamma_1 = 1{\cdot}963$$
$$\beta_2 = 8{\cdot}333, \quad \gamma_2 = 5{\cdot}333$$

indicating considerable skewness and leptokurtosis.

Example 3.19

From the formulae for the moments of the binomial distribution considered in Example 3.2 we find

$$\gamma_1 = \frac{q-p}{\sqrt{(npq)}}$$
$$\gamma_2 = \frac{1-6pq}{npq}$$

so that, as $n \to \infty$, γ_1 and $\gamma_2 \to 0$. This is in accordance with a result we shall prove later, that the binomial distribution in standard measure tends to the normal form as n tends to infinity.

Moments as characteristics of a distribution

3.33 The use of moments and cumulants in determining the nature of a frequency-distribution will be abundantly illustrated in later chapters, but some general remarks may be made at this stage.

It has been noted that the characteristic function determines the moments when they exist, and it will be proved in Chapter 4 that the characteristic function also determines the distribution function. It does not, however, follow that the moments completely determine the distribution, even when moments of all orders exist. Only under certain conditions will a set of moments determine a distribution uniquely, but, fortunately for statisticians, those conditions are obeyed by all the distributions arising in statistical practice. For all ordinary purposes, therefore, a knowledge of the

moments, when they all exist, is equivalent to a knowledge of the distribution function : equivalent, that is, in the sense that it should be possible *theoretically* to exhibit all the properties of the distribution in terms of the moments.

3.34 In particular we expect that if two distributions have a certain number of moments in common they will bear some resemblance to each other. If, say, moments up to those of order n are identical we know that as n tends to infinity the distributions approach identity, and consequently we expect that by identifying the lower moments of two distributions we bring them to approximate equality. Some mathematical support for this so-called Principle of Moments may be derived from the following approach.

It is known that a function which is continuous in a finite range a to b can be represented in that range by a uniformly convergent series of polynomials in x, say $\sum_{n=0}^{\infty} P_n(x)$ where $P_n(x)$ is of degree n. Suppose we wish to represent such a function approximately by the *finite* series of powers $\sum_{n=0}^{s} a_n x^n$. The coefficients a_n may be determined by the principle of least squares, i.e. so as to make

$$\int_a^b (f - \Sigma a_n x^n)^2 \, dx \tag{3.91}$$

a minimum. Differentiating by a_j we have

$$2\int_a^b (f - \Sigma a_n x^n) x^j \, dx = 0$$

or

$$\int_a^b f x^j \, dx = \mu'_j = \int_a^b \Sigma a_n x^{n+j} \, dx. \tag{3.92}$$

If now two distributions have moments up to order s equal they must have the same least-squares approximation, for the coefficients a_n are determined by the moments in virtue of (3.92). Furthermore, if in the range the distribution f_1 differs from $\Sigma a_n x^n$ by ε_1 and f_2 by ε_2, then f_1 differs from f_2 by not more than $\varepsilon_1 + \varepsilon_2$.

A similar line of approach may be adopted when the range is infinite, the distributions in such cases being, under certain general conditions, capable of representation by a series of terms such as $e^{-x^2} P_n(x)$.(*) The same conclusion is reached.

Thus distributions which have a finite number of the lower moments in common will, in a sense, be approximations one to another. We shall encounter many cases where, although we cannot determine a distribution function explicitly, we may ascertain its moments at least up to some order ; and hence we shall be able to approximate to the distribution by finding another distribution of known form which has the same lower moments. In practice, approximations of this kind often turn out to be remarkably good, even when only the first three or four moments are equated.

(*) A theorem of Vera Myller-Lebedeff (1907, *Math. Ann.*, **64**, 388) states that a function can be expanded in a series of derivatives of $\exp(-x^2)$. This is not to be confused with the Gram-Charlier series of Chapter 6, where a function is expanded in terms of derivatives of $\exp(-\frac{1}{2}x^2)$.

Inequalities of the Tchebycheff type

3.35 It is also possible, given certain moments or expected values of a distribution, to make precise statements about the frequency or distribution functions in terms of inequalities. Let F represent a distribution with an existent second moment. For any real t we have

$$\int_{-\infty}^{\infty}\left(1-\frac{x^2}{t^2}\right)dF = \left\{\int_{-\infty}^{-t}+\int_{t}^{\infty}\right\}\left(1-\frac{x^2}{t^2}\right)dF+\int_{-t}^{t}\left(1-\frac{x^2}{t^2}\right)dF.$$

In the first two integrals on the right $1-x^2/t^2$ is $\leqslant 0$ and hence

$$1-\frac{\mu_2'}{t^2} = \int_{-\infty}^{\infty}\left(1-\frac{x^2}{t^2}\right)dF \leqslant \int_{-t}^{t}\left(1-\frac{x^2}{t^2}\right)dF$$
$$\leqslant \int_{-t}^{t} dF.$$

Thus the frequency from $-t$ to t is no less than $1-\mu_2'/t^2$. In particular, if we take an origin at the mean of the distribution,

$$F(t)-F(-t)\geqslant 1-\mu_2/t^2. \tag{3.93}$$

In the language of probability this is more usually written in the form, taking $\mu_2 = \sigma^2$ and $t = \lambda\sigma$,

$$\mathrm{P}\{|x-\mathrm{E}(x)|\leqslant\lambda\sigma\}\geqslant 1-1/\lambda^2 \tag{3.94}$$

or

$$\mathrm{P}\{|x-\mathrm{E}(x)|\geqslant\lambda\sigma\}\leqslant 1/\lambda^2. \tag{3.95}$$

This is the inequality due to Bienaymé, who published it in 1853, and Tchebycheff, who independently discovered it later. It is usually known as the Bienaymé-Tchebycheff inequality. It gives us a lower limit to the proportion of frequency lying within a range of $\lambda\sigma$ on either side of the mean. Being expressed with great generality, the upper limits are usually too wide to be of much use in determining frequencies in particular cases, but such inequalities are invaluable in proving limiting properties and are not without their uses for determining probabilities in a very rough way.

3.36 There are by now a large number of inequalities of the Tchebycheff type. We shall meet several of them in the sequel. Here we will record a few of the commoner ones for reference.

If ν_r is the absolute moment of order r it is easy to show by a direct extension of the foregoing method that

$$\mathrm{P}\{|x-\mathrm{E}(x)|/\nu_r^{1/r}\leqslant\lambda\}\geqslant 1-1/\lambda^r. \tag{3.96}$$

This inequality, given by K. Pearson (1919), gives (3.94) when $r = 2$. When the variate does not take negative values we may likewise show (cf. Exercise 2.15) that its d.f. satisfies, for all $t > 0$,

$$F(t) > 1-\mu_1'/t \tag{3.97}$$

which is sometimes called Markoff's inequality. This gives (3.94) if we take $\{x-\mathrm{E}(x)\}^2$ as the variate and $t = \lambda^2\sigma^2$.

For general distributions these inequalities cannot be improved upon, but it is rather remarkable that by imposing quite broad restrictions on the form of the frequency function we can improve them considerably. For example, if a continuous frequency function has a single mode, say m_0, and the origin is taken there,

$$P\{|x-m_0| \leqslant \lambda\tau\} \geqslant 1-4/(9\lambda^2) \tag{3.98}$$

where τ is the square root of the second moment about the mode, as compared with (3.94). This expression, in a different form, is traceable to Gauss.

3.37 For an account of these inequalities see a comprehensive review of the subject by Godwin (1955). By assuming that two mean values (e.g. two moments) are known we may obtain more refined inequalities. Mallows (1956), by imposing further restrictions on the frequency function, such as the condition that there is only one point of inflection on each side of the mode, has obtained some very fair approximations to the distribution function itself. The subject is mathematically attractive and might well be accorded a separate monograph.

EXERCISES

3.1 Show that the discontinuous (Poisson) distribution whose frequencies corresponding to the values 0, 1, . . , j, . . are

$$e^{-\lambda}\left(1, \frac{\lambda}{1!}, \ldots, \frac{\lambda^j}{j!}, \ldots\right)$$

has, for the moments about the mean,

$$\mu_2 = \lambda,\ \mu_3 = \lambda,\ \mu_4 = \lambda(1+3\lambda),\ \mu_5 = \lambda(1+10\lambda),\ \mu_6 = \lambda(1+25\lambda+15\lambda^2).$$

3.2 For the distribution

$$dF = kx^{-p}e^{-\gamma/x}\,dx, \qquad 0 \leqslant x \leqslant \infty\ ;\ \gamma > 0,$$

show that, about the origin,

$$\mu_r' = \frac{\gamma^r\,\Gamma(p-r-1)}{\Gamma(p-1)}$$

if $r < p-1$, and does not exist in the contrary case.

3.3 In the distribution

$$dF = k\left(1+\frac{x^2}{a^2}\right)^{-m} \exp\{-\nu \arctan(x/a)\}\,dx, \qquad -\infty \leqslant x \leqslant \infty,$$

show that, about the origin,

$$\mu_r' = ka^{r+1}\int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} \cos^{2m-r-2}\theta \sin^r\theta\, e^{-\nu\theta}\,d\theta$$

and hence that

$$\mu_r' = \frac{a}{2m-r-1}\{(r-1)a\mu_{r-2}' - \nu\mu_{r-1}'\}.$$

3.4 $f(x)$ is a continuous frequency function, symmetrical about $x = 0$ and ranging from $-a$ to $+a$. It has a single maximum at $x = 0$. Show that

$$\mu_{2r} < \frac{a^{2r}}{2r+1}, \qquad r \geqslant 1.$$

3.5 Show that for any symmetrical distribution the cumulants of odd order (except κ_1) vanish if they exist.

3.6 Show that for the distribution

$$dF = e^{-x/\sigma}\,dx/\sigma, \qquad 0 \leqslant x \leqslant \infty\ ;\ \sigma > 0,$$
$$\kappa_r = \sigma^r(r-1)!$$

3.7 Show that e^{tx} may be expanded in an infinite series, valid in $-\infty \leqslant x \leqslant \infty$,

$$1+(e^{it}-1)\,x^{[1]}+(e^{it}-1)^2\,\frac{x^{[2]}}{2!}+\ldots+(e^{it}-1)^r\,\frac{x^{[r]}}{r!}+\ldots$$

the factorials being taken with unit interval; and hence that

$$\mu'_{[r]} = [D^r\,\phi(t)]_{t\to 0}$$

where

$$D = \frac{d}{d(e^{it})}.$$

Hence show that, for the binomial $(q+p)^n$ about the origin,

$$\mu'_{[r]} = n^{[r]}p^r.$$

3.8 Show that the distribution whose frequency at the variate-value $\pm 2r$ (r integral) is

$$e^{-2a}\left\{\frac{a^{2r}}{0!\,(2r)!}+\frac{a^{2r+2}}{1!\,(2r+1)!}+\frac{a^{2r+4}}{2!\,(2r+2)!}+\ldots\right\}$$

and at $\pm(2r+1)$ is

$$e^{-2a}\left\{\frac{a^{2r+1}}{(2r+1)!}+\frac{a^{2r+3}}{1!\,(2r+2)!}+\frac{a^{2r+5}}{2!\,(2r+3)!}+\ldots\right\}$$

has odd-order cumulants equal to zero and even-order cumulants equal to $2a$.

3.9 Show that

$$\mu'_r = \sum_{j=1}^{r}\binom{r-1}{j-1}\mu'_{r-j}\,\kappa_j$$

and hence that

$$\kappa_r = (-1)^{r-1}\begin{vmatrix} \mu'_1 & 1 & 0 & 0 & \cdot & \cdot & 0 \\ \mu'_2 & \binom{1}{0}\mu'_1 & 1 & 0 & \cdot & \cdot & 0 \\ \mu'_3 & \binom{2}{0}\mu'_2 & \binom{2}{1}\mu'_1 & 1 & \cdot & \cdot & 0 \\ & \cdot & \cdot & \cdot & \cdot & \cdot & 1 \\ \mu'_r & \binom{r-1}{0}\mu'_{r-1} & \binom{r-1}{1}\mu'_{r-2} & \cdot & \cdot & \cdot & \binom{r-1}{r-2}\mu'_1 \end{vmatrix}$$

3.10 Show that for the distribution

$$dF = dx, \qquad 0 \leqslant x \leqslant 1,$$

grouped into an integral number of intervals of equal width h, the corrections to the second and fourth moments about the mean are

$$\mu_2 = \bar{\mu}_2 + \frac{h^2}{12}$$

$$\mu_4 = \bar{\mu}_4 + \bar{\mu}_2\frac{h^2}{2} + \frac{h^4}{80}.$$

(Cf. Elderton, 1938*b*. Note that the first is exactly, and the second approximately, the Sheppard correction *with sign reversed.*)

3.11 If ∂_p stands for the operator such that

$$\begin{aligned} \partial_p \mu'_r &= r^{[p]} \mu'_{r-p} \qquad r \geqslant p \\ &= 0, \qquad\qquad r < p, \end{aligned}$$

and ∂_p is distributive when applied to products, e.g.

$$\partial_p(AB) = B(\partial_p A) + A(\partial_p B),$$

show that ∂_p annihilates every cumulant (considered as a function of the moments) except κ_p, and that

$$\partial_p \kappa_p = p!$$

3.12 If $f(x)$ is an odd function of x of period $\frac{1}{2}$, show that

$$\int_0^\infty x^r x^{-\log x} f(\log x)\, dx = 0$$

for all integral values of r. Hence show that the distributions

$$dF = x^{-\log x}\{1 - \lambda \sin(4\pi \log x)\}\, dx, \qquad 0 \leqslant x \leqslant \infty\ ;\ \ 0 \leqslant \lambda \leqslant 1,$$

have the same moments whatever the value of λ. (Stieltjes, 1918)

3.13 Show that if the frequencies of a discontinuous distribution are distributed at equal intervals $\frac{h}{m}$, m in each grouping-interval h, the average grouping corrections to the cumulants are given by

$$\kappa_r = \bar{\kappa}_r - \frac{B_r h^r}{r}\left(1 - \frac{1}{m^r}\right).$$

(Craig, 1936)

3.14 Show that for the bivariate distribution

$$dF = \frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{\frac{1}{2}}} \exp\left[-\frac{1}{2(1-\rho^2)}\left\{\frac{x_1^2}{\sigma_1^2} - \frac{2\rho x_1 x_2}{\sigma_1\sigma_2} + \frac{x_2^2}{\sigma_2^2}\right\}\right] dx_1\, dx_2, \qquad -\infty \leqslant x_1,\ x_2 \leqslant \infty,$$

all cumulants κ_{rs}, r, $s > 2$, vanish; and further, if

$$\lambda_{rs} = \frac{\mu_{rs}}{\sigma_1^r \sigma_2^s}$$

$$\lambda_{rs} = (r+s-1)\rho\,\lambda_{r-1,\,s-1} + (r-1)(s-1)(1-\rho^2)\,\lambda_{r-2,\,s-2}$$

$$\lambda_{2r,\,2s} = \frac{(2r)!\,(2s)!}{2^{r+s}} \sum_{j=0}^{t} \frac{(2\rho)^{2j}}{(r-j)!\,(s-j)!\,(2j)!}$$

$$\lambda_{2r+1,\,2s+1} = \frac{(2r+1)!\,(2s+1)!}{2^{r+s}}\,\rho \sum_{j=0}^{t} \frac{(2\rho)^{2j}}{(r-j)!\,(s-j)!\,(2j+1)!}$$

$$\lambda_{2r,\,2s+1} = \lambda_{2r+1,\,2s} = 0,$$

where t is the smaller of r and s. In particular,

$$\begin{aligned} &\lambda_{11} = \rho, \quad \lambda_{31} = 3\rho, \quad \lambda_{51} = 15\rho, \quad \lambda_{71} = 105\rho, \quad \lambda_{91} = 945\rho\,; \\ &\lambda_{22} = (1+2\rho^2), \quad \lambda_{24} = 3(1+4\rho^2), \quad \lambda_{26} = 15(1+6\rho^2), \\ &\qquad \lambda_{28} = 105(1+8\rho^2), \quad \lambda_{2,\,10} = 945(1+10\rho^2)\,; \\ &\lambda_{33} = 3\rho(3+2\rho^2), \quad \lambda_{35} = 15\rho(3+4\rho^2), \\ &\lambda_{44} = 3(3+24\rho^2+8\rho^4). \end{aligned}$$

3.15 (Liapounoff's inequality for absolute moments.) Beginning with one form of Hölder's inequality

$$\Sigma x^\alpha y^\beta \leqslant \Sigma x^\alpha \Sigma y^\beta\,; \qquad \alpha + \beta = 1,$$

put $x = u^c$, $y = u^a$, $\beta = 1-\alpha$, $\alpha = (a-b)/(a-c)$ and hence deduce the inequality

$$\nu_b^{a-c} \leqslant \nu_c^{a-b}\nu_a^{b-c}.$$

3.16 Verify the expressions for κ_{42} and κ_{33} in equation (3.81).

3.17 A distribution is grouped in intervals of h. Show that as the grouping grid moves along the variate axis the raw moments have period h. Hence, writing

$$\begin{aligned}\mu'_r &= \sum_{j=-\infty}^{\infty} \zeta^r \int_{\zeta-\frac{1}{2}h}^{\zeta+\frac{1}{2}h} f(x)\,dx \\ &= A_0 + \sum_{j=1}^{\infty} A_j \sin j\theta + \sum_{j=1}^{\infty} B_j \cos j\theta\end{aligned}$$

where $\zeta = (j+\theta/2\pi)h,$

show that

$$A_0 = \frac{1}{h}\int_{-\infty}^{\infty} x^r \int_{-\frac{1}{2}h}^{\frac{1}{2}h} f(x+\zeta)\,d\zeta\,dx$$

and hence that this leads to Sheppard's corrections.

Show also that

$$A_s = \frac{2}{h}\int_{-\infty}^{\infty} f(x)\,dx \int_{x-\frac{1}{2}h}^{x+\frac{1}{2}h} \zeta^r \sin(2\pi s\zeta/h)\,d\zeta$$

with a similar expression for B_s.

For the distribution $\dfrac{1}{\sqrt{(2\pi)}}e^{-\frac{1}{2}x^2}dx, \quad -\infty \leqslant x \leqslant \infty$, show that, for corrections to the mean, the B's vanish and $A_s = (-1)^{s+1}\dfrac{h}{\pi s}\exp\left(-\dfrac{2s^2\pi^2}{h^2}\right)$ and hence that, even for a coarse grouping $h = 1$, the correction to the mean is not greater than $e^{-2\pi^2}/\pi$ approximately.

(Fisher, 1921*a*)

3.18 (The Gauss-Winckler inequality.) By considering $1-f(x)/f(\tilde{x})$ as a distribution function and using Liapounoff's inequality, show that for a frequency function which is differentiable and has a single mode $\tilde{x}$, the absolute moments about $\tilde{x}$ obey the relation

$$\{(r+1)\nu'_r\}^{1/r} \leqslant \{(n+1)\nu'_n\}^{1/n}, \quad r < n.$$

In particular show that $\beta_2 \geqslant 1{\cdot}8$.

3.19 Show that

$$\begin{vmatrix} \mu'_0 & \mu'_1 & \mu'_2 \\ \mu'_1 & \mu'_2 & \mu'_3 \\ \mu'_2 & \mu'_3 & \mu'_4 \end{vmatrix} > 0$$

and hence that $\beta_2 > 1+\beta_1$.

3.20 Consider the two distributions

$$\begin{aligned} f_1(x) &= \frac{1}{\sqrt{(2\pi)}}e^{-\frac{1}{2}x^2} \\ f_2(x) &= \frac{1}{\sqrt{(2\pi)}}e^{-\frac{1}{2}x^2}\{1 + A(x^8 - 22x^6 + 116x^4 - 126x^2 + 3)\}, \end{aligned}$$

where A is a small positive constant.

Show that both have zero mean, unit variance and only one point of inflexion between 0 and ∞, on suitable choice of A. Show also that $f_2 > f_1$ when x is large and when x is near zero ; but that f_2 is platykurtic. (Dyson, 1943)

3.21 The following four unimodal symmetrical f.f.'s have zero mean and unit variance. Show that they have the values of μ_4 and the maximum ordinates indicated.

		Value of μ_4	Maximum ordinate
(1)	$\frac{1}{3\sqrt{\pi}}\left(\frac{9}{4}+x^4\right)e^{-x^2}$	2·75	0·423
(2)	$\frac{3}{2\sqrt{(2\pi)}}e^{-\frac{1}{2}x^2}-\frac{1}{6\sqrt{\pi}}\left(\frac{9}{4}+x^4\right)e^{-x^2}$	3·125	0·387
(3)	$\frac{1}{6\sqrt{\pi}}(e^{-\frac{1}{2}x^2}+4e^{-x^2})$	4·5	0·470
(4)	$\frac{3\sqrt{3}}{16\sqrt{\pi}}(2+x^2)\,e^{-\frac{1}{2}x^2}$	2·667	0·366

Comparing these with the normal values $\mu_4 = 3$, max. ordinate = 0·399, show that any combination of higher or lower ordinate at the centre and lepto- or platykurtosis is possible.
(Kaplansky, 1945)

3.22 Defining, by analogy with **3.31**,

$$s = \text{(mean} - \text{median)/standard deviation,}$$

show that for any distribution

$$-1 \leqslant s \leqslant 1.$$

(Hotelling and Solomons, 1932)

3.23 Using (3.61) show that the c.f. of the rectangular distribution $dF = dx/h$, $-\frac{1}{2}h \leqslant x \leqslant \frac{1}{2}h$, is given by

$$\phi(t) = \sinh(\tfrac{1}{2}ith)/(ith)$$

with cumulants

$$\kappa_{2r-1} = 0,$$
$$\kappa_{2r} = B_{2r}h^{2r}/(2r), \qquad r = 1, 2, \ldots$$

Similarly for the discrete rectangular distribution

$$\mathrm{P}(x = r) = \frac{1}{n}, \qquad r = 1, 2, \ldots, n,$$

show that

$$\phi(t) = \sinh(\tfrac{1}{2}itn)/\{n \sinh(\tfrac{1}{2}it)\}$$

with cumulants

$$\kappa_1 = \tfrac{1}{2}(n+1)$$
$$\kappa_{2r+1} = 0,$$
$$\kappa_{2r} = B_{2r}(n^{2r}-1)/(2r), \qquad r = 1, 2, \ldots$$

3.24 Show that if the variate takes zero or positive integer values only, (3.29) may be inverted to give

$$f_j = \frac{1}{j!}\sum_{r=0}^{n-j}(-1)^r\mu'_{[j+r]}/r!$$

Invert the factorial moments obtained in Example 3.8 by this means.

CHAPTER 4

CHARACTERISTIC FUNCTIONS

4.1 In the previous chapter we have introduced the characteristic function

$$\phi(t) = \int_{-\infty}^{\infty} e^{itx}\, dF \tag{4.1}$$

as a moment-generating function. It has many other useful and important properties which give it a central rôle in statistical theory and in this chapter we shall give an account of them. To establish our general theorems we shall, however, have to use somewhat advanced mathematics; and the reader who is interested principally in the statistical applications may prefer to omit the proofs, especially at first reading, and merely note the results. Broadly speaking, this part of our theory may be regarded as a special case of the general theory of Fourier transforms—special because our functions are frequency and distribution functions. Several of the proofs in this chapter are based on the book by P. Lévy, *Calcul des Probabilités*, Gauthier-Villars, Paris, 1925.(*)

4.2 We recall, in the first instance, that $\phi(t)$ always exists, since

$$\left| \int_{-\infty}^{\infty} e^{itx}\, dF \right| \leqslant \int_{-\infty}^{\infty} |\, e^{itx}\,|\, dF = \int_{-\infty}^{\infty} dF = 1, \tag{4.2}$$

so that the defining integral converges absolutely. Further, $\phi(t)$ is uniformly continuous in t and differentiable j times under the integral sign if the resulting expressions exist and are uniformly convergent, for which it is sufficient that ν_j exists. For then

$$\begin{aligned} |\, \phi^{(j)}(t)\,| &= \left| \int_{-\infty}^{\infty} x^j e^{ixt}\, dF \right| \\ &\leqslant \int_{-\infty}^{\infty} |\, x^j\,|\, dF = \nu_j. \end{aligned} \tag{4.3}$$

The Inversion Theorem

4.3 We now prove the fundamental theorem of the theory of characteristic functions, which will be called the Inversion Theorem, namely that the characteristic function (c.f.) uniquely determines the distribution function; more precisely, if $\phi(t)$ is given by (4.1) then

$$F(x) - F(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi(t) \frac{1 - e^{-ixt}}{it}\, dt, \tag{4.4}$$

the integral being understood as a principal value, i.e. as

$$\lim_{c \to \infty} \frac{1}{2\pi} \int_{-c}^{c} \phi(t) \frac{1 - e^{-ixt}}{it}\, dt.$$

(*) Lukacs (*Characteristic Functions*, 1960, Griffin) has recently published an elegant monograph on the subject.

Further, if $F(x)$ is continuous everywhere and $dF = f(x)\,dx$

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \phi(t)\, e^{-ixt}\, dt, \tag{4.5}$$

the integral, as before, being a principal value if there is not separate convergence at the limits. Equation (4.5) may be compared with the form

$$\phi(t) = \int_{-\infty}^{\infty} f(x)\, e^{ixt}\, dx, \tag{4.6}$$

the comparison exhibiting the kind of reciprocal relationship which exists between $f(x)$ and $\phi(t)$.

Equation (4.4) contains $F(0)$ and we may avoid introducing this quantity by an alternative form

$$F(x) = \frac{1}{2} + \frac{1}{2\pi}\int_0^{\infty} \frac{e^{itx}\phi(-t) - e^{-itx}\phi(t)}{it}\, dt. \tag{4.7}$$

We require as a preliminary result the theorem that for a real number ξ,

$$\frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\sin t\xi}{t}\, dt = \left.\begin{array}{ll} -1, & \xi < 0 \\ 0, & \xi = 0 \\ 1, & \xi > 0 \end{array}\right\} \tag{4.8}$$

The quantity on the right of (4.8) is written sgn ξ, read as "signum ξ." The result is, perhaps, most easily established by evaluating the complex integral $\int \frac{e^{iz\xi}}{iz}\, dz$ round a contour consisting of a large semicircle, the real axis and a small semicircle round the origin. We note that for a variate u and some fixed x

$$\begin{aligned} \int_{-\infty}^{\infty} \operatorname{sgn}(u-x)\, dF(u) &= -\int_{-\infty}^{x} dF(u) + \int_{x}^{\infty} dF(u) \\ &= 1 - 2F(x). \end{aligned} \tag{4.9}$$

Now we have, for a positive number c, the uniformly convergent integral

$$\begin{aligned} I_c \equiv \int_{-c}^{c} \frac{1-e^{-ixt}}{it}\,\phi(t)\, dt &= \int_{-c}^{c} \frac{1-e^{-ixt}}{it}\int_{-\infty}^{\infty} e^{itu}\, dF(u)\, dt \\ &= \int_{-c}^{c}\int_{-\infty}^{\infty} \frac{e^{itu} - e^{it(u-x)}}{it}\, dF(u)\, dt \\ &= \int_{-c}^{c}\int_{-\infty}^{\infty} \frac{\sin tu - \sin\{t(u-x)\}}{t}\, dF(u)\, dt. \end{aligned}$$

Because the integral is uniformly convergent, we may change the order of integration to obtain

$$I_c = \int_{-\infty}^{\infty}\int_{-c}^{c} \left\{\frac{\sin tu}{t} - \frac{\sin\{t(u-x)\}}{t}\right\} dt\, dF(u). \tag{4.10}$$

The integral with respect to t is continuous in c and bounded. We may therefore let c tend to infinity to obtain, from (4.8) and (4.9),

$$\begin{aligned} \lim_{c\to\infty} I_c &= \pi\int_{-8}^{\infty} \{\operatorname{sgn} u - \operatorname{sgn}(u-x)\}\, dF(u) \\ &= \pi[1 - 2F(0) - \{1 - 2F(x)\}] = 2\pi\{F(x) - F(0)\}. \end{aligned}$$

Equation (4.4) follows. Equation (4.7) may be obtained in a similar way by use of the result

$$\frac{1}{2\pi}\int_0^\infty \frac{\sin t\xi}{t}\,dt = \operatorname{sgn}\xi. \tag{4.11}$$

Further, if $f(x)$ exists we may differentiate (4.4) under the integral sign, since the integral converges uniformly, to obtain (4.5). We also have, on differentiating (4.7),

$$f(x) = \frac{1}{2\pi}\int_0^\infty \{e^{itx}\phi(-t)+e^{-itx}\phi(t)\}\,dt. \tag{4.12}$$

4.4 From the definition of the c.f. we note that $\phi(t)$ and $\phi(-t)$ are conjugate quantities and hence, if $\mathscr{R}(t)$ and $\mathscr{I}(t)$ are the real and imaginary parts of $\phi(t)$ we have, from (4.4),

$$F(x)-F(0) = \frac{1}{2\pi}\int_{-\infty}^\infty \frac{\mathscr{R}(t)\sin xt+\mathscr{I}(t)(1-\cos xt)}{t}\,dt, \tag{4.13}$$

and, from (4.7),

$$F(x) = \tfrac{1}{2}+\frac{1}{\pi}\int_0^\infty \frac{\mathscr{R}(t)\sin xt-\mathscr{I}(t)\cos xt}{t}\,dt. \tag{4.14}$$

Similarly we have

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^\infty \{\mathscr{R}(t)\cos xt+\mathscr{I}(t)\sin xt\}\,dt. \tag{4.15}$$

4.5 Consider now the expression J_c defined by

$$J_c = \frac{1}{2c}\int_{-c}^{c} e^{-ixt}\phi(t)\,dt. \tag{4.16}$$

If the distribution function $F(x)$ has a derivative $f(x)$ we have, from (4.5),

$$\lim_{c\to\infty} J_c = 2\pi \lim f(x)/2c = 0$$

and hence J_c tends to zero at all points where $F(x)$ is continuous and differentiable, i.e. if the frequency function is continuous.

If the distribution function is discontinuous, consider one point of discontinuity, say where the frequency is f_j. The contribution of this part of the frequency to $\phi(t)$ is $f_j \exp(itx_j)$ and thus the contribution to J_c is

$$\frac{1}{2c}\int_{-c}^{c} f_j \exp(itx_j - itx)\,dt$$
$$= \frac{1}{2c} f_j \left[\frac{\exp\{it(x_j-x)\}}{i(x_j-x)}\right]_{-c}^{c}.$$

If $x_j \neq x$ this clearly tends to zero as $c\to\infty$. But if $x_j = x$ the integral is simply

$$\frac{1}{2c}\int_{-c}^{c} f_j\,dt = f_j.$$

Thus J_c tends to f_j at $x = x_j$. Hence, if J_c tends to zero at a point, there is no discontinuity in the distribution function at that point; but if it tends to a positive number

f_j the d.f. is discontinuous at that point and the frequency is f_j. This gives us a criterion whether a given c.f. represents a continuous distribution or not.

4.6 Looking similarly to (4.6) we see that if the distribution has discontinuous frequencies at x_j ($j = 1, 2, \ldots$) the c.f. is given by

$$\phi(t) = \Sigma f_j \exp(itx_j)$$

and

$$\begin{aligned} F(x) &= \frac{1}{2} + \frac{1}{2\pi}\int_0^\infty \frac{\Sigma f_j \exp\{it(x-x_j)\} - \Sigma f_j \exp\{-it(x-x_j)\}}{it}\,dt \\ &= \frac{1}{2} + \frac{1}{\pi}\Sigma f_j \int_0^\infty \frac{\sin t(x-x_j)}{t}\,dt \\ &= \tfrac{1}{2} + \tfrac{1}{2}\Sigma f_j \operatorname{sgn}(x-x_j). \end{aligned} \tag{4.17}$$

This has a saltus at $x = x_j$ equal to f_j. We note, however, that for $x = x_k$

$$\begin{aligned} F(x_k) &= \tfrac{1}{2} + \tfrac{1}{2}\sum_{j=1}^{k-1} f_j - \tfrac{1}{2}\sum_{j=k+1}^{\infty} f_j \\ &= \tfrac{1}{2}\left(1 - \sum_{k+1}^{\infty} f_j\right) + \tfrac{1}{2}\sum_{1}^{k-1} f_j \\ &= \sum_{1}^{k-1} f_j + \tfrac{1}{2} f_k. \end{aligned} \tag{4.18}$$

The value assigned to $F(x)$ at the saltus is thus the frequency up to that point plus one-half of the increase in frequency at that point. On our ordinary definition, with $F(x)$ continuous on the right, the value of $F(x_k)$ would be $\sum_1^k f_j$. This is a point to remember when the distribution function of a discontinuous distribution is obtained from the characteristic function.

Example 4.1

We found in Example 3.4 that the characteristic function of the normal distribution

$$dF = \frac{1}{\sigma\sqrt{(2\pi)}} e^{-x^2/2\sigma^2}\,dx, \qquad -\infty \leqslant x \leqslant \infty,$$

is

$$\phi(t) = \exp(-\tfrac{1}{2}t^2\sigma^2).$$

Suppose we are given this function and require to find the distribution, if any, of which it is the characteristic function.

In the first place we note that the distribution, if any, is continuous. For

$$J_c = \frac{1}{2c}\int_{-c}^{c} \exp\{-\tfrac{1}{2}t^2\sigma^2 - itx\}\,dt$$

and the integral is less in modulus than $\int_{-c}^{c} \exp(-\tfrac{1}{2}t^2\sigma^2)\,dt$, which is less than

$\int_{-\infty}^{\infty} \exp(-\frac{1}{2}t^2\sigma^2)\,dt = \sqrt{(2\pi)}/\sigma$. Thus $J_c \to 0$ everywhere. We have for the frequency function, if any,

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp\{-\tfrac{1}{2}t^2\sigma^2 - itx\}\,dt$$
$$= \frac{1}{2\pi}e^{-x^2/2\sigma^2}\int_{-\infty}^{\infty} \exp\{-\tfrac{1}{2}(t\sigma + ix/\sigma)^2\}\,dt.$$

This may be regarded as an integral in the complex plane along a line parallel to the real axis. Taking $t\sigma + ix/\sigma = u$ as a new variable we find that the integral is $\sqrt{(2\pi)}/\sigma$. Thus

$$f(x) = \frac{1}{\sigma\sqrt{(2\pi)}}e^{-\frac{1}{2}x^2/\sigma^2}.$$

This is everywhere positive and $\int_{-\infty}^{\infty} dF$ converges to unity. Hence it is, in fact, a frequency function with the given expression for its c.f.

Example 4.2

To find the frequency function, if any, for which

$$\phi(t) = e^{-|t|}.$$

We note that J_c tends to zero and that the distribution, if any, is continuous. We then have for $f(x)$, if it exists,

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-|t|}e^{-itx}\,dt$$
$$= \frac{1}{2\pi}\left[\int_0^{\infty} e^{-t-itx}\,dt + \int_{-\infty}^{0} e^{t-itx}\,dt\right]$$
$$= \frac{1}{2\pi}\int_0^{\infty} e^{-t}(e^{itx} + e^{-itx})\,dt$$
$$= \frac{1}{\pi}\int_0^{\infty} e^{-t}\cos tx\,dt.$$

This may be evaluated by two partial integrations. We find

$$f(x) = \frac{1}{\pi}\left[-e^{-t}\cos tx\right]_0^{\infty} - \frac{x}{\pi}\int_0^{\infty} e^{-t}\sin tx\,dt$$
$$= \frac{1}{\pi} + \frac{x}{\pi}\left[e^{-t}\sin tx\right]_0^{\infty} - \frac{x^2}{\pi}\int_0^{\infty} e^{-t}\cos tx\,dt$$
$$= \frac{1}{\pi} - x^2 f(x).$$

Thus

$$f(x) = \frac{1}{\pi(1+x^2)}, \qquad -\infty \leqslant x \leqslant \infty.$$

As before, this function can represent a frequency function, and it may be verified that $f(x)$ has, in fact, the required characteristic function.

Example 4.3

Does there exist a frequency function for which

$$\phi(t) = e^{it}?$$

We have

$$J_c = \frac{1}{2_c}\int_{-c}^{c} e^{it-itx}\,dt = \frac{1}{2c}\int_{-c}^{c} e^{it(1-x)}\,dt.$$

If $1-x$ is not zero the integral is

$$\int_{-c}^{c} [\cos\{(1-x)t\} + i\sin\{(1-x)t\}]\,dt.$$

Since $\sin t$ is an odd function this is equal to

$$\int_{-c}^{c} \cos\{(1-x)t\}\,dt = \left[-\frac{\sin(1-x)t}{1-x}\right]_{-c}^{c}.$$

This does not converge, but it is bounded and hence $J_c \to 0$.

If, however, $x = 1$, the integral is simply

$$\int_{-c}^{c} dt$$

and thus $J_c = 1$.

Thus there is unit frequency at $x = 1$ and it is seen at once that this accounts for the whole of the frequency, so that there is no frequency elsewhere. The distribution thus consists of a unit at $x = 1$. This is otherwise evident from the consideration that $\log \phi(t) = it$, so that the second cumulant is zero and there is no dispersion.

Example 4.4

For what distribution, if any, are the cumulants given by $\kappa_r = (r-1)!$?

The series

$$\sum_{j=1}^{\infty} \kappa_j \frac{(it)^j}{j!} = \Sigma \frac{(it)^j}{j}$$

converges absolutely for $|t| < 1$ and is thus equal to the cumulant g.f., say $\psi(t)$, if such a function exists. We have

$$\psi(t) = \Sigma \frac{(it)^j}{j} = -\log(1-it),$$

and thus

$$\phi(t) = \frac{1}{1-it}.$$

If the frequency function exists we have

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{-ixt}}{1-it}\,dt.$$

This integral may be evaluated by integrating the complex function $\dfrac{e^{-ixz}}{1-iz}$ round a contour consisting of the real axis and the infinite semicircle below that axis. The first

part reduces to the integral we are seeking with its sign reversed. On the semicircle of radius R we have $z = R(\cos\theta + i\sin\theta)$ and the integrand becomes

$$\frac{\exp(-ixR\cos\theta + xR\sin\theta)}{1 - iR\cos\theta + R\sin\theta}.$$

θ here lies between π and 2π and hence $\sin\theta$ is negative. Hence *if x is positive* the expression is less in modulus than

$$\frac{e^{-xR|\sin\theta|}}{R},$$

i.e. tends to zero as $R \to \infty$.

Now the function $\dfrac{e^{-ixz}}{1-iz}$ has a pole within the domain of integration at $z = -i$ and the residue there is ie^{-x}. Hence

$$f(x) = \frac{1}{2\pi}.2\pi e^{-x}$$
$$= e^{-x}, \qquad 0 \leqslant x \leqslant \infty.$$

More generally, if $\kappa_r = p(r-1)!$, $p > 0$, it will be found that the residue of $\dfrac{e^{-ixz}}{(1-iz)^p}$ is $\dfrac{ix^{p-1}e^{-x}}{\Gamma(p)}$, so that the distribution is

$$f(x) = \frac{e^{-x}x^{p-1}}{\Gamma(p)}, \qquad 0 \leqslant x \leqslant \infty\ ;\ \ p > 0.$$

Example 4.5

For what distribution, if any, are all cumulants of odd order zero and those of even order a constant, say $2a$?

We have for the c.g.f.

$$\psi(t) = 2a\left\{\frac{(it)^2}{2!} + \frac{(it)^4}{4!} + \ldots\right\}.$$

This series converges and

$$\psi(t) = 2a(\cos t - 1).$$

Hence
$$\phi(t) = e^{2a(\cos t - 1)}$$

If we try to integrate $\displaystyle\int_{-\infty}^{\infty} e^{2a(\cos t-1)}\, e^{-itx}\, dt$ in the ordinary way, we fail. Let us then look into the question of continuity of the distribution function.

We have

$$J_c = \frac{e^{-2a}}{2c}\int_{-c}^{c} e^{2a\cos t}\, e^{-ixt}\, dt$$
$$= \frac{e^{-2a}}{2c}\int_{-c}^{c} \sum_{j=0}^{\infty} \frac{(2a)^j}{j!}\cos^j t\, e^{-ixt}\, dt.$$

The series is uniformly convergent and hence

$$J_c = \frac{e^{-2a}}{2c}\sum_{j=0}^{\infty}\int_{-c}^{c}\frac{(2a)^j}{j!}\cos^j t\, e^{-ixt}\,dt$$
$$= \frac{e^{-2a}}{2c}\sum_{=0}^{\infty}\int_{-c}^{c}\frac{(2a)^j}{j!}\cos^j t\cos xt\,dt,$$

since $\sin xt$ is an odd function.

Consider now the integral $\int_{-\infty}^{\infty} 2^j\cos^j t\cos xt\,dt$. By a well-known expansion

$$2^j\cos^j t\cos xt = \tfrac{1}{2}(e^{it}+e^{-it})^j(e^{ixt}+e^{-ixt})$$
$$= \tfrac{1}{2}(e^{ixt}+e^{-ixt})\left\{e^{ijt}+\binom{j}{1}e^{(j-2)it}+\ldots+e^{-ijt}\right\}.$$

The only part of this expression of present interest is the constant term, the others not contributing more than a finite amount. The coefficient of e^0 is zero unless x is integral in absolute value, and in that case is

$$\frac{1}{2}\left\{\binom{j}{\frac{1}{2}(j-x)}+\binom{j}{\frac{1}{2}(j+x)}\right\} = \binom{j}{\frac{1}{2}(j-x)}.$$

Thus J_c tends to zero unless x is integral in absolute value, and in the latter case

$$J_c \to e^{-2a}\sum_{j=x}^{\infty}\frac{a^j}{j!}\binom{j}{\frac{1}{2}(j-x)}.$$

Thus, if x is even, say $2r$, the frequency at $x = \pm 2r$ is

$$e^{-2a}\left\{\frac{a^{2r}}{(2r)!}+\frac{a^{2r+2}}{(2r+2)!}\binom{2r+2}{1}+\frac{a^{2r+4}}{(2r+4)!}\binom{2r+4}{2}+\ldots\right\}$$
$$= e^{-2a}\left\{\frac{a^{2r}}{(2r)!}+\frac{a^{2r+2}}{(2r+1)!\,1!}+\frac{a^{2r+4}}{(2r+2)!\,2!}+\ldots\right\},$$

and if x is odd the frequency at $x = 2r+1$ is

$$e^{-2a}\left\{\frac{a^{2r+1}}{(2r+1)!}+\frac{a^{2r+3}}{(2r+2)!\,1!}+\frac{a^{2r+5}}{(2r+3)!\,2!}+\ldots\right\}.$$

We may now verify that these frequencies account for the whole of the characteristic function and hence that all frequencies have been found.

Conditions for a function to be a characteristic function

4.7 Any function which is not negative in its range of definition and which is integrable in the Stieltjes sense can be a frequency function; and any non-decreasing function which increases from 0 to 1 in its range of definition can be a distribution function. There are much more restrictive conditions to be obeyed before a given function can be a characteristic function.

In the first place, let us note that it is a necessary and sufficient condition for a function $\phi(t)$ to be a characteristic function that

$$F(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\phi(t)\frac{1-e^{-ixt}}{it}\,dt$$

shall (except for an additive constant $F(0)$) be a distribution function. This, however, is not a very helpful criterion in practice.

Looking to the definition of $\phi(t)$ as $\int_{-\infty}^{\infty} e^{itx}\,dF$, we see that necessary conditions for $\phi(t)$ to be a characteristic function are

(a) that $\phi(t)$ must be continuous in t,
(b) that $\phi(t)$ is defined in every finite t-interval,
(c) that $\phi(0) = 1$,
(d) that $\phi(t)$ and $\phi(-t)$ shall be conjugate quantities,
(e) that $|\phi(t)| \leqslant \int_{-\infty}^{\infty} |e^{itx}|\,dF \leqslant 1 = \phi(0)$.

4.8 A general theorem of Cramér (1937) states that for a bounded and continuous function $\phi(t)$ to be a c.f. it is necessary and sufficient that $\phi(0) = 1$ and that

$$\int_0^A \int_0^A \phi(t-u) \exp\{ix(t-u)\}\,dt\,du$$

is real and non-negative for all real x and all $A > 0$.

There are also some notable specialized theorems on the subject. For example Marcinkiewicz (1938) proved that $\exp P(t)$, where $P(t)$ is a polynomial in t, cannot be a c.f. unless $P(t)$ is of the first or second degree. Lukacs and Szász (1952) have given conditions under which a *rational* function can be a c.f.

Limiting properties of distribution and characteristic functions

4.9 We now proceed to discuss a set of results involving the tendency of d.f.'s and c.f.'s to limiting forms. Suppose there is given a sequence of distribution functions $F_n(x)$ depending on a parameter n which can increase indefinitely. To each F_n there will correspond a characteristic function ϕ_n. The type of question to be discussed is this : if F_n tends to a limit G, will ϕ_n tend to a limit ϕ and is ϕ the characteristic function of G ? Conversely, if ϕ_n tends to a limit ϕ, does F_n tend to a limit G and is G a distribution function having ϕ for its characteristic function ? The answers to these questions, as will be seen below, are affirmative under certain general conditions.

4.10 Let us first of all consider what is meant by a distribution function tending to another. If both are continuous, $F_n(x)$ is said to tend to $G(x)$ if, given any ε, there is an n_0 such that $|F_n(x) - G(x)| < \varepsilon$ for all $n > n_0$ and for all x. If there are discontinuities present, F_n will be said to tend to G if it does so in every point of continuity of G. Since by definition our functions are taken to be continuous on the right at saltuses, this evidently conforms to the definition for the continuous case and to the commonsense requirements of the situation.

4.11 We require a number of results concerning the convergence of d.f.'s. The first is that if a sequence of d.f.'s $\{F_n(x)\}$ converges to the d.f. $G(x)$ at all its points of continuity, then the convergence is uniform in every closed interval of continuity of $G(x)$.

Let $G(x)$ be continuous in the range a to b. Divide the range into a finite number of parts, say at $a = \xi_1, \xi_2, \ldots, \xi_h = b$, such that the increase of $G(x)$ in each interval is at most ε. Choose n_0 so large that $|F_n(\xi_j) - G(\xi_j)| < \varepsilon$, $j = 1, \ldots, h$, $n > n_0$. Then

for such an n and any x in the interval (a, b) there is an r such that x lies in (ξ_r, ξ_{r+1}) and

$$G(x)-2\varepsilon \leqslant G(\xi_r)-\varepsilon < F_n(\xi_r) \leqslant F_n(x) \leqslant F_n(\xi_{r+1}) < G(\xi_{r+1})+\varepsilon \leqslant G(x)+2\varepsilon \qquad (4.19)$$

and hence

$$|F_n(x)-G(x)|<2\varepsilon.$$

This is true for all x in (a, b) and hence the theorem is established.

In particular, if $G(x)$ is continuous the convergence is uniform throughout the whole interval $-\infty \leqslant x \leqslant \infty$.

4.12 The second theorem we require (the Montel-Helly theorem) is that if the sequence $\{F_n(x)\}$ is monotonic and bounded for all x (which is so for distribution functions) then we can pick out a subsequence $\{F_{n'}(x)\}$ which converges to some monotonic increasing function G (not necessarily a distribution function itself, for it may not vary from 0 to 1).

Consider first of all a series of values $x_1, x_2, \ldots$ It is known that every bounded set of numbers contains a convergent sequence. Hence we can pick out from the sequence $\{F_n(x_1)\}$ a convergent sequence, say $\{F_{n_1}(x_1)\}$. Then from the subsequence $\{F_{n_1}(x_2)\}$ we can pick out a subsequence $\{F_{n_2}(x_2)\}$ and $\{F_{n_2}(x)\}$ is thus convergent at both x_1 and x_2. Continuing in this way we may, by picking out the first function in $\{F_{n_1}(x)\}$, the second in $\{F_{n_2}(x)\}$, and so on, arrive at a sequence of functions $G_1(x)$, $G_2(x)\ldots$ which converges at each of the values $x_1, x_2, \ldots$, etc. This is the so-called Weierstrassian diagonal process.

It follows that the sequence G_n is convergent at every *rational* point x. Since $G_n(a) \leqslant G_n(x) \leqslant G_n(b)$ for every x between a and b, we see that if $G_n(a)$ and $G_n(b)$ converge, the limiting values of $G_n(x)$ lie between those limits, say $G(a)$ and $G(b)$.

Then the function $u(x)$ = upper bound of $G_n(x)$ (x not necessarily rational) is well defined and non-decreasing and so has no more than an enumerable number of points of discontinuity. If u is continuous at x, we take y and z such that $y<x<z$ and $u(z)-u(y)<\varepsilon$. Then if a and b are rational points such that $y<a<x<b<z$ it follows that $u(y) \leqslant G(a) \leqslant G(b) \leqslant u(z)$. Moreover, as all the limiting values of $G_n(x)$ are between $G(a)$ and $G(b)$, they are between $u(y)$ and $u(z)$. Hence, as ε can be arbitrarily small, we see that $G(x)$ tends to $u(x)$ at every point of continuity of u. Finally, by the diagonal process, we can select a sequence which will also be convergent at the points of discontinuity of $u(x)$. The theorem is established.

4.13 It is not obvious that if a sequence of functions $\{F_n(x)\}$ all lie between 0 and 1 then their limit $G(x)$ can attain those values. A counter-example is furnished by the distribution

$$\left.\begin{aligned} F_n(x) &= 0, && x<-n, \\ &= \tfrac{1}{2}, && -n \leqslant x \leqslant n, \\ &= 1, && x>n. \end{aligned}\right\} \qquad (4.20)$$

The limit of $F_n(x)$ here is $\frac{1}{2}$, $-\infty \leqslant x \leqslant \infty$. A condition which is sufficient (but not necessary) has been given by Kendall and Rao (1950). If $F_n(x)$ has a second moment $\mu'_{2(n)} \leqslant$ some finite A, all $n>n_0$, then $G(x)$ is a d.f.

For, from the Bienaymé-Tchebycheff inequality (3.35),

$$F_n(-X)+1-F_n(X)\leqslant A/X^2, \qquad n>n_0,$$

and hence, for sufficiently large X,

$$F_n(-X)+1-F_n(X)<\varepsilon$$

and since this is so uniformly in X

$$G(-X)+1-G(X)\leqslant\varepsilon \tag{4.21}$$

provided that (as we can ensure without loss of generality) $G(x)$ is continuous at $\pm X$. Thus $G(-\infty)=0$ and $G(\infty)=1$.

The First Limit Theorem

4.14 We now prove the theorem: if a sequence of distribution functions $\{F_n\}$ tends to a continuous distribution function G, then the corresponding sequence of characteristic functions ϕ_n tends to ϕ uniformly in any finite t-interval, where ϕ is the characteristic function of G.

It is required to prove that, given ε, there is an n_0 independent of t such that

$$|\phi(t)-\phi_n(t)| = \left|\int_{-\infty}^{\infty} e^{itx}(dG-dF_n)\right|<\varepsilon, \qquad n>n_0.$$

Select two points of continuity of G, X and $-X$. We can make X as large as we please. We then split the integral

$$\int_{-\infty}^{\infty} e^{itx}(dG-dF_n) \tag{4.22}$$

into two parts, that in the range $-X$ to $+X$ and that in the remaining portion of the range. Now

$$\left|\int_{x<-X}^{x>X} e^{itx}\,dG\right| \leqslant \int_{x<-X}^{x>X} dG \leqslant 1-G(X)+G(-X),$$

and by taking X large enough we can make this quantity less than $\varepsilon/6$.

Similarly

$$\left|\int_{x<-X}^{x>X} e^{itx}\,dF_n\right| \leqslant 1-F_n(X)+F_n(-X),$$

and since F_n tends to G (and that uniformly) this, for some large X, will be less than $\varepsilon/3$. Hence for some n_0 the portion of (4.22) outside the range $-X$ to $+X$ will be less in modulus than $\varepsilon/6+\varepsilon/3=\frac{1}{2}\varepsilon$. Consider now the other part

$$\int_{-X}^{X} e^{itx}(dG-dF_n). \tag{4.23}$$

This expression is the limit of the sum

$$\Sigma\, e^{itx_j}[\{G(\xi_{j+1})-G(\xi_j)\}-\{F_n(\xi_{j+1})-F_n(\xi_j)\}], \tag{4.24}$$

ξ_j, ξ_{j+1} being the boundaries of the interval into which the range is subdivided and x_j a value in that interval. The difference between this sum and the limiting value can be made less than $\frac{1}{2}\varepsilon$ if the intervals are small enough; for if they are less than η in width the difference of e^{itx_j} and $e^{it\xi_j}$ is less in modulus than $\eta|t|$, by the mean value

theorem, and thus in any t-range $\pm T$ the difference of (4.23) and (4.24) is less in modulus than

$$\eta T\,|\,\Sigma[\{G(\xi_{j+1})-G(\xi_j)\}-\{F_n(\xi_{j+1})-F_n(\xi_j)\}]\,|<2\eta T,$$

which is less than $\frac{1}{4}\varepsilon$ if $\eta<\varepsilon/8T$.

Now the sum (4.24) will itself be less than $\frac{1}{4}\varepsilon$ for some $n>n_0$, for it is the sum of a finite number of terms each of which tends to zero. Consequently (4.23) is less than $\frac{1}{2}\varepsilon$ and hence

$$|\,\phi(t)-\phi_n(t)\,|<\varepsilon, \qquad n>n_0.$$

Converse of the First Limit Theorem

4.15 The converse result is even more important.

Let $\{\phi_n\}$ be a sequence of characteristic functions corresponding to the sequence of distribution functions $\{F_n\}$. Then if $\phi_n(t)$ tends to $\phi(t)$ for all real t, and uniformly in a finite t-interval $|\,t\,|<a$,(*) $\{F_n\}$ tends to a distribution function G and ϕ is the characteristic function of G.

As a preliminary lemma, let us prove that if G is a distribution function with characteristic function ϕ, then for all real ξ and all $h>0$

$$\frac{1}{h}\int_{\xi}^{\xi+h}G(u)\,du-\frac{1}{h}\int_{\xi-h}^{\xi}G(u)\,du=\frac{1}{\pi}\int_{-\infty}^{\infty}\left(\frac{\sin t}{t}\right)^2 e^{-\frac{2it\xi}{h}}\,\phi\left(\frac{2t}{h}\right)dt. \qquad (4.25)$$

In fact, put

$$H(x)=\frac{1}{h}\int_{x}^{x+h}G(u)\,du.$$

This is a continuous distribution function and its characteristic function is

$$\int_{-\infty}^{\infty}e^{itx}\,dH=\frac{1}{h}\int_{-\infty}^{\infty}e^{itx}\,\{G(x+h)-G(x)\}\,dx,$$

which by a partial integration becomes

$$\frac{1}{h}\left[\frac{\{G(x+h)-G(x)\}\,e^{itx}}{it}\right]_{-\infty}^{\infty}-\frac{1}{ith}\int_{-\infty}^{\infty}e^{itx}\,\{dG(x+h)-dG(x)\}$$

$$=-\frac{1}{ith}\int_{-\infty}^{\infty}\{e^{it(x-h)}\,dG(x)-e^{itx}\,dG(x)\}$$

$$=\phi(t)\,\frac{1-e^{-ith}}{ith}.$$

Substituting for $H(x)$ in (4.4) we get

$$\frac{1}{h}\int_{\xi+h}^{\xi+2h}G(u)\,du-\frac{1}{h}\int_{\xi}^{\xi+h}G(u)\,du$$

$$=\frac{1}{2\pi h}\int_{-\infty}^{\infty}\left(\frac{1-e^{-ith}}{it}\right)^2 e^{-it\xi}\,\phi(t)\,dt,$$

whence, writing ξ for $\xi+h$, we find, after a little re-arrangement, equation (4.25).

Reverting now to the theorem required to be proved, note that it is sufficient to

(*) Or equivalently, if $\phi(t)$ is continuous at $t=0$ or if $\phi(t)$ is a characteristic function.

establish that if $\phi_n \to \phi$ uniformly in *some* interval $|t| < a$, then $\{F_n\}$ tends to some distribution function G in every point of continuity of G. When this is established it follows from the First Limit Theorem that ϕ is the characteristic function of G and that ϕ_n converges to ϕ uniformly in *every* finite t-interval.

As shown in **4.12**, given a sequence $\{F_n\}$ we may always choose from it a subsequence $\{F_{n'}\}$ such that $\{F_{n'}\}$ converges to a non-decreasing function G in every continuity point of G.

Let us then choose such a sequence. We have of necessity $0 \leqslant G \leqslant 1$, and G may be supposed everywhere continuous on the right. It is then a distribution function if $G(+\infty) - G(-\infty) = 1$, and this we proceed to prove. From (4.25) with $\xi = 0$ we have

$$\frac{1}{h}\int_0^h F_{n'}(u)\,du - \frac{1}{h}\int_{-h}^0 F_{n'}(u)\,du = \frac{1}{\pi}\int_{-\infty}^{\infty}\left(\frac{\sin t}{t}\right)^2 \phi_{n'}\left(\frac{2t}{h}\right)dt.$$

By hypothesis ϕ_n tends uniformly to ϕ for $|t| < a$ and hence $\phi_{n'}$ does so, and it is easily seen that the integral on the right is uniformly convergent. Thus, given ε, we can find h_0 such that for $h > h_0$

$$\left(\frac{1}{h}\int_0^h - \frac{1}{h}\int_{-h}^0\right) G(u)\,du = \frac{1}{\pi}\int_{-\frac{1}{2}ah}^{\frac{1}{2}ah}\left(\frac{\sin t}{t}\right)^2 \phi\left(\frac{2t}{h}\right)dt + \eta,$$

where $|\eta| < \varepsilon$. Now let h tend to infinity. As G is a non-decreasing function the left-hand side tends to $G(+\infty) - G(-\infty)$. The right-hand side tends, in virtue of the uniformity of $\phi_{n'}$ and the consequent continuity of ϕ near $t = 0$, to

$$\frac{1}{\pi}\int_{-\infty}^{\infty}\left(\frac{\sin t}{t}\right)^2 dt,$$

which is equal to unity.

Hence G, the limit of the subsequence $\{F_{n'}\}$, is a distribution function whose characteristic function is ϕ.

But any sequence of ϕ_n tends to ϕ, in virtue of the uniformity of the convergence, and hence any convergent subsequence of $\{F_n\}$ tends to G. Consequently $\{F_n\}$ tends to G in every point of continuity of G and the theorem follows.

Example 4.6

In Example 3.5, it was shown in effect that the binomial distribution has the characteristic function

$$\phi(t) = (q + pe^{it})^n.$$

Now the frequency at $x = j$ is $\binom{n}{j} q^{n-j} p^j$. This is greater than the ordinate at $x = j+1$ if

$$\binom{n}{j} q^{n-j} p^j > \binom{n}{j+1} q^{n-j-1} p^{j+1}$$

or

$$j > pn - q.$$

For large n the maximum frequency will then be in the neighbourhood of $j = pn$, and is there

$$\binom{n}{pn} q^{qn} p^{pn}.$$

In virtue of Stirling's approximation to the factorial this approximates to

$$\frac{n^n e^{-n} \sqrt{(2\pi n)}\, q^{qn} p^{pn}}{(pn)^{pn} e^{-pn} \sqrt{(2\pi pn)}\,(qn)^{qn} e^{-qn} \sqrt{(2\pi qn)}} \sim \frac{1}{\sqrt{(2\pi pqn)}}$$

and therefore tends to zero.

Thus every frequency in the binomial distribution tends to zero and the distribution does not tend to any limiting distribution.

Suppose, however, that we express the distribution in standard measure. Putting $y = (x-\mu_1')/\sigma$ we have

$$\phi_x(t) = \int_{-\infty}^{\infty} e^{itx}\, dF(x) = \int_{-\infty}^{\infty} \exp\{it(\sigma y + \mu_1')\}\, dF(y)$$
$$= e^{it\mu_1'}\, \phi_y(\sigma t).$$

Hence $$\phi_y(t) = \exp(-it\mu_1'/\sigma)\, \phi_x(t/\sigma).$$

The effect on $\phi(t)$ of transferring to standard measure is then to replace t by t/σ and to multiply by $\exp(-it\mu_1'/\sigma)$.

For the binomial, $\mu_1' = np$, $\mu_2 = npq$, and thus the characteristic function of the binomial expressed in standard measure is

$$\exp\left\{-\frac{itnp}{(npq)^{\frac{1}{2}}}\right\}\left\{q + p\exp\frac{it}{(npq)^{\frac{1}{2}}}\right\}^n.$$

Thus

$$\log\phi = \frac{-itnp}{(npq)^{\frac{1}{2}}} + n\log\left\{1 + p\left(\exp\frac{it}{(npq)^{\frac{1}{2}}} - 1\right)\right\}$$
$$= \frac{-itnp}{(npq)^{\frac{1}{2}}} + n\log\left\{1 + p\frac{it}{(npq)^{\frac{1}{2}}} - \frac{pt^2}{2npq} + \frac{p\theta t^3}{6(npq)^{\frac{3}{2}}}\right\}, \qquad 0 \leqslant |\theta| \leqslant 1,$$
$$= n\left\{\frac{-pt^2}{2npq} - \frac{1}{2}\frac{(pit)^2}{npq}\right\} + O(t^3 n^{-\frac{1}{2}})$$
$$= -\tfrac{1}{2}t^2 + O(t^3 n^{-\frac{1}{2}}).$$

Thus for any finite t, $\log\phi$ tends uniformly to $-\frac{1}{2}t^2$ and hence

$$\phi(t) \longrightarrow e^{-\frac{1}{2}t^2}.$$

Thus the binomial distribution expressed in standard measure tends to the distribution whose characteristic function is $e^{-\frac{1}{2}t^2}$, i.e. to the form

$$dF = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2}\, dx, \qquad -\infty \leqslant x \leqslant \infty.$$

Multivariate characteristic functions

4.16 The characteristic function of a bivariate distribution $F(x_1, x_2)$ is defined as

$$\phi(t_1, t_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{it_1x_1 + it_2x_2}\, dF(x_1, x_2) \tag{4.26}$$

and generally, that of a multivariate distribution $F(x_1, x_2, \ldots, x_n)$ as

$$\phi(t_1, t_2, \ldots, t_n) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{it_1x_1+it_2x_2\ldots+it_nx_n}\, dF(x_1, x_2, \ldots, x_n), \tag{4.27}$$

If $x_1, x_2, \ldots, x_n$ are independent we have

$$\begin{aligned}\phi(t_1, t_2, \ldots, t_n) &= \int_{-\infty}^{\infty} e^{it_1x_1}\, dF_1(x_1)\int_{-\infty}^{\infty} e^{it_2x_2}\, dF_2(x_2)\ldots\int_{-\infty}^{\infty} e^{it_nx_n}\, dF_n(x_n)\\ &= \phi_1(t_1)\,\phi_2(t_2)\ldots\phi_n(t_n)\end{aligned} \tag{4.28}$$

where ϕ_r is the c.f. of F_r. Similarly

$$\psi(t_1, t_2, \ldots, t_n) = \sum_{j=1}^{n} \log\phi(t_j). \tag{4.29}$$

Thus the characteristic function of the joint distribution of a number of independent variables is the product of their characteristic functions; and the cumulative function is the sum of their cumulative functions. This is a fundamentally important result in the theory of sampling.

4.17 In generalization of (4.4) we have

$$F(x_1, x_2, \ldots, x_n) - F(0, 0, \ldots, 0) = \frac{1}{(2\pi)^n}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\frac{1-e^{ix_1t_1}}{it_1}\cdots\frac{1-e^{ix_nt_n}}{it_n}\,\phi(t_1, t_2, \ldots, t_n)\, dt_1\ldots dt_n. \tag{4.30}$$

The integrals are to be understood as principal values $\lim_{c\to\infty}\int_{-c}^{c}\cdots\int_{-c}^{c}$.

The proof is similar to that for the univariate case. For example, with two variates we have, in an obvious generalization of (4.10),

$$I_c = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-c}^{c}\int_{-c}^{c}\left\{\frac{\sin t_1u_1}{t_1} - \frac{\sin t_1(u_1-x_1)}{t_1}\right\}\left\{\frac{\sin t_2u_2}{t_2} - \frac{\sin t_2(u_2-x_2)}{t_2}\right\} dt_1\, dt_2\, dF(u_1, u_2)$$

and

$$\begin{aligned}\lim I_c &= \lim\int_{-\infty}^{\infty}\int_{-c}^{c}\left\{\frac{\sin t_1u_1}{t_1} - \frac{\sin t_1(u_1-x_1)}{t_1}\right\} dt_1(2\pi)\,\{dF(u_1, x_2) - dF(u_1, 0)\}\\ &= (2\pi)^2\,\{F(x_1, x_2) - F(0, 0)\},\end{aligned} \tag{4.31}$$

and so on. As in the univariate case, the c.f. defines the distribution uniquely and the corresponding analogues of the limit theorems also remain true. Thus the converse of the result of **4.16** is ensured by the unique correspondence of the c.f. and the d.f.: if a joint c.f. factorises, the variables are independent.

4.18 If we have a distribution $F(x)$ and some function of the variate such as $y\ (= y(x))$, we define the c.f. of y as

$$\phi_y(t) = \int_{-\infty}^{\infty}\exp(ity)\, dF(x). \tag{4.32}$$

The distribution of y is then given by inversion ; for instance, the frequency function of y, say $g(y)$, is given by

$$g(y) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-ity)\,\phi_y(t)\,dt, \tag{4.33}$$

and a similar result holds for a function g of several variates.

The problem of moments

4.19 It is of some interest to consider how far a set of moments (assuming that they all exist) determine a distribution uniquely. To give some point to the discussion let us note that in some circumstances it is possible for two different distributions to have the same set of moments.

Consider the integral

$$\int_0^{\infty} t^{p-1} e^{-qt}\,dt = \Gamma(p)/q^p, \qquad p>0, \qquad \mathscr{R}(q)>0.$$

Put $p = (n+1)/\lambda$, n a non-negative integer ; $0<\lambda<\frac{1}{2}$; $q = \alpha+i\beta$; $\beta/\alpha = \tan\lambda\pi$; $x^\lambda = t$. We find on substitution, since $(1+i\tan\lambda\pi)^{(n+1)/\lambda}$ is real, that the imaginary part of the integral is zero and thus that

$$\int_0^{\infty} x^n \exp(-\alpha x^\lambda)\sin(\beta x^\lambda)\,dx = 0. \tag{4.34}$$

Hence the distributions

$$f(x) = k\exp(-\alpha x^\lambda)\{1+\varepsilon\sin(\beta x^\lambda)\}, \quad 0\leqslant x\leqslant\infty ; \quad \alpha>0 ; \quad 0<\lambda<\tfrac{1}{2} ; \quad |\varepsilon|<1 ; \tag{4.35}$$

have the same moments for all ε in the range $|\varepsilon|<1$.

The range of the distribution is infinite in one direction. We may derive a similar family with range infinite in both directions by putting $p = (2n+1)/\rho$, $q = \alpha+i\beta$, $\beta/\alpha = \tan\frac{1}{2}\rho\pi$, $x^\rho = t$, $\rho = 2s/(s+1)$, s a positive integer. The family then becomes

$$f(x) = k\exp\{-\alpha|x|^\rho\}\{1+\varepsilon\cos(\alpha|x|^\rho)\}, \qquad -\infty\leqslant x\leqslant\infty. \tag{4.36}$$

4.20 The problem of moments in its full generality considers a set of constants $c_0, c_1, c_2, \ldots$ and inquires whether they can be the moments of a distribution. For statistical purposes this is not of particular interest.(*) We are more concerned with the problem : given that the set of constants are, in fact, the moments of a distribution, can any other distribution have the same set ?

Note in the first instance that this problem need only be considered when absolute moments of all orders exist. It is not difficult to see that more than one distribution can exist having a limited number of moments finite and the remainder infinite. Furthermore, if any moment of even order exists, those of lower order must exist. In particular, if μ'_{2r} exists the absolute moment of order $2r$ exists. Thus we consider only the case when all absolute moments exist.

4.21 We will prove in the first place the theorem that a set of moments determines

(*) For some results on the general problem, see Stieltjes (1918) and Hamburger (1920).

a distribution uniquely if the series $\sum_{j=0}^{\infty} \nu_j t^j/j!$ converges for some real non-zero t. We write ν without the prime in this and the following sections to save printing, but the result is true for moments about any origin.

The characteristic function is continuous in t and its derivatives exist at $t = 0$ if the moments exist. We have then in the neighbourhood of $t = 0$

$$\phi(t) = \sum_{j=0}^{r} (it)^j \mu'_j/j! + R_r, \tag{4.37}$$

where R_r is less in absolute value than $\dfrac{\nu_{r+1}|t|^{r+1}}{(r+1)!}$

Thus if $\Sigma \nu_j t^j/j!$ converges, $\nu_j t^j/j!$ tends to zero and hence $\phi(t)$ is equal to the sum of the infinite series $\sum_{j=0}^{\infty} (it)^j \mu'_j/j!$ if it exists. Moreover, this series is majorated by $\Sigma \nu_j t^j/j!$ and hence is absolutely convergent if the latter is convergent. Hence we have

$$\phi(t) = \sum_{j=0}^{\infty} (it)^j \mu'_j/j! \tag{4.38}$$

and thus $\phi(t)$ is uniquely determined in the neighbourhood of $t = 0$. In the neighbourhood of $t = t_0$ we have

$$\phi(t) = \sum_{j=0}^{r} \left\{ \frac{i^j (t-t_0)^j}{j!} \int_{-\infty}^{\infty} x^j e^{it_0 x} dF \right\} + R'_r$$

and the modulus of the coefficient of $(t-t_0)^j/j!$ is not greater than ν_j. Consequently $\phi(t)$ can be expanded everywhere as a convergent Taylor series and is equal to the sum of that series. Hence $\phi(t)$ may be extended from the neighbourhood $t = t_0$ by analytic continuation through any finite t-interval. Hence $\phi(t)$ is everywhere uniquely defined.

But $\phi(t)$ determines the distribution function and hence the latter is uniquely determined.

4.22 A few simple but effective results follow as corollaries.

(a) The moments uniquely determine the distribution if the upper limit of $(\nu_n^{1/n})/n$ is finite. For the series whose general term is $\nu_n t^n/n!$ is convergent if

$$\limsup (\nu_n t^n/n!)^{1/n} < 1.$$

Replacing the factorial by its Stirling approximation we see that this will be true if

$$\limsup (\nu_n^{1/n})/n < k/t$$

where k is some constant. If the upper limit is finite the inequality can be satisfied for some non-zero t.

(b) It is also sufficient for the upper limit of $(\mu'_{2n}{}^{1/2n})/2n$ to be finite, a form which enables us to disregard the absolute moments. It is, in fact, easy to see that this upper limit and the upper limit of $(\nu_n^{1/n})/n$ are finite or infinite together.

(c) The moments uniquely determine the distribution if the range is finite. For, taking an origin at the start of the distribution (of range h, say), we see that $\nu_r \leqslant h^r$ and hence $(\nu_n^{1/n})/n \leqslant h/n$, which tends to zero.

4.23 Two further criteria may be mentioned. The first is due to Carleman (1925). A set of moments determines a distribution uniquely if (in the case of limits $-\infty$ to $+\infty$)

$$\sum_{j=0}^{\infty} \frac{1}{(\mu_{2j})^{1/2j}} \tag{4.39}$$

diverges. For the limits 0 to ∞ the corresponding series is

$$\sum_{j=0}^{\infty} \frac{1}{(\mu_j)^{1/2j}}. \tag{4.40}$$

Secondly, if there exists a frequency function, the moments determine it uniquely if, for limits $-\infty$ to $+\infty$,

$$f(x) < M|x|^{\beta-1}\exp(-\alpha|x|^{\lambda}) \text{ for } |x| > x_0, \qquad M, \beta, \alpha, > 0, \ \lambda \geqslant 1, \tag{4.41}$$

and for limits 0 to ∞,

$$f(x) < M|x|^{\beta-1}\exp(-\alpha|x|^{\lambda}) \text{ for } |x| > x_0, \qquad M, \beta, \alpha, > 0, \ \lambda \geqslant \tfrac{1}{2}. \tag{4.42}$$

This result is due ultimately to Stieltjes. It follows without difficulty from the Carleman criterion.

It is interesting to note that if for some x_0

$$f(x) > \exp(-\alpha|x|^{\lambda}), \qquad \alpha > 0, \ x > x_0, \tag{4.43}$$

then the problem of moments is necessarily indeterminate (as usual, $\lambda < \frac{1}{2}$ for the range 0 to ∞ and $\lambda < 1$ for the range $-\infty$ to $+\infty$). This follows from the examples in equations (4.35) and (4.36), for we can add to (4.43), *without rendering any frequency negative*, a function all of whose moments are zero.

Example 4.7

The moments of the distribution

$$dF = \frac{1}{\sigma\sqrt{(2\pi)}}\exp(-x^2/2\sigma^2), \qquad -\infty \leqslant x \leqslant \infty,$$

are given (cf. Example 3.4) by

$$\mu_{2r+1} = 0, \qquad \mu_{2r} = \frac{(2r)!}{2^r r!}\sigma^{2r}.$$

Thus the upper limit of $(\mu_{2n}{}^{1/2n})/2n$ is, from the Stirling approximation to the factorial, asymptotically equivalent to

$$\frac{\sigma}{2n\sqrt{2}}\left[\frac{e^{-2n}(2n)^{2n}\sqrt{(4\pi n)}}{\{e^{-n}n^n\sqrt{(2\pi n)}\}}\right]^{1/2n} \sim \frac{\sigma}{(2en)^{\frac{1}{2}}}.$$

The upper limit is then zero and the distribution is uniquely determined by its moments.

4.24 Cramér and Wold (1936) have extended Carleman's criterion to multivariate distributions. We need only consider the moments of the marginal distributions, $\mu'_{r00\ldots}$, $\mu'_{0r0\ldots}$, $\mu'_{00r\ldots}$, etc. If the sum of these for all variates is λ_r then the criteria of (4.39) and (4.40) remain with λ_r instead of μ_r; for example, a distribution ranging from $-\infty$ to ∞ is completely determined by its moments if

$$\sum_{j=0}^{\infty} \frac{1}{(\lambda_{2j})^{1/2j}} \tag{4.44}$$

diverges.

4.25 If a moment of order r exists it must be given by the rth derivative of the c.f. at $t = 0$. Thus, if $\phi(t)$ can be expanded as an infinite Taylor series, that series can only be $\Sigma (it)^j \mu'_j/j!$. And if this series does not converge, $\phi(t)$ cannot be so expanded. But it can always be expanded asymptotically in the finite form with remainder

$$\phi(t) = \sum_{j=0}^{r} (it)^j \mu'_j/j! + R.$$

This illustrates the source of a difficulty in discussing limiting properties when the infinite series does not converge, for it is known that there exist an infinite number of functions which have a given set of coefficients in an asymptotic expansion. For instance, if $\alpha(t)$ has an asymptotic expansion in t the functions $\alpha(t) + kt^{-\log t}$ all have the same expansion. It is therefore hardly surprising that when $\Sigma (it)^j \mu'_j/j!$ fails to converge, there may be more than one frequency or distribution function with the same set of moments.

But it does not follow from what has been said that there *must* be more than one frequency-distribution. There must be more than one function, but those functions may not qualify as frequency-distributions, e.g. they may be negative in part of the range. In the example just given, $t^{-\log t}$ cannot be a characteristic function, for it does not obey the well-known condition that $\phi(t)$ and $\phi(-t)$ should be conjugate.

4.26 We now proceed to what is known as the Second Limit Theorem, which is concerned with the way in which a sequence of distribution functions $\{F_n(x)\}$ tends to a limit if the corresponding sequence of moments of order j, say $\mu_j(n)$, tends to a limit μ_j. (To save printing, we shall throughout write the moments without primes.) Our method of proof is due to Kendall and Rao (1950).

We require in the first place a result concerning the expansion of a c.f. which is rather more precise than those which have served our purpose so far. If μ_{2m} exists then the c.f. can be expressed in the form

$$\phi(t) = \sum_{j=0}^{2m-1} (it)^j \mu_j/j! + \rho (it)^{2m} \mu_{2m}/(2m)! \tag{4.45}$$

for all real t, ρ being such that $|\rho| < 1$. Also $\lim_{t \to 0} \rho = 1$.

For $m = 0$ this is true trivially. For $m > 0$ we have for real x, t,

$$e^{ixt} = \sum_{j=1}^{n} \frac{(ixt)^j}{j!} + \rho' \frac{(ixt)^{n+1}}{(n+1)!}, \qquad |\rho'| \leqslant 1, \tag{4.46}$$

and on substitution in the integral defining $\phi(t)$ we get, for $n = 2m-1$, equation (4.45). For t, μ_{2m} zero we may define $\rho = 1$; for other values, considering the difference of (4.46) taken to n and $n+1$ terms we find

$$|\rho' - 1| \leqslant |xt|/(n+2). \tag{4.47}$$

Since

$$|\rho - 1| \mu_{2m} = \int_{-\infty}^{\infty} (\rho' - 1) x^{2m} dF$$

we have

$$|\rho - 1| \mu_{2m} < \varepsilon + \int_{-X}^{X} |\rho' - 1| x^{2m} dF,$$

for any given ε, if X is large enough (for otherwise μ_{2m} would not converge). But for fixed X the integral can be made as small as we please by taking t sufficiently small, in virtue of (4.47). Thus we have

$$|\rho - 1|\mu_{2m} < 2\varepsilon$$

and thus $\rho \to 1$ when $t \to 0$.

4.27 A second result we require (a converse of the previous one) is as follows: If for some positive integer m

$$\phi(t) = \sum_{j=0}^{2m} (it)^j \lambda_j / j! + o(t^{2m}) \tag{4.48}$$

as t tends to zero through real values, then the first $2m$ moments exist and $\mu_r = \lambda_r$, $r = 0, 1, \ldots, 2m$.

Let δ^2 denote the operation of taking central differences with interval $2h$, i.e.

$$\delta^2 A(t) = A(t+2h) - 2A(t) + A(t-2h).$$

Then if the remainder term in (4.48) is $t^{2m}B(t)$, where $B(t)$ tends to zero as $t \to \infty$, a little calculation shows that for $t = 0$

$$\left(\frac{\delta^2}{4h^2}\right)^m t^{2m} B(t) = \sum_{j=0}^{2m} \binom{2m}{j} (-1)^j (m-j)^{2m} B(2mh - 2jh).$$

Thus for $|B(t)| < \varepsilon$

$$\left(\frac{\delta^2}{4h^2}\right)^m t^{2m} B(t) \leqslant \sum_{j=0}^{2m} \binom{2m}{j} |m-j|^{2m} \varepsilon \tag{4.49}$$

when $|h| < \eta/2m$ and $|t| \leqslant \eta$. Hence if (4.48) is true

$$\left[\lim_{h \to 0} \left(\frac{\delta^2}{4h^2}\right)^m t^{2m} B(t)\right]_{t=0} = 0. \tag{4.50}$$

But $\lim_{h \to 0} \left(\frac{\delta^2}{4h^2}\right)^m$, $= D^{2m}$ say, acting on a polynomial is easily seen to be the $(2m)$th derivative. Thus applying the operation to (4.48) we find, remembering (4.50), that when $t = 0$

$$D^{2m}\phi(t) = (-1)^m \lambda_{2m}. \tag{4.51}$$

Now

$$\delta^2 e^{ixt} = -(2 \sin xh)^2 e^{ixt}$$

and thus

$$\lim_{h \to 0} \int_{-\infty}^{\infty} \left(\frac{\sin xh}{xh}\right)^{2m} x^{2m}\, dF = \lambda_{2m}. \tag{4.52}$$

From the uniform convergence to unity of $(\sin xh/xh)^{2m}$ we see that for any finite interval (a, b)

$$\int_a^b x^{2m}\, dF \leqslant \lambda_{2m},$$

and thus the moment of order $2m$ must be finite. Hence from (4.45), as t tends to zero,

$$\phi(t) = \sum_{j=0}^{2m-1} (it)^j \mu_j / j! + \frac{(it)^{2m} \mu_{2m}}{(2m)!}[1 + o(1)]. \tag{4.53}$$

Comparing this with (4.48) we see that

$$\mu_r = \lambda_r, \qquad r = 0, 1, \ldots, 2m.$$

The result of this section implies that if the $(2m)$th derivative of the c.f. exists at $t = 0$, the $(2m)$th moment also exists. We have seen in **3.5** that if the moments exist the corresponding derivative of the c.f. will exist at $t = 0$, a statement of which the above is the converse for *even* moments. For odd moments the existence of the derivative at $t = 0$ is not sufficient for the existence of the corresponding moment. Pitman (1956, *Ann. Math. Statist.*, **27**, 1156) has shown that necessary and sufficient conditions for the existence of the derivative are (1) that the principal value of μ_r' exists and (2) that

$$\lim_{x\to\infty} x^r \{F(-x) + 1 - F(x)\} = 0.$$

4.28 The results of the two preceding sections are necessary to establish the uniqueness of asymptotic expansions of the c.f. under certain conditions. We now prove some theorems on sequences of distributions and moments.

Let $F_n(x)$, the nth member of a sequence $\{F_n(x)\}$, possess a finite moment of the jth order, say $\mu_j(n)$, for all $n > n_0$. Let

$$\lim_{n\to\infty} \mu_j(n) = \lambda_j$$

for every value of j; and let $\{F_n(x)\}$ converge to a limit function $G(x)$ at all its points of continuity. ($G(x)$ is necessarily bounded, monotonically increasing and may be taken to be continuous on the right.) Then $G(x)$ is a distribution function possessing moments of all orders; $\{\lambda_j\}$ is a sequence of moments; and λ_j is the jth moment of $G(x)$.

That $G(x)$ is a d.f. follows from **4.13**, for the second moments $\mu_2(n)$ form a convergent and therefore bounded sequence. Hence, by the First Limit Theorem of **4.14**, the sequence of c.f.'s $\{\phi_n(t)\}$ converges to the c.f. of $G(x)$, say $\phi(t)$.

Now when $n > n_0$

$$\phi_n(t) = \sum_{j=0}^{2m-1} (it)^j \mu_j(n)/j! + \rho(it)^{2m} \mu_{2m}(n)/(2m)! \tag{4.54}$$

and the last term on the right must approach a limit since all the other terms do so. Thus

$$\phi(t) = \sum_{j=0}^{2m-1} (it)^j \lambda_j/j! + R(it)^{2m} \lambda_{2m}/(2m)! \quad |R| \leqslant 1. \tag{4.55}$$

This is true for all m and by comparing each of these formulae with the next following one we have, as in **4.26**, $R \to 1$ as $t \to 0$. From the results of **4.27** the theorem now follows.

Example 4.8

Note that the result of **4.28** follows even if $F_n(x)$ does not possess moments of all orders. For example, the distribution

$$dF = \frac{k\,dt}{(1 + t^2/\nu)^{\frac{1}{2}(\nu+1)}}, \qquad -\infty \leqslant t \leqslant \infty\,;\ \nu > 1,$$

possesses moments only up to and including $\mu_{\nu-1}$. In fact (cf. Example 3.3)

$$\mu_{2r} = \frac{\Gamma(\frac{1}{2}\nu - r)\,\Gamma(r+\frac{1}{2})}{\Gamma(\frac{1}{2}\nu)\,\Gamma(\frac{1}{2})}\,\nu^r. \tag{4.56}$$

By use of the Stirling approximation to the Gamma-function we find that as $\nu \to \infty$

$$\mu_{2r} \to \frac{(2r)!}{2^r r!}, \tag{4.57}$$

namely to the moments of the standardized normal distribution (Example 3.4), all of which exist. If, then, our limit function converges it must do so to the normal form, i.e. to a form for which all moments exist.

It is, in fact, easy to see directly that the frequency function tends to $k \exp(-\frac{1}{2}t^2)$.

The Second Limit Theorem

4.29 Let $\{F_n(x)\}$ converge to $G(x)$; let $\mu_j(n)$ exist for $n > n_0$ and for all $j \geqslant 0$; and let $\mu_j(n)$ be bounded above by some constant A_j. Then all the moments λ_j of $G(x)$ exist and $\mu_j(n) \to \lambda_j$ as $n \to \infty$.

From the First Limit Theorem we know that the sequence of c.f.'s $\{\phi_n(t)\}$ converges to $\phi(t)$, the c.f. of $G(x)$. Also

$$\phi_n(t) = 1 + it\,\mu_1(n) + O(t^2), \qquad n > n_0,$$

where the constant implied by $O(t^2)$ can be taken as independent of n and t. Thus if m is some other value $> n_0$,

$$|\mu_1(n) - \mu_1(m)| \leqslant A_1 |t| + \delta_{m,n}$$

where $\delta_{m,n} \to 0$ as m, n tend independently to infinity. This, being true for every non-zero t, implies that

$$\lim_{m,\,n \to \infty} |\mu_1(n) - \mu_1(m)| = 0$$

and hence there exists a constant λ_1 such that

$$\lim \mu_1(n) = \lambda_1. \tag{4.58}$$

We now proceed inductively by the same kind of argument to establish the limit of $\mu_2(n)$. We require

$$\phi_n(t) = 1 + it\,\mu_1(n) + (it)^2 \mu_2(n)/2! + O(t^3),$$

an expression which follows from (4.54) because the sixth moment is bounded. We then find

$$|\mu_2(n) - \mu_2(m)| \leqslant \tfrac{2}{3} A_6^{\frac{1}{3}} |t| + \delta'_{m,n}$$

and, as before, this implies the limit:

$$\lim \mu_2(n) = \lambda_2. \tag{4.59}$$

Thus we establish that the moment-sequence $\{\lambda_j\}$ exists. It then follows from the theorem of **4.28** that these moments are, in fact, the moments of $G(x)$.

4.30 For most purposes we require a converse form of the Second Limit Theorem: let the moments $\mu_j(n)$ exist and the limits $\lim \mu_j(n) = \lambda_j$ be the moments of a distribution $G(x)$ *which is uniquely determined by its moments.* Then $\{F_n(x)\}$ converges to $G(x)$ in all points of continuity.

We can always find a sequence in $\{F_n(x)\}$ which converges to a distribution function. It then follows from the theorem of **4.28** that this function must be the unique function having the moments λ_j and thus it is $G(x)$.

Suppose now that at some point of continuity of x, say a,

$$\alpha \equiv \limsup F_n(a) \neq G(a).$$

We can choose a subsequence converging to α at $x = a$; and a further sub-subsequence which converges both to α and also to $G(a)$, because $G(x)$ is continuous at $x = a$. This contradiction can only be avoided if

$$\limsup F_n(a) = G(a);$$

and by a similar argument

$$\liminf F_n(a) = G(a).$$

The converse theorem is thus established.

Example 4.9

The discontinuous distribution whose frequency at $x = j$ ($j = 0, 1, \ldots$) is $e^{-\lambda}\lambda^j/j!$ has a characteristic function (cf. Example 3.10)

$$\phi(t) = \exp\{\lambda(e^{it} - 1)\},$$

and hence all cumulants equal to λ.

The distribution is evidently the only one with such cumulants, for $\Sigma \kappa_j (it)^j/j!$ is convergent and equals $\lambda(e^{it} - 1)$, so that the cumulative function and the characteristic function are uniquely determined.

Now as λ tends to infinity the frequency at x_j, $e^{-\lambda}\lambda^j/j!$, tends to zero and thus the distribution does not tend to a limit. This is consistent with the behaviour of the cumulants, which increase without limit.

Suppose, however, we express the distribution in standard measure. Then

$$\kappa_1 = 0, \qquad \kappa_r = \lambda/\kappa_2^{\frac{1}{2}r} = \lambda^{-(\frac{1}{2}r-1)}, \qquad r \geqslant 2.$$

Hence as $\lambda \to \infty$ all cumulants higher than the second tend to zero, and the cumulants of the distribution tend to those of the "normal distribution"

$$dF = \frac{1}{\sqrt{(2\pi)}} \exp(-\tfrac{1}{2}x^2)\, dx, \qquad -\infty \leqslant x \leqslant \infty.$$

Now we know that this distribution is completely determined by its moments (Example 4.7). We also know that the cumulants determine the moments and vice versa, so that if the cumulants of the discontinuous distribution tend to those of the normal distribution, the moments will tend to the moments of that distribution. Hence the converse form of the Second Limit Theorem is applicable, and the discontinuous distribution does in fact tend to the normal form *when expressed in standard measure.*

4.31 In connection with the dual role of the characteristic function and the distribution function, it is worth remarking that the behaviour of one in the neighbourhood of the origin is related to the behaviour of the other at infinity. In fact, the mth derivative of $\phi(t)$ at $t = 0$ is the mth moment, the existence of which depends on the behaviour of $x^m f(x)$ at infinity. Conversely, from (4.5) we see that the mth derivative of $f(x)$, if it

exists, is given by $\int_{-\infty}^{\infty} (-it)^m \phi(t) e^{-ixt}\, dt/2\pi$, which in modulus is not greater than $\int_{-\infty}^{\infty} t^m \phi(t)\, dt/2\pi$, which depends for its convergence on the behaviour of $t^m \phi(t)$ at infinity.

4.32 It is natural to inquire whether there is some other transform of the distribution function which enjoys properties similar to those of the c.f. The answer appears to be in the negative. Lukacs (1952) has proved a theorem to the following effect: Let $\kappa(x, t)$ be a complex function defined for all real x, t, bounded and measurable in t. Define

$$\phi(t) = \int_{-\infty}^{\infty} \kappa(x, t)\, dF(x);$$

and let the two following conditions hold:

(1) $\phi_1(t) \equiv \phi_2(t)$ if and only if $F_1(x) = F_2(x)$.

(2) If $F(x) = \int_{-\infty}^{\infty} F_1(x-y)\, dF_2(y)$, then $\phi(t) = \phi_1(t)\phi_2(t)$.

Then $\kappa(x, t)$ must have the form $\exp\{it A(x)\}$ where $A(x)$ is a real-valued function which assumes all values of a set which is dense on a real line. It is very remarkable that we do not require among the conditions either an inversion theorem or limit theorems. The Fourier transform seems, as it were, to present us gratuitously with these useful and desirable properties.

EXERCISES

4.1 Show that if a frequency function $f(x)$ is symmetrical the characteristic function, apart from a term involving the mean, is an even function, i.e. $\phi(t) = \phi(-t)$, and that therefore $\phi(t)$ is real; and conversely, if $\phi(t)$ is real the frequency function, if any, is symmetrical.

4.2 Show that the characteristic function

$$\phi(t) = \left(\frac{e^{it}-1}{it}\right)^n, \quad n \text{ a positive integer,}$$

is that of the frequency function

$$f(x) = \frac{1}{(n-1)!} \sum_{j=0}^{[x]} (-1)^j \binom{n}{j} (x-j)^{n-1}$$

4.3 Show that the distribution whose c.f. is $1/(1+t^2)$ has frequency function $\frac{1}{2}e^{-|x|}$, $-\infty \leqslant x \leqslant \infty$, and that the distribution whose c.f. is $\cos \frac{1}{2}\pi t/(1-t^2)$ has f.f. $\frac{1}{2}\cos x$, $-\frac{1}{2}\pi \leqslant x \leqslant \frac{1}{2}\pi$.

4.4 If for a certain distribution

$$\kappa_r = ba^r,$$

a and b being positive constants, show that the distribution is discontinuous with variate-values $0, a, \ldots, ra, \ldots$ and the frequency at ra equal to $e^{-b} b^r/r!$.

4.5 Show that the function $\exp(-t^\alpha)$ cannot be a characteristic function unless $\alpha = 2$.

4.6 Show that there is only one distribution with moments given by

$$\mu'_r = \Gamma(\nu+r)/\Gamma(\nu)$$

and that it is

$$dF = \frac{1}{\Gamma(\nu)} e^{-x} x^{\nu-1}\, dx, \quad 0 \leqslant x \leqslant \infty.$$

4.7 Show that the distribution

$$dF = \frac{\sqrt{2}}{\pi}\frac{dx}{1+x^4}, \qquad -\infty \leqslant x \leqslant \infty,$$

has a c.f.

$$\phi(t) = \sqrt{2}\exp(-|t|/\sqrt{2})\sin(|t|/\sqrt{2}+\tfrac{1}{4}\pi).$$

4.8 Show that the distribution

$$dF = \frac{nb^{2n-1}\sin(\pi/2n)}{\pi(b^{2n}+x^{2n})}\,dx, \qquad -\infty \leqslant x \leqslant \infty\ ;\ b>0\ ;\ n \text{ integral},$$

has a c.f.

$$\phi(t) = \sum_{s=0}^{n-1}\exp\left(-b|t|\sin\frac{2s+1}{2n}\pi\right)\sin\left(\frac{2s+1}{2n}\pi + b|t|\cos\frac{2s+1}{2n}\pi\right)\sin(\pi/2n).$$

4.9 Show that the distribution

$$dF = \frac{dx}{\cosh \pi x}, \qquad -\infty \leqslant x \leqslant \infty,$$

has a c.f.

$$\phi(t) = \operatorname{sech} \tfrac{1}{2}t.$$

4.10 Show that $(1+t^{2k})^{-1}$ cannot be a characteristic function for $k>1$; and also that $(1-t^2)^{-1}$ cannot be a c.f.

4.11 Using the theorem of Marcinkiewicz referred to in 4.8 show that, if all cumulants κ_r vanish for r greater than some $r_0>2$, then all cumulants vanish for $r>2$; and hence that the distribution is normal.

4.12 Show that the distribution

$$F_n(x) = \left(1-\frac{1}{n}\right)\frac{1}{\sqrt{(2\pi)}}\int_{-\infty}^{x} e^{-\frac{1}{2}u^2}\,du + \frac{1}{2n}\{1+\operatorname{sgn}(x-n)\}$$

tends to normality as $n \longrightarrow \infty$ but that the higher moments tend to infinity.

(Rao and Kendall, 1950)

4.13 A theorem due to Weierstrass states that any function continuous in the range (a, b) can be represented by a uniformly convergent series of polynomials $\sum_{n=0}^{\infty} P_n(x)$, $P_n(x)$ being of degree n in x. Deduce that if two continuous frequency functions, f_1 and f_2, have the same moments of all orders,

$$\int_a^b (f_1-f_2)^2\,dx = 0,$$

and hence that the moments determine a distribution uniquely if it is continuous and of finite range.

4.14 If θ is a non-negative function of the variate x and

$$\alpha(t) = \int_{-\infty}^{\infty} \theta^t\,dF(x),$$

show that the frequency function of θ, if any, is given by

$$f(\theta) = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} \theta^{-t-1}\alpha(t)\,dt.$$

4.15 Show that if a characteristic function $\phi(t)$ possesses derivatives up to and including the second order, then

$$\left|\left(\frac{d\phi}{dt}\right)_{t=0}^2\right| \leqslant \left|\left(\frac{d^2\phi}{dt^2}\right)_{t=0}\right|$$

and generalize this result.

4.16 λ is a random variable with c.f. equal to $\phi(t)$. Another variable y is conditionally distributed in the (Poisson) form of Exercise 3.1 with parameter equal to the value of λ. Show that the unconditional c.f. of y is

$$\phi\{(e^{it}-1)/i\}.$$

4.17 Show that the distribution

$$dF = \frac{dx_1\,dx_2}{2\pi(1-\rho^2)^{\frac{1}{2}}}\exp\left\{\frac{-1}{2(1-\rho^2)}(x_1^2-2\rho x_1x_2+x_2^2)\right\}, \qquad -\infty \leqslant x_1, x_2 \leqslant \infty,$$

is uniquely determined by its moments.

4.18 If a variate is distributed in the normal form

$$dF = \frac{1}{\sigma\sqrt{(2\pi)}}e^{-\frac{1}{2}x^2/\sigma^2}\,dx, \qquad -\infty \leqslant x \leqslant \infty,$$

show that the c.f. of its square is given by

$$\phi(t) = (1-2it\sigma^2)^{-\frac{1}{2}}.$$

4.19 If two variates are distributed as in Exercise 4.17 show that the joint c.f. of x_1^2 and x_2^2 (with variables t_1 and t_2 respectively) is

$$\{(1-2it_1)(1-2it_2)+4\rho^2t_1t_2\}^{-\frac{1}{2}}.$$

4.20 A discontinuous distribution defined at integral values in the range $-\frac{1}{2}n(n-1)$ to $\frac{1}{2}n(n-1)$ has a frequency-generating function

$$P(t) = \prod_{j=1}^{n}\frac{t^j-t^{-j}}{j(t-t^{-1})}.$$

By considering the cumulant-generating function, and using the relation for Bernoulli numbers

$$|B_{2k}| = \frac{2(2k)!}{(2\pi)^{2k}}\sum_{j=1}^{\infty}\left(\frac{1}{j^{2k}}\right),$$

show that in standard measure the distribution tends to the normal form.

4.21 Show that the c.f. of the "logistic" distribution

$$dF = \frac{dx}{4\cosh^2(\frac{1}{2}x)} = \frac{\exp(-x)\,dx}{\{1+\exp(-x)\}^2}, \qquad -\infty \leqslant x \leqslant \infty,$$

is

$$\phi(t) = \Gamma(1-it)\,\Gamma(1+it)$$

and that its variance is $\pi^2/3$.

4.22 If x is an integer-valued random variable, show that the limits of integration in the inversion formula (4.5) may be replaced by $-\pi$ and π. Use this fact to invert the c.f. of a Poisson variate, given in Example 3.10.

CHAPTER 5

STANDARD DISTRIBUTIONS—(1)

5.1 There are certain distribution and frequency functions which, for both theoretical and practical reasons, occupy a central position in statistical theory. In this and the next chapter we shall consider their properties, leaving their statistical uses to be developed and illustrated later in the book. We shall, however, indicate briefly some of the ways in which they arise, even at the expense of anticipating ideas introduced at a subsequent stage. This will not impair the logical continuity of our development and will give concreteness to a treatment which might otherwise appear somewhat abstract.

The binomial distribution

5.2 Suppose we have a large population of members each of which exhibits either some quality P or a complementary quality Q (= not-P), for example, a population of men who are either blue-eyed or not-blue-eyed. Suppose that the proportion of individuals with quality P is p and that with quality Q is q, where of course $p+q = 1$. If we take a random sample of N members from the population we expect that on the average Np members will exhibit P and Nq will exhibit Q. We may thus array the members according to the quality as

$$N(p+q).$$

Now suppose we choose N *pairs* of individuals. There will be pairs PP, pairs PQ, pairs QP and pairs QQ. Of the Np pairs for which the first member is P there will, on the average, be a proportion p for which the second member is P and q for which it is Q. Similarly for the Nq exhibiting Q in the first member. Thus the pairs may be arrayed as

$$Np(p+q)+Nq(p+q) = N(p+q)^2.$$

Generally if we choose N sets of n the array will be $N(p+q)^n$. That is to say, the proportion of cases containing j P's and $(n-j)$ Q's will be $\binom{n}{j}p^j q^{n-j}$, the term in $p^j q^{n-j}$ in $(p+q)^n$. We are then led to consider the binomial distribution given by expanding

$$f = (p+q)^n \tag{5.1}$$

as a discontinuous frequency-distribution, the variate being the number of P's in the set of n, which may vary from n to 0. If, as is frequently more convenient, we wish to consider the variate as increasing from 0 to n, the distribution is inverted, i.e. becomes

$$f = (q+p)^n. \tag{5.2}$$

5.3 Distributions very close to the binomial form occur in practice, particularly in artificial experiments with coin-tossing or dice-throwing. Table 5.1 shows some

data due to Weldon, who threw 12 dice 26,306 times and noted the values at each throw. This is equivalent to the drawing of samples of 12 from a large population. The occurrence of a 5 or a 6 on any die was regarded as the exhibition of the quality P, a " success " as we may call it.

Table 5.1—Frequency-distribution of 26,306 throws of 12 dice, the occurrence of a 5 or 6 being counted a success

No. of successes	Observed frequency	Theoretical frequency from the binomial $26{,}306\,(0{\cdot}6623+0{\cdot}3377)^{12}$	No. of successes	Observed frequency	Theoretical frequency from the binomial $26{,}306\,(0{\cdot}6623+0{\cdot}3377)^{12}$
0	185	187	6	3,067	3,043
1	1,149	1,146	7	1,331	1,330
2	3,265	3,215	8	403	424
3	5,475	5,465	9	105	96
4	6,114	6,269	10 and over	18	16
5	5,194	5,115			
			TOTALS	26,306	26,306

If the dice were perfect (a condition rarely realized in practice) the proportion p of successes would be $\frac{1}{3}$; and the appropriate binomial would be, in the form (5.2), $(\frac{2}{3}+\frac{1}{3})^{12}$. In this particular case the dice were not quite perfect, the proportion of cases exhibiting a 5 or a 6 being 0·3377. Taking this as the value of p, we get the frequency function $(0{\cdot}6623+0{\cdot}3377)^{12}$, which when multiplied by the total frequency 26,306 gives the theoretical frequencies shown in the third column of Table 5.1. The agreement with observation is evidently fairly good.

5.4 We have already found the moments and factorial moments of the distribution in Examples 3.2 and 3.8. The characteristic function of the distribution is given by

$$\phi(t) = \sum_{j=0}^{n} \binom{n}{j} q^{n-j} p^j e^{ijt}$$
$$= (q+p\,e^{it})^n. \tag{5.3}$$

Taking logarithms and expanding we have for the cumulant-generating function (writing $\theta = it$)

$$\psi(t) = n\log(1+p\theta+\tfrac{1}{2}p\theta^2+\tfrac{1}{6}p\theta^3+\tfrac{1}{24}p\theta^4+\ldots).$$

Expanding the logarithm and identifying powers in θ we then find

$$\kappa_1 = \mu_1' = np\,;\quad \kappa_2 = \mu_2 = np(1-p) = npq\,;\quad \kappa_3 = \mu_3 = np(1-p)(1-2p)$$
$$= npq(q-p)\,;\quad \kappa_4 = np(1-p)(1-6p+6p^2) = npq(1-6pq)\,;\quad \text{etc.} \tag{5.4}$$

so that

$$\mu_4 = 3n^2p^2q^2+npq(1-6pq)\,; \tag{5.5}$$

$$\gamma_1 = \mu_3/\mu_2^{3/2} = \frac{q-p}{(npq)^{\frac{1}{2}}} = \sqrt{\beta_1}\,; \tag{5.6}$$

$$\gamma_2 = \kappa_4/\kappa_2^2 = \frac{1-6pq}{npq} = \beta_2-3. \tag{5.7}$$

5.5 Further formulae are not often required, but when they are can be derived from some interesting recurrence relations connecting the moments of the binomial.

Writing $\theta = it$ we have, for the characteristic function referred to the mean as origin,

$$\phi(t) = e^{-np\theta}(q+pe^{\theta})^n. \tag{5.8}$$

Differentiating with respect to θ we find

$$\sum_{j=1}^{\infty} \frac{\mu_j \theta^{j-1}}{(j-1)!} = -npe^{-np\theta}(q+pe^{\theta})^n + ne^{-np\theta}(q+pe^{\theta})^{n-1}pe$$

$$= -np \sum_{j=0}^{\infty} \frac{\mu_j \theta^j}{j!} + \frac{npe^{\theta}}{q+pe^{\theta}} \sum_{j=0}^{\infty} \frac{\mu_j \theta^j}{j!},$$

and hence, after a little re-arrangement,

$$(q+pe^{\theta})\left\{\sum_{j=1}^{\infty} \frac{\mu_j \theta^{j-1}}{(j-1)!}\right\} - npq(e^{\theta}-1) \sum_{j=0}^{\infty} \frac{\mu_j \theta^j}{j!} = 0.$$

Identifying coefficients in θ^{r-1} we get

$$\mu_r = npq \sum_{j=0}^{r-2} \binom{r-1}{j} \mu_j - p \sum_{j=0}^{r-2} \binom{r-1}{j} \mu_{j+1}, \tag{5.9}$$

giving the moment of order r about the mean in terms of those of lower orders.

Furthermore, writing the moment about the mean as

$$\mu_r = \sum_{j=0}^{n} (j-np)^r \binom{n}{j} (1-p)^{n-j} p^j$$

we have, differentiating with respect to p,

$$\frac{d\mu_r}{dp} = -rn\Sigma(j-np)^{r-1}\binom{n}{j}q^{n-j}p^j - \Sigma(j-np)^r\binom{n}{j}q^{n-j-1}(n-j)p^j$$

$$+\Sigma(j-np)^r\binom{n}{j}q^{n-j}jp^{j-1}.$$

The first term on the right is $-rn\mu_{r-1}$. The sum of the other two will be found to be $\frac{1}{pq}\Sigma(j-np)^{r+1}\binom{n}{j}q^{n-j}p^j = \frac{1}{pq}\mu_{r+1}$. Hence we find

$$\mu_{r+1} = pq\left(nr\mu_{r-1} + \frac{d\mu_r}{dp}\right). \tag{5.10}$$

For example, $\mu_1 = 0$, $\mu_2 = npq = np(1-p)$ and hence

$$\mu_3 = pq(n-2np)$$
$$= npq(q-p)$$

as stated in (5.4).

For factorial moments about the origin the expressions assume a particularly simple form. We have already found (Example 3.8) that

$$\mu'_{[r]} = p^r n^{[r]}. \tag{5.11}$$

5.6 If $p = q$ the binomial distribution is obviously symmetrical. If $p \neq q$ the distribution is skew. But in both cases it will be unimodal unless pn is small. For

the frequency of the $(r+1)$th term is greater than that of the rth so long as

$$\binom{n}{r+1}q^{n-r-1}p^{r+1} > \binom{n}{r}q^{n-r}p^r$$

or

$$\frac{n!}{(n-r)!\,r!}\cdot\frac{(n-r-1)!\,(r+1)!}{n!} < \frac{p}{q}$$

or

$$\frac{r+1}{n-r} < \frac{p}{q},$$

which is equivalent to

$$\frac{r+1}{n+1} < p. \tag{5.12}$$

Hence the frequency increases until the point when $(r+1) > p(n+1)$ and then declines again. Some typical distributions are shown in Table 5.2.

Table 5.2—Terms of the binomial distribution for $n = 20$ and values of p from 0·1 to 0·5 (each term has been multiplied by 10,000)

Number of successes	$p = 0{\cdot}1$ $q = 0{\cdot}9$	$p = 0{\cdot}2$ $q = 0{\cdot}8$	$p = 0{\cdot}3$ $q = 0{\cdot}7$	$p = 0{\cdot}4$ $q = 0{\cdot}6$	$p = 0{\cdot}5$ $q = 0{\cdot}5$
0	1216	115	8	—	—
1	2702	576	68	5	—
2	2852	1369	278	31	2
3	1901	2054	716	123	11
4	898	2182	1304	350	46
5	319	1746	1789	746	148
6	89	1091	1916	1244	370
7	20	545	1643	1659	739
8	4	222	1144	1797	1201
9	1	74	654	1597	1602
10	—	20	308	1171	1762
11	—	5	120	710	1602
12	—	1	39	355	1201
13	—	—	10	146	739
14	—	—	2	49	370
15	—	—	—	13	148
16	—	—	—	3	46
17	—	—	—	—	11
18	—	—	—	—	2
19	—	—	—	—	—
20	—	—	—	—	—

5.7 The ordinates of the binomial are most directly calculated from the formula $\binom{n}{j}q^{n-j}p^j$; for low values of n the calculation is straightforward and for high values assistance can be derived from tables of $\log n!$. The calculation of the distribution function, which is equivalent to the summation of terms of the binomial, is tedious to perform directly, but use may be made of the Tables of the Incomplete B-function. We have, for Taylor's series with the integral form of remainder,

$$f(a+h) = \sum_{j=0}^{r-1}\frac{h^j}{j!}f^{(j)}(a) + \int_0^1 \frac{h^r(1-t)^{r-1}}{(r-1)!}f^{(r)}(a+th)\,dt. \tag{5.13}$$

Putting $a = q$, $h = p$ and $f(a+h) = (q+p)^n$ we have

$$(q+p)^n = \sum_{j=0}^{r-1} \binom{n}{j} q^{n-j} p^j + R_r,$$

where R_r is the remainder after r terms and equals

$$R_r = \int_0^1 \frac{p^r (1-t)^{r-1}}{(r-1)!} \frac{n!}{(n-r)!} (q+pt)^{n-r} \, dt. \tag{5.14}$$

The substitution $t = 1 - x/p$ reduces this to

$$\begin{aligned} R_r &= \frac{n!}{(r-1)!\,(n-r)!} \int_0^p x^{r-1} (1-x)^{n-r} \, dx \\ &= \frac{\Gamma(n+1)}{\Gamma(r)\,\Gamma(n-r+1)} B_p(r, n-r+1) \\ &= \frac{B_p(r, n-r+1)}{B(r, n-r+1)} \\ &= I_p(r, n-r+1) \end{aligned} \tag{5.15}$$

in the usual notation. The sum of the first r terms is therefore

$$(q+p)^n - R_r = 1 - I_p(r, n-r+1) = I_q(n-r+1, r), \tag{5.16}$$

by a well-known property of the B-function.

The sum of the first $r+1$ terms is, similarly, $I_q(n-r, r+1)$, and hence the $(r+1)$th term is

$$\binom{n}{r} q^{n-r} p^r = I_q(n-r, r+1) - I_q(n-r+1, r). \tag{5.17}$$

Example 5.1

When $n = 20$, $r = 11$, $p = 0{\cdot}4$ we have for the remainder after 11 terms $I_{0{\cdot}4}(11, 10)$ which from the tables is found to be 0·127,521,2. The value given by summing the last six terms in the appropriate column of Table 5.2 is 0·1276, the error in the last place being due to rounding up. The remainder after 12 terms is $I_{0{\cdot}4}(12, 9)$ = 0·056,526,4. The 12th term (11 " successes ") is then the difference of these two remainders = 0·0710, as shown in Table 5.2 for the frequency per 10,000 of 11 successes.

Tables of the binomial distribution

Several detailed tables of the distribution have been prepared.

(a) The *Biometrika Tables* give the individual terms of the distribution for p = 0·01, 0·02 (0·02) 0·1 (0·01) 0·5 and n = 5 (5) 30, to 5 decimal places.

(b) *Tables of the Binomial Probability Distribution* (National Bureau of Standards, Applied Mathematics Series, 6, Washington, 1950) gives the individual terms and the distribution function for p = 0·01 (0·01) 0·50 and n = 2 (1) 49, to 7 d.p.

(c) Romig's *50–100 Binomial Tables* (Wiley, New York, 1953) extends (b) above to n = 50 (5) 100, to 6 d.p.

(d) *Tables of the Cumulative Binomial Probabilities* (U.S. Army Ordnance Corps Pamphlet ORD P20–1, Washington, 1952) gives the distribution function for p = 0·01 (0·01) 0·50 and n = 1 (1) 150, to 7 d.p.

(e) *Tables of the Cumulative Binomial Probability Distribution* (Harvard University Press, 1955) gives the distribution function for $p = 0{\cdot}01\,(0{\cdot}01)\,0{\cdot}50$, $p = \frac{1}{16}(\frac{1}{16})\frac{1}{2}$, $p = \frac{1}{12}(\frac{1}{12})\frac{1}{2}$ and $n = 1\,(1)\,50\,(2)\,100\,(10)\,200\,(20)\,500\,(50)\,1000$.

(f) Those edited by J. C. P. Miller (1954) give $\binom{n}{r}$ for $2 \leqslant r \leqslant \frac{1}{2}n \leqslant 100$; $r = 2\,(1)\,12$, $n \leqslant 500$; $r = 2\,(1)\,11$, $n = 500\,(1)\,1000$; $r = 2\,(1)\,5$, $n = 1000\,(1)\,2000$; $r = 2, 3$, $n = 2000\,(1)\,5000$.

As we have seen in Example 4.6, the binomial distribution tends to the "normal" distribution as n increases. Raff (1956) has shown that if $np^{3/2} > 1{\cdot}07$, the error in using the normal distribution function instead of the binomial never exceeds 0·05 for any r.

The Poisson distribution

5.8 Cases sometimes occur in which the proportion p of "successes" in the population is very small. We may suppose our number n large enough to render np itself appreciable though p is small; and we are thus led to consider the limiting form of the binomial distribution as $p \to 0$ subject to the condition that np remains finite, and equal to λ, say.

Under these conditions the term

$$\begin{aligned}\binom{n}{r}p^r q^{n-r} &= \frac{n!}{(n-r)!\,r!}\frac{\lambda^r}{n^r}\left(1-\frac{\lambda}{n}\right)^{n-r} \\ &\sim \frac{\sqrt{(2\pi)\,e^{-n}n^{n+\frac{1}{2}}}}{\sqrt{(2\pi)\,(n-r)^{n-r+\frac{1}{2}}e^{-n+r}n^r}}\cdot\frac{\lambda^r}{r!}e^{-\lambda} \\ &\sim \frac{1}{\left(1-\frac{r}{n}\right)^n e^r}\cdot\frac{\lambda^r}{r!}e^{-\lambda} \sim \frac{\lambda^r}{r!}e^{-\lambda}.\end{aligned}$$

Thus the terms of the binomial become the successive terms

$$e^{-\lambda}\left(1, \frac{\lambda}{1!}, \ldots, \frac{\lambda^j}{j!}, \ldots\right) \tag{5.18}$$

The distribution whose frequencies at $0, 1, 2, \ldots$ are arrayed by (5.18) is called the Poisson distribution, having been given first by Poisson in 1837.

From the point of view of characteristic functions we have (writing $\theta = it$)

$$\begin{aligned}\phi(t) &= \lim_{n\to\infty}(q+pe^{\theta})^n \\ &= \lim_{n\to\infty}\left\{1+\frac{\lambda}{n}(e^{\theta}-1)\right\}^n \\ &= \exp\{\lambda(e^{\theta}-1)\}\end{aligned} \tag{5.19}$$

which has already been shown in Example 3.10 to be the c.f. of the distribution (5.18).

Thus the cumulative function is given by

$$\psi(t) = \lambda(e^{\theta}-1) = \lambda\Sigma\theta^j/j!$$

and hence all cumulants of the Poisson distribution are equal to λ. We thus find

$$\left.\begin{aligned} \mu_1' &= \lambda, \\ \mu_2 &= \lambda, \\ \mu_3 &= \lambda, \\ \mu_4 &= \lambda + 3\lambda^2. \end{aligned}\right\} \tag{5.20}$$

If we let $n \to \infty$ in (5.9) and (5.10) we find

$$\mu_r = \lambda \sum_{j=0}^{r-2} \binom{r-1}{j} \mu_j \tag{5.21}$$

and

$$\mu_{r+1} = r\lambda\mu_{r-1} + \frac{\lambda\, d\mu_r}{d\lambda}. \tag{5.22}$$

5.9 A comparison of the successive terms $\lambda^r/r!$ and $\lambda^{r+1}/(r+1)!$ shows that the frequency increases up to the point for which $r+1 < \lambda$ and then declines to zero. For low values of λ the frequency-polygons of the distribution are very skew, being J-shaped for $\lambda < 1$, but become nearer to unimodal symmetry as λ increases.

The summation of r terms in the Poisson distribution may be carried out in a manner similar to that of **5.7**. The remainder after r terms of the distribution is found to be, from (5.13),

$$\begin{aligned} R_r &= \frac{\lambda^r}{\Gamma(r)} \int_0^1 e^{-\lambda(1-t)} (1-t)^{r-1}\, dt \\ &= \frac{\Gamma_\lambda(r)}{\Gamma(r)} \\ &= I\left(\frac{\lambda}{\sqrt{r}}, r-1\right), \end{aligned} \tag{5.23}$$

in the notation of Pearson's tables of the Incomplete Γ-function. The argument used in these tables is a difficult one to work with in the present case and, though formula (5.23) may be used for summing a number of terms in the Poisson distribution, it is easier to calculate $e^{-\lambda}\lambda^r/r!$ directly rather than to use an expression analogous to (5.17) in the form

$$r\text{th term} = I\left(\frac{\lambda}{\sqrt{(r-1)}}, r-2\right) - I\left(\frac{\lambda}{\sqrt{r}}, r-1\right).$$

Biometrika Tables for Statisticians, Part I, gives the terms $e^{-\lambda}\lambda^r/r!$ for $\lambda = 0{\cdot}1\,(0{\cdot}1)\,15$, $r = 0\,(1)\,k$ to 6 d.p. where k is the value for which the function is not zero to six places of decimals. The sum of terms can be obtained from tables of the χ^2 distribution and *Biometrika Tables for Statisticians* effectively gives $\sum_{j=0}^{r-1} e^{-\lambda}\lambda^j/j!$ for $\lambda = 0{\cdot}0005\,(0{\cdot}0005)$ $0{\cdot}005\,(0{\cdot}005)\,0{\cdot}05\,(0{\cdot}05)\,1{\cdot}0\,(0{\cdot}1)\,5{\cdot}0\,(0{\cdot}25)\,10{\cdot}0\,(0{\cdot}5)\,20\,(1)\,60$; all to five decimal places. These can be used to supplement the tables of the individual terms, especially in the range $0 \leqslant \lambda \leqslant 1$. Tables by Molina (1942) give the terms and the sums to 6 or 7 decimal places for $\lambda = 0{\cdot}001\,(0{\cdot}001)\,0{\cdot}010\,(0{\cdot}010)\,0{\cdot}30\,(0{\cdot}10)\,15\,(1)\,100$.

5.10 We now consider a generalization of the binomial and the Poisson distributions. In **5.2** our approach was based on the drawing of sets of n from the same

population. Suppose, however, that each set of n consists of an observation drawn from each of n different populations with proportions $p_1, p_2, \ldots, p_n$. Then our proportional frequencies will be arrayed by the form

$$(p_1+q_1)(p_2+q_2)\ldots(p_n+q_n) = \prod_{j=1}^{n}(p_j+q_j) \tag{5.24}$$

which of course reduces to the binomial distribution if all the p's are equal.

The characteristic function of this distribution is

$$\phi(t) = \Pi(q_j+p_j e^{\theta})$$

from which we have

$$\begin{aligned}\psi(t) &= \Sigma\log(q_j+p_j e^{\theta}) \\ &= \Sigma\log(1+p_j\theta+\tfrac{1}{2}p_j\theta^2+\ldots) \\ &= \theta\Sigma p_j+\tfrac{1}{2}\theta^2\Sigma(p_i-p_j^2)+\ldots, \text{ etc.},\end{aligned}$$

giving

$$\left.\begin{aligned}\mu_1' &= \Sigma p_j \\ \kappa_2 = \mu_2 &= \Sigma p_j q\end{aligned}\right\}. \tag{5.25}$$

Writing now $\bar{p}$ for the mean of the p's in the different populations, we have

$$\begin{aligned}\mu_1' &= n\bar{p} \\ \mu_2 &= \Sigma pq = \Sigma p-\Sigma p^2 \\ &= \Sigma p-\frac{1}{n}(\Sigma p)^2-\left\{\Sigma p^2-\frac{1}{n}(\Sigma p)^2\right\} \\ &= n\bar{p}-n\bar{p}^2-n\sigma_p^2\end{aligned}$$

where σ_p^2 is written for the variance of p)

$$= n\bar{p}\bar{q}-n\sigma_p^2. \tag{5.26}$$

A comparison of these results with those for the binomial distribution shows that the variance of the distribution arrayed by (5.24) is *less* than that of the binomial with the same average p by an amount equal to n times the variance of p.

Similarly we see that, for the Poisson distribution in such a case,

$$\mu_1' = \bar{\lambda}, \text{ the mean of the } \lambda\text{'s}, \tag{5.27}$$

and

$$\begin{aligned}\mu_2 &= \bar{\lambda}-\sigma_\lambda^2/n \\ &= \bar{\lambda}+O(n^{-1}).\end{aligned}$$

The Poisson value for the variance thus holds for (5.24) notwithstanding the inequality of the p's, provided that the variance of λ is small compared with n, which will be so if all the p's are small.

5.11 Consider now the case when each complete set of n is drawn from one of k different populations, with proportions $p_1, p_2, \ldots, p_k$. (In the previous case we supposed any set of n obtained by taking one observation from each of n populations. We now suppose that any set is drawn from one population only, but that different sets come from different populations.) Our array of frequencies will now be

$$\frac{1}{k}\sum_{j=1}^{k}(q_j+p_j)^n \tag{5.28}$$

and evidently the moments of this array are the means of the moments of the $(\bar{q}+p)^n$, that is to say, from (5.4),

$$\mu_1' = \frac{1}{k}\Sigma np,$$

$$\mu_2' = \frac{1}{k}\{\Sigma np + \Sigma n(n-1)p^2\}.$$

Writing $\bar{p}$ for the mean of the p's as before, we have

$$\begin{aligned}\mu_1' &= n\bar{p},\\ \mu_2 &= n\bar{p} + \frac{1}{k}\Sigma n(n-1)p^2 - n^2\bar{p}^2\\ &= n\bar{p}\bar{q} + \frac{1}{k}\Sigma n(n-1)p^2 - n(n-1)\bar{p}^2\\ &= n\bar{p}\bar{q} + n(n-1)\left(\frac{1}{k}\Sigma p^2 - \bar{p}^2\right)\\ &= n\bar{p}\bar{q} + n(n-1)\sigma_p^2. \end{aligned} \tag{5.29}$$

In this case the variance is *greater* than what it would be if the distribution were of the ordinary binomial type by an amount $n(n-1)$ var p.

For the Poisson distribution we have, on taking limits,

$$\left.\begin{aligned}\mu_1' &= \bar{\lambda}\\ \mu_2 &= \bar{\lambda} + \sigma\end{aligned}\right\} \tag{5.30}$$

and here also the variance of the distribution is affected.

5.12 We shall consider results of this kind in more detail when we discuss the theory of sample surveys, but a few comments may be helpful at this stage. It is often found in practice that data derived from the sampling of attributes over an extended area or an extended period of time do not conform to the simple binomial type. For example, suppose we regard the possession of blue eyes as a success, and take a number of samples of n from the population of the United Kingdom in different localities. We should probably find that the proportions in these samples did not conform to the simple binomial form. The variance calculated from the known n and average observed p would probably turn out to be smaller than the variance observed among the different samples. If so, we should conclude from (5.29) that the proportion p varied from place to place in the population, the excess in the variance of the proportions observed being due to the variance of p itself in the sections of the population from which the samples were chosen. We are assuming for the time being that these differences are not explicable on the basis of sampling fluctuation alone; but a full discussion will have to wait until later chapters.

5.13 The same effect is found in distributions which at first sight might be expected to be of the Poisson type. For example, suicide is a rare event and it might be supposed that if we took a series of large samples, say the population of the United Kingdom in successive years, the frequencies of suicides would follow the Poisson

distribution. This, however, is not necessarily so, for all members of the population are not equally exposed to risk, and the temptation to suicide may vary from year to year, e.g. being greater in years of trade depression. This inequality of risk is typical of one field in which the Poisson distribution has been freely applied, namely, industrial accidents. Table 5.3 shows, in the second column, the frequency of accidents occurring to women working on the manufacture of shells. The Poisson frequencies shown in the third column provide a very poor fit. One possible reason is that the liability of individuals to accident varies.

Table 5.3—Accidents to 647 women working on H.E. shells in 5 weeks (*Greenwood and Yule* (1920))

Number of accidents	Observed frequency	Poisson distribution with same mean	Distribution given by (5.33)
0	447	406	442
1	132	189	140
2	42	45	45
3	21	7	14
4	3	1	5
5 and over	2	0·1	2
TOTALS	647	648	648

As a working hypothesis (cf. Greenwood and Yule, 1920) suppose that the population is composed of individuals with different degrees of accident proneness, represented by different values of λ in a Poisson distribution; and suppose that in the population the distribution of λ is given by

$$dF = \frac{c^p}{\Gamma(p)} e^{-c\lambda} \lambda^{p-1} d\lambda, \qquad 0 \leqslant \lambda \leqslant \infty\ ;\ c > 0. \tag{5.31}$$

The frequency of j successes is then

$$\int_0^\infty \frac{c^p}{\Gamma(p)} e^{-c\lambda} \lambda^{p-1} e^{-\lambda} \frac{\lambda^j}{j!} d\lambda,$$

or the coefficient of t^j in

$$\frac{c^p}{\Gamma(p)} \int_0^\infty e^{-c\lambda} \lambda^{p-1} e^{-\lambda+\lambda t} d\lambda,$$

which, on the substitution of $(c+1-t)\lambda = u$, becomes

$$\frac{c^p}{(c+1-t)^p} = \left(\frac{c}{c+1}\right)^p \left(1-\frac{t}{c+1}\right)^{-p}. \tag{5.32}$$

The frequency of 0, 1, 2, . . . successes is therefore

$$\left(\frac{c}{c+1}\right)^p \left\{1, \frac{p}{c+1}, \frac{p(p+1)}{2!(c+1)^2}, \ldots\right\}. \tag{5.33}$$

The mean is thus

$$\mu_1' = \left(\frac{c}{c+1}\right)^p \left\{\frac{p}{c+1} + \frac{p(p+1)}{(c+1)^2} + \dots\right\}$$
$$= \left(\frac{c}{c+1}\right)^p \cdot \frac{p}{c+1} \cdot \left(\frac{c+1}{c}\right)^{p+1}$$
$$= \frac{p}{c}. \tag{5.34}$$

Similarly $\mu_2' = \frac{p}{c}\left(\frac{p+c+1}{c}\right)$, so that

$$\mu_2 = \frac{p(c+1)}{c^2} = \frac{p}{c} + \frac{p}{c^2}. \tag{5.35}$$

If we now put the observed mean and variance of Table 5.3 equal to the values of (5.34) and (5.35) we have two equations which can be solved for p and c. The distribution (5.33) can then be found. The frequencies are given in the fourth column of Table 5.3 and evidently give a much better agreement with the facts.

The negative binomial distribution

5.14 An interesting feature of the distribution arrayed in (5.33) is that it is generated by a binomial *with negative index.* In the approach adopted in **5.2** the index is necessarily positive. Cases not infrequently arise, however, in which data are represented by " negative binomials." We shall encounter one very simple generative process when we come to consider sequential sampling. Another classical case, the earliest to bring this type of distribution to the notice of statisticians, was given by Yule (1910); it does not depend on any arbitrary assumptions about distributions such as that embodied in (5.31).

Suppose that we have a population subjected to recurring attacks of a disease, and suppose that each attack affects (as if they were chosen at random) a proportion p of the population. After n attacks, the proportions of individuals who have had the disease 0, 1,... times will be the terms of the binomial $(q+p)^n$. If r attacks are fatal to the individual, the proportion of survivors after n attacks will be the sum of the first r terms of the binomial. The proportion dying during the nth exposure will be those who have had the disease $r-1$ times in the first $n-1$ exposures and contract it again, i.e. will be

$$\binom{n-1}{r-1} q^{n-r} p^{r-1} \times p = \binom{n-1}{r-1} q^{n-r} p^r.$$

Thus, since death does not commence until the rth exposure, the proportions of deaths at the rth, $(r+1)$th, . . . exposures are

$$p^r \left\{1, \quad rq, \quad \frac{r(r+1)}{2!} q^2, \dots\right\}$$

i.e. successive terms in $p^r(1-q)^{-r}$, a binomial with negative index. A law of this kind has been found to operate in experiments on the killing of bacteria by disinfectants.

5.15 Writing the negative binomial array in the form $p^n(1-q)^{-n}$, we have for the c.f. (θ as usual being written for it)

$$\phi(t) = p^n(1-qe^{\theta})^{-n}. \tag{5.36}$$

Thus for the cumulative function

$$\begin{aligned}\psi(t) &= n\log p - n\log(1-qe^{\theta}) \\ &= -n\log\left\{1-\frac{q}{p}(e^{\theta}-1)\right\}.\end{aligned} \tag{5.37}$$

Expanding and identifying coefficients we have

$$\left.\begin{aligned}&\kappa_1 = \frac{nq}{p}; \quad \kappa_2 = \frac{nq}{p^2}; \quad \kappa_3 = \frac{nq(1+q)}{p^3}; \\ &\kappa_4 = \frac{nq(1+4q+q^2)}{p^4}.\end{aligned}\right\} \tag{5.38}$$

These relations may be derived from equations (5.4) by putting therein q/p for p and $1/p$ for q.

The f.m.g.f. is given by

$$\begin{aligned}\omega(t) &= p^n\{1-q(1+t)\}^{-n} \\ &= (1-qt/p)^{-n}.\end{aligned} \tag{5.39}$$

Hence we find, for the factorial moments about the origin

$$\mu'_{[r]} = \left(\frac{q}{p}\right)^r (n+r-1)^{[r]}. \tag{5.40}$$

It is to be noted that for the negative binomial $\mu_2 > \mu'_1$, whereas for the ordinary binomial $\mu_2 < \mu'_1$, and for the Poisson $\mu_2 = \mu'_1$. A preliminary calculation of μ_2/μ'_1 is often a good guide to which of the three is likely to fit the data best.

The logarithmic distribution

5.16 There is an interesting limiting form of the negative binomial. Suppose that q represents the probability of the presence of an attribute and p of its absence. In some circumstances we may not observe—we may not even be capable of observing—the situation in which there is complete absence of the attribute. We are then led to consider a negative binomial with the zero class missing, i.e. the distribution with frequencies proportional to

$$p^n\left\{nq, \quad \frac{n(n+1)}{2!}q^2, \ldots\right\}. \tag{5.41}$$

The total frequency is $1-p^n$ and we may write the distribution as

$$\frac{np^n}{1-p^n}\left\{q, \quad \frac{(n+1)}{2!}q^2, \quad \frac{(n+1)(n+2)}{3!}q^3, \ldots\right\}. \tag{5.42}$$

Now let n tend to zero. For the initial factor we have

$$\lim\frac{n}{p^{-n}-1} = \lim\frac{n}{\exp(-n\log p)-1} = \frac{1}{-\log p},$$

in virtue of the rule known as l'Hôpital's. The distribution then tends to the form

$$\frac{-1}{\log(1-q)}\left\{q, \frac{q^2}{2}, \frac{q^3}{3}, \ldots\right\}, \tag{5.43}$$

that is to say, the frequency at integral values $r \geqslant 1$ is the coefficient of t^r in the frequency-generating function

$$P(t) = \frac{1}{\log(1-q)} \log(1-qt). \tag{5.44}$$

This distribution is due to R. A. Fisher *et al.* (1943). As we have introduced it, the derivative process looks rather artificial; for we obtained the negative binomial with integral index n and we are now letting n tend to zero. The justification for the procedure will be clearer when we discuss sampling theory. In fact, the sum of a number N of independent observations, each distributed in the same binomial form with index n, is itself binomially distributed with index Nn, a result which we state here without proof. We may regard the generative scheme of the logarithmic distribution (5.43) as the limit of a set of N observations as N tends to infinity but n tends to zero such that Nn tends to zero. An example of a distribution of the logarithmic type is given in Table 5.4; but this is to be regarded as illustrative of goodness of fit, not as evidence that the data are generated by a scheme of the kind we have just mentioned.

Table 5.4—Distribution of butterflies in Malaya, with theoretical frequencies given by the logarithmic distribution

(From Fisher, Corbet and Williams, *Jour. An. Ecology*, 1943, **12**, 42)

No. of species	Theoretical frequency	Observed frequency
1	135·05	118
2	67·33	74
3	44·75	44
4	33·46	24
5	26·69	29
6	22·17	22
7	18·95	20
8	16·53	19
9	14·65	20
10	13·14	15
11	11·91	12
12	10·89	14
13	10·02	6
14	9·28	12
15	8·63	6
16	8·07	9
17	7·57	9
18	7·13	6
19	6·74	10
20	6·38	10
21	6·06	11
22	5·77	5
23	5·50	3
24	5·25	3

5.17 Replacing t by e^{θ} in (5.44) we find the characteristic function of the distribution to be

$$\phi(t) = 1+\alpha\log\left\{1-\frac{q}{p}\theta-\frac{q}{p}\frac{\theta^2}{2!}-\frac{q}{p}\frac{\theta^3}{3!}-\frac{q}{p}\frac{\theta^4}{4!}\cdots\right\}$$

where $\alpha = 1/\log(1-q)$. Noting the similarity of this to (5.37) we have for the moments about the origin (not the cumulants)

$$\mu_1' = -\frac{\alpha q}{p};\quad \mu_2' = -\frac{\alpha q}{p^2};\quad \mu_3' = -\frac{\alpha q(1+q)}{p^3};\quad \mu_4' = -\frac{\alpha q(1+4q+q^2)}{p^4} \tag{5.45}$$

whence

$$\left.\begin{aligned}\mu_2 &= -\frac{\alpha q(1+\alpha q)}{p^2},\\ \mu_3 &= -\frac{\alpha q(1+q+3\alpha q+2\alpha^2 q^2)}{p^3},\\ \mu_4 &= -\frac{\alpha q\{1+4q+q^2+4\alpha q(1+q)+6\alpha^2 q^2+3\alpha^3 q^3\}}{p^4}.\end{aligned}\right\} \tag{5.46}$$

The hypergeometric distribution

5.18 Consider now the generalization of the approach of **5.2** when samples of n are drawn from a population of N individuals, where N is not necessarily large. If we take a sample which contains r P's and $n-r$ Q's, it can arise in

$$\binom{n}{r}\frac{Np(Np-1)\ldots(Np-r+1)\,Nq(Nq-1)\ldots(Nq-n+r+1)}{N(N-1)\ldots(N-n+1)}$$

$$= \binom{n}{r}\frac{(Np)^{[r]}(Nq)^{[n-r]}}{N^{[n]}} \tag{5.47}$$

ways. For there are $\binom{N}{n}$ ways of selecting the sample, and the r P's can be chosen in $\binom{Np}{r}$ ways, and the $n-r$ Q's in $\binom{Nq}{n-r}$ ways, if $r \leqslant Np$ and $n-r \leqslant Nq$, the expression given in (5.47) being equal to $\binom{Np}{r}\binom{Nq}{n-r}\Big/\binom{N}{n}$.

Hence we are led to consider the discontinuous distribution whose general term is

$$f_j = \frac{1}{N^{[n]}}\binom{n}{j}(Np)^{[j]}(Nq)^{[n-j]},\quad j=0,1,..,n;\ Np \geqslant n \leqslant Nq, \tag{5.48}$$

a form in which the analogy with the binomial distribution is evident. As $N\to\infty$ the form (5.48) approaches the binomial form.

The frequency-generating function

$$P(t) = \frac{1}{N^{[n]}}\sum_{j=0}^{n}\left\{\binom{n}{j}(Np)^{[j]}(Nq)^{[n-j]}\,t^j\right\} \tag{5.49}$$

is equal to

$$\frac{(Nq)^{[n]}}{N^{[n]}}\sum\frac{(Np)^{[j]}n^{[j]}\ t^j}{(Nq-n+j)^{[j]}\,j!},$$

that is to say, to the hypergeometric function

$$\frac{(Nq)^{[n]}}{N^{[n]}}F(\alpha,\beta\,;\ \gamma,t)$$

if
$$\alpha = -n,\quad \beta = -Np,\quad \gamma = Nq-n+1. \tag{5.50}$$

The distribution (5.48) is therefore called hypergeometric. We have

$$F(\alpha,\beta\,;\ \gamma,t) = 1+\frac{\alpha\beta}{\gamma}\frac{t}{1!}+\frac{\alpha(\alpha+1)\beta(\beta+1)}{\gamma(\gamma+1)}\frac{t^2}{2!}+\cdots$$

and it is well known that this function satisfies the differential equation

$$t(1-t)\frac{d^2F}{dt^2}+\{\gamma-(\alpha+\beta+1)t\}\frac{dF}{dt}-\alpha\beta F = 0, \tag{5.51}$$

a fact which may be readily verified from the equation itself.

If in (5.49) we put $t = e^\theta$ we evidently have the characteristic function of the distribution. On making this substitution in (5.51) we find, after some reduction and replacement of the values of α, β, γ by those of (5.50),

$$(1-e^\theta)\left\{\frac{d^2\phi}{d\theta^2}-(n+Np)\frac{d\phi}{d\theta}+nNp\phi\right\}-Nnp\phi+N\frac{d\phi}{d\theta} = 0. \tag{5.52}$$

Since $\phi = \Sigma\mu_j'\theta^j/j!$ we find, from the coefficient of θ^0 in this expression,

$$-Nnp+N\mu_1' = 0$$
$$\mu_1' = np, \tag{5.53}$$

the same result as for the binomial. The mean of the hypergeometric series is independent of N.

Taking now the distribution about its mean, and hence substituting $e^{np\theta}\phi$ for ϕ in (5.52), we find

$$(1-e^\theta)\left[\frac{d^2\phi}{d\theta^2}+\frac{d\phi}{d\theta}\{n(p-q)-Np\}+(N-n)pqn\phi\right]+N\frac{d\phi}{d\theta} = 0 \tag{5.54}$$

whence, identifying coefficients in θ, θ^2, θ^3, we find

$$\left.\begin{aligned}
\mu_2 &= \frac{npq(N-n)}{N-1},\\
\mu_3 &= \frac{npq(q-p)(N-n)(N-2n)}{(N-1)(N-2)},\\
\mu_4 &= \frac{npq(N-n)}{(N-1)(N-2)(N-3)}\\
&\qquad\times[N(N+1)-6n(N-n)+3pq\{N^2(n-2)-Nn^2+6n(N-n)\}],
\end{aligned}\right\} \tag{5.55}$$

and generally, if E denotes the operation of raising the order of a moment by unity, i.e. $E\mu_r = \mu_{r+1}$, we have

$$N\mu_{r+1} = \{(1+E)^r-E^r\}[\mu_2-\{Np+n(q-p)\}\mu_1+\{npq(N-n)\mu_0\}]. \tag{5.56}$$

As we expect, when $N\to\infty$ these values tend to those of the binomial distribution.

5.19 An example of the occurrence of the hypergeometric series in practice is given in Table 5.5, giving the frequency of occurrence of cards of a certain suit in

hands of whist. Here N is the number of cards in the pack, 52, and $n = 13$, $p = \frac{1}{4}$. The appropriate frequencies are thus

$$\frac{1}{52^{[13]}}\binom{13}{j}13^{[j]}39^{[n-j]}$$

giving the figures shown in the third column. The agreement is reasonably good.

Table 5.5—Distribution of 3400 first hands at whist according to number of trumps in the hand (*K. Pearson* (*1924a*))

Number of cards in the hand	Observed frequency	Frequency of hypergeometric distribution	Number of cards in the hand	Observed frequency	Frequency of hypergeometric distribution
0	35	43·5	5	444	424·0
1	290	272·2	6	115	141·3
2	696	700·0	7	21	30·0
3	937	973·5	8	11	4·0
4	851	811·3	9 and over	0	0·2
			TOTALS	3400	3400

G. J. Lieberman and D. B. Owen (1961, *Tables of the Hypergeometric Probability Distribution*, Stanford U.P.) give the d.f. and f.f. for $N = 1(1)50(10)100$ and all n; and for $N = 200(100)2000$ and selected n. Wise (1954) has developed a quickly convergent expansion for the d.f. in terms of the Incomplete Beta-function. Sandiford (1960, *J. Amer. Statist. Ass.*, **55**, 718) shows that a close approximation is obtained from a binomial distribution with the same mean and variance.

The normal distribution

5.20 We have already noted in Examples 4.6 and 4.9 that the binomial and the Poisson distributions tend, when expressed in standard measure, to the form

$$dF = \frac{1}{\sqrt{(2\pi)}}e^{-\frac{1}{2}x^2}dx, \qquad -\infty \leqslant x \leqslant \infty. \tag{5.57}$$

The more general form

$$dF = \frac{1}{\sigma\sqrt{(2\pi)}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu_1')^2\right\}dx, \quad -\infty \leqslant x \leqslant \infty\,;\ \sigma > 0, \tag{5.58}$$

is known as the normal distribution.(*) It is the most important theoretical distribution in statistics. The expression (5.57) is the normal distribution in standard measure. We recall some of its properties which have been given in examples.

The c.f. of (5.58) is easily found to be

$$\phi(t) = \exp(it\mu_1' - \tfrac{1}{2}t^2\sigma^2) \tag{5.59}$$

giving

$$\left.\begin{aligned}\mu_{2r} &= \frac{(2r)!}{2^r r!}\sigma^{2r},\\ \mu_{2r+1} &= 0, \qquad r \geqslant 1\end{aligned}\right\}. \tag{5.60}$$

(*) The description of the distribution as the "normal," due to K. Pearson, is now almost universal among English writers. Continental writers refer to it variously as the second law of Laplace, the Laplace distribution, the Gauss distribution, the Laplace-Gauss distribution and the Gauss-Laplace distribution. As an approximation to the binomial it was reached by Demoivre in 1738 but he did not discuss its properties.

For the cumulant-generating function we have

$$\psi(t) = it\mu_1' - \tfrac{1}{2}t^2\sigma^2$$

so that

$$\kappa_2 = \sigma^2, \qquad \kappa_r = 0, \qquad r > 2. \tag{5.61}$$

We also have $\beta_2 = 3$, $\gamma_2 = 0$, which accounts for the standard adopted for mesokurtosis in **3.32**.

5.21 The frequency function of the standardized normal distribution will occur so often that we shall introduce a special symbol for it. We write

$$\alpha(x) = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2}. \tag{5.62}$$

For the distribution function we have

$$F(x) = \int_{-\infty}^{x} \alpha(y)\,dy = \tfrac{1}{2} + \int_{0}^{x} \alpha(y)\,dy. \tag{5.63}$$

We may expand $\alpha(x)$ as a power series in x and integrate term by term to obtain

$$F(x) = \tfrac{1}{2} + \frac{1}{\sqrt{(2\pi)}}\left(x - \frac{x^3}{2.3} + \frac{x^5}{2!\,2^2.5} - \ldots\right). \tag{5.64}$$

This, however, is not a very useful expression for the calculation of $F(x)$. It converges too slowly except when x is small.

5.22 For large x an asymptotic series may be employed. Let us note that

$$\frac{\exp\{-\frac{1}{2}x^2(1+u^2)\}}{1+u^2} = \int_{\frac{1}{2}x^2}^{\infty} \exp\{-v(1+u^2)\}\,dv.$$

We may integrate both sides for u from $-\infty$ to ∞, and in virtue of the uniform convergence on the right-hand side, may there invert the order of integration. We then have

$$\begin{aligned}\int_{-\infty}^{\infty} \frac{\exp\{-\frac{1}{2}x^2(1+u^2)\}}{1+u^2}\,du &= \int_{\frac{1}{2}x^2}^{\infty}\int_{-\infty}^{\infty} \exp\{-v(1+u^2)\}\,du\,dv \\ &= \sqrt{(2\pi)}\int_{\frac{1}{2}x^2}^{\infty} \frac{1}{\sqrt{(2v)}} e^{-v}\,dv \\ &= \sqrt{(2\pi)}\int_{x}^{\infty} e^{-\frac{1}{2}u^2}\,du.\end{aligned}$$

Hence

$$\begin{aligned}1 - F(x) &= \frac{1}{\sqrt{(2\pi)}}\int_{x}^{\infty} e^{-\frac{1}{2}u^2}\,du \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\exp\{-\frac{1}{2}x^2(1+u^2)\}}{1+u^2}\,du \\ &= \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2} \cdot \frac{1}{\sqrt{(2\pi)}}\int_{-\infty}^{\infty} \frac{\exp(-\frac{1}{2}x^2u^2)}{1+u^2}\,du.\end{aligned} \tag{5.65}$$

The ratio of the "tail area" of the distribution, $1-F(x)$, to its bounding ordinate

$\alpha(x)$, is known as Mills' ratio (*) and denoted by $R(x)$. Thus

$$R(x) = \{1-F(x)\}\Big/\frac{1}{\sqrt{(2\pi)}}e^{-\frac{1}{2}x^2}$$

$$= e^{\frac{1}{2}x^2}\int_x^\infty e^{-\frac{1}{2}u^2}\,du. \tag{5.66}$$

From (5.65) we have

$$R(x) = \frac{1}{\sqrt{(2\pi)}}\int_{-\infty}^{\infty}\frac{\exp(-\frac{1}{2}x^2u^2)}{1+u^2}\,du, \tag{5.67}$$

and on replacing $\frac{1}{2}x^2u^2$ by t,

$$R(x) = \frac{x}{2\sqrt{\pi}}\int_0^\infty \frac{e^{-t}t^{-\frac{1}{2}}}{t+\frac{1}{2}x^2}dt.$$

Expanding the denominator and expressing the resultant integrals in terms of Gamma-functions we have

$$R(x) = \frac{x}{2\sqrt{\pi}}\int_0^\infty \frac{2}{x^2}\left(1-\frac{2t}{x^2}+\frac{4t^2}{x^4}-\ldots\right)e^{-t}t^{-\frac{1}{2}}\,dt$$

$$= \frac{1}{x}-\frac{1}{x^3}+\frac{1.3}{x^5}-\ldots(-1)^j\frac{1.3.5\ldots(2j-1)}{x^{2j+1}}+R_j(x), \tag{5.68}$$

where

$$|R_j(x)| = \frac{2x^{-(2j+1)}}{2\sqrt{\pi}}\int_0^\infty \frac{t^{2j+2}e^{-t}t^{-\frac{1}{2}}}{t+x^2}dt.$$

This series does not converge but

$$|R_j(x)| < \frac{2x^{-(2j+1)}}{2\sqrt{\pi}}\int_0^\infty t^{2j+\frac{1}{2}}e^{-t}\,dt$$

$$< x^{-(2j+1)}1.3.5\ldots(2j-1). \tag{5.69}$$

Thus the remainder at any point in the summation is less in absolute value than the last term taken into account, and the series is asymptotic. For large x the expansion is reasonably effective.

5.23 The most useful forms for the calculation of $F(x)$, or equivalently of $R(x)$, however, are continued fractions. One such, given by Laplace in 1805, is as follows. Define $z(t)$ by

$$\frac{1}{x}z(t) = R\{x(1-t)\} = \frac{1}{\sqrt{(2\pi)}}\int_{-\infty}^{\infty}\frac{\exp\{-\frac{1}{2}x^2u^2(1-t)^2\}\,du}{1+u^2}. \tag{5.70}$$

Then

$$\frac{1}{x}\frac{dz}{dt} = -x\{(1-t)z-1\}. \tag{5.71}$$

(*) It was tabulated by Mills in 1926, but had, in effect, been considered by earlier writers.

Hence if z has the expansion in powers of t

$$z = \sum_{j=0}^{\infty} y_j t^j$$

we find, on substitution in (5.70) and identification of coefficients,

$$\frac{1}{x^2}(r+1)y_{r+1}+y_r-y_{r-1} = 0. \tag{5.72}$$

Hence

$$\frac{y_r}{y_{r-1}} = \frac{1}{1+\dfrac{r+1}{x^2}\dfrac{y_{r+1}}{y_r}}.$$

Repeated applications of this formula give

$$\frac{y_1}{y_0} = \frac{x}{x+}\,\frac{2}{x+}\,\frac{3}{x+}\cdots\frac{n}{x+}\cdots$$

Also, from (5.70) we find

$$y_0 = xR(x), \qquad y_1/x^2 = 1-y_0.$$

Hence, substituting for y_1 we have, after a little re-arrangement,

$$R(x) = y_0/x = \frac{1}{x+}\,\frac{1}{x+}\,\frac{2}{x+}\,\frac{3}{x+}\cdots\frac{n}{x+}\cdots \tag{5.73}$$

This expression was used by Sheppard (1939, posthumous) in calculating the standard tables of the normal distribution, described in **5.25** below. At the end of this volume we give some tables which will suffice to illustrate the theory and examples given in this book.

5.24 Another continued fraction which converges much more rapidly than Laplace's has been given by Shenton (1954). Defining

$$\overline{R}(x) = \int_0^x \alpha(y)\,dy/\alpha(x) = \frac{1}{2\alpha(x)} - R(x),$$

Shenton finds

$$\overline{R}(x) = \frac{x}{1-}\,\frac{x^2}{3+}\,\frac{2x^2}{5-}\,\frac{3x^2}{7+}\cdots \tag{5.74}$$

Tables of the normal distribution

5.25 (a) Sheppard's tables give, among other things, Mills' ratio to 12 decimal places by intervals of 0·01 of x; the same to 24 decimal places by intervals of 0·1; and the negative natural logarithm of the tail area to 16 places by intervals 0·1.

(b) The *Biometrika Tables*, in a table based on Sheppard's work, give the ordinate and the distribution function of the standardized distribution, to 7 decimal places for $x = 0\ (0{\cdot}01)\ 4{\cdot}50$ and to 10 d.p. for $x = 4{\cdot}50\ (0{\cdot}01)\ 6{\cdot}00$, as well as an auxiliary table which extends to $x = 500$ and inverse tables giving the deviate and the ordinate from the value of the distribution function.

(c) *Tables of Normal Probability Functions* (National Bureau of Standards, Applied Mathematics Series, 23, Washington, 1953) gives the ordinate and $\{2F(x)-1\}$, where F is the distribution function, for $x = 0\ (0{\cdot}0001)\ 1\ (0{\cdot}001)\ 7{\cdot}800$ to 15 d.p., and the ordinate and $2\{1-F(x)\}$ to 7 *significant figures* for $x = 6\ (0{\cdot}01)\ 10$, as well as auxiliary tables for large values of x.

(d) *The Kelley Statistical Tables* (Harvard U.P., 1948) gives an inverse table of the

standardized deviate and the ordinate, each to 8 d.p., for values of the distribution function 0·0001 (0·0001) 0·9999, with a finer tabulation for 10 lower and 10 higher values extending to 0·000,000,001 and 0·999,999,999.

5.26 The shape of the standardized normal distribution

$$\alpha(x) = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2}$$

is illustrated in Fig. 5.1. It is symmetrical and unlimited in range, falling off to zero very rapidly as the variate increases. There are points of inflection at unit distance on either side of the mean.

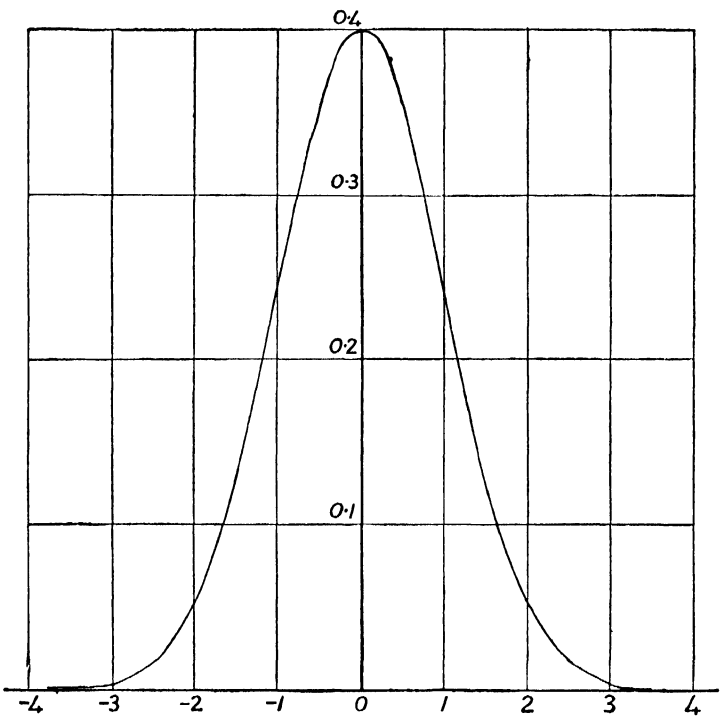


Fig. 5.1—The normal curve $y = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2}$

For the mean deviation we have

$$\frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} |x| e^{-\frac{1}{2}x^2} dx = \sqrt{\frac{2}{\pi}} \int_0^{\infty} x e^{-\frac{1}{2}x^2} dx$$

$$= \sqrt{\frac{2}{\pi}} = 0{\cdot}79788. \qquad (5.75)$$

The variance is of course unity, because the distribution is expressed in standard measure. The quartiles are distant 0·674,489,75 from the mean, as may be found from the Tables.

5.27 As an illustration of the occurrence in practice of a distribution which is very close to the normal, the height data of Table 1.7 may be taken. Table 5.6 shows

the actual frequencies and those given by the normal curve with the same mean and standard deviation (67·46 and 2·56 inches respectively).

The correspondence is evidently fairly good. It must, however, be noted that whereas the theoretical distribution has infinite range, the practical distribution has not, since it is impossible to have a negative height. In this particular case the relative frequency of the normal distribution outside the range 57–77 inches is so small that the point is unimportant; but when distributions of finite range are represented by those of infinite range it is as well to remember that the fit near the tails may not be very close.

Table 5.6—Frequency-distribution of 8585 men according to height (Table 1.7) compared with theoretical frequencies of a normal distribution with the same mean and variance

Height (inches)	Observed frequency	Theoretical frequency	Height (inches)	Observed frequency	Theoretical frequency
57–	2	1	68–	1230	1234
58–	4	3	69–	1063	989
59–	14	11	70–	646	682
60–	41	33	71–	392	405
61–	83	88	72–	202	207
62–	169	200	73–	79	91
63–	394	395	74–	32	34
64–	669	669	75–	16	11
65–	990	976	76–	5	3
66–	1223	1227	77–	2	1
67–	1329	1326			
			TOTALS	8585	8586

5.28 Since the normal distribution may be considered as the limit of the binomial, it is natural to inquire into the limiting forms, if any, of the hypergeometric distribution. From (5.48) we see that the difference between two successive terms in the distribution is

$$\frac{1}{N^{[n]}}\frac{n!}{r!(n-r-1)!}(Np)^{[r]}(Nq)^{[n-r-1]}\left\{\frac{Np-r}{r+1}-\frac{Nq-n+r+1}{n-r}\right\}$$
$$=\frac{1}{N^{[n]}}\frac{n!}{r!(n-r)!}(Np)^{[r]}(Nq)^{[n-r]}\left\{\frac{Nnp-Nq+n-1-r(N+2)}{(r+1)(Nq-n+r+1)}\right\}.$$

The ratio of this difference to the $(r+1)$th term is then

$$\frac{\Delta y_r}{y_r}=\frac{A+Br}{C+Dr+Er^2},$$

where the quantities $A \ldots E$ are constants. In the limit when the distribution is expressed in standard measure, Δy_r is the increment when r increases by a small quantity, and we are thus led to consider the differential equation defining a frequency function

$$\frac{df}{f}=\frac{A+Bx}{C+Dx+Ex^2}dx. \tag{5.76}$$

This is the equation of a family of functions—the Pearson distributions—which will be considered from a slightly different standpoint in the next chapter.

Some extensions

5.29 The distributions we have studied in this chapter may be generalized in various ways. The extension of the normal distribution to the multivariate case is so important that we shall later devote a chapter to it. We conclude our present account with a brief reference to two extensions of the binomial distribution.

The multinomial distribution

5.30 Suppose that the members of a population, instead of being classified into P's and not-P's as in **5.2**, can be classified into $k+1$ classes, $P_1, P_2, \ldots, P_k$ and neither $-P_1$ nor $-P_2 \ldots$ nor $-P_k$, which we may denote by P_0. Thus, for example, with $k = 3$, P_1 might denote the possession of blue eyes, P_2 that of grey eyes, P_3 that of brown eyes and P_0 that of eyes of some other colour, or perhaps no eyes at all. If we denote the corresponding proportions by $p_1, \ldots, p_k, p_0$, it is easy to see by an extension of the argument of **5.2** that the frequencies in sets of n individuals are given by the expansion of the multinomial

$$(p_0+p_1+p_2+\ldots+p_k)^n, \tag{5.77}$$

that is to say, the frequency of a set having r_0 P_0's, r_1 P_1's, etc., is the term

$$\frac{n!}{r_0!r_1!\ldots r_k!}\, p_0^{r_0}p_1^{r_1}\ldots p_k^{r_k}. \tag{5.78}$$

The characteristic function, attaching variables t_i to p_i $(r = 1, \ldots, k)$, is given by

$$(p_0+p_1e^{it_1}+p_2e^{it_2}+\ldots+p_ke^{it_k})^n \tag{5.79}$$

and we may obtain the moments and product-moments in the usual way. It will be sufficient if we write down certain of the cumulants up to the fourth order for $k = 3$. The others follow by symmetry.

$$\left.\begin{aligned}
&\kappa_{1000} = np_0\,; \quad \kappa_{2000} = np_0q_0\,; \quad \kappa_{1100} = -np_0p_1\,;\\
&\kappa_{3000} = np_0q_0(q_0-p_0)\,; \quad \kappa_{2100} = -np_0p_1(q_0-p_0)\,; \quad \kappa_{1110} = 2np_0p_1p_2\,;\\
&\kappa_{4000} = np_0q_0(1-6p_0q_0)\,; \quad \kappa_{3100} = -np_0p_1(1-6p_0q_0)\,;\\
&\kappa_{2200} = -np_0p_1\{(q_0-p_0)(q_1-p_1)+2p_0p_1\}\,;\\
&\kappa_{2110} = 2np_0p_1p_2(q_0-2p_0)\,; \quad \kappa_{1111} = -6np_0p_1p_2p_3.
\end{aligned}\right\} \tag{5.80}$$

Here, as usual, we write $q_i = 1-p_i$ for brevity.

The bivariate binomial distribution

5.31 We may also have a population classified by two qualities, e.g. blue-eyed or not-blue-eyed and male or female. (More generally, we could have several qualities and several categories in each, but this would give us a cumbrous complexity without raising any essentially new points.)

Suppose the presence or absence of one quality is denoted by P, Q $(= 1-P)$ and

those of the other quality by $P', Q' (= 1-P')$. The proportions of the four possible combinations may be represented thus:

	P'	Q'	TOTAL
P	p_{11}	p_{10}	p
Q	p_{01}	p_{00}	q
TOTAL	p'	q'	1

(5.81)

From some points of view we may regard this distribution as a multinomial arrayed by

$$(p_{00}+p_{01}+p_{10}+p_{11})^n. \tag{5.82}$$

We are, however, usually more interested in the distribution of p and p' than those of p_{00}, etc. The joint c.f. of the numbers of P's and P''s is given by

$$\phi(t) = (p_{00}+p_{01}e^{\theta_1}+p_{10}e^{\theta_2}+p_{11}e^{\theta_1+\theta_2})^n, \tag{5.83}$$

where, as usual, $\theta_1 = it_1$, $\theta_2 = it_2$. If we now define

$$p_{(11)} = p_{11}-pp' \tag{5.84}$$

we find, on expansion of $\log\phi(t)$, the bivariate cumulants

$$\left.\begin{aligned} \kappa_{11} = np_{(11)}; \quad \kappa_{21} = np_{(11)}(q-p); \quad \kappa_{31} = np_{(11)}(1-6pq); \\ \kappa_{22} = np_{(11)}\{(q-p)(q'-p')-2p_{(11)}\}. \end{aligned}\right\} \tag{5.85}$$

Other bivariate cumulants such as κ_{12} are obtainable by symmetry; and, of course, such cumulants as κ_{20} are the cumulants of the univariate (marginal) binomials $(q+p)^n$ and $(q'+p')^n$.

5.32 A little calculation shows that if $p_{(11)}$ of equation (5.84) is zero the c.f. becomes

$$(q+pe^{\theta_1})(q'+p'e^{\theta_2}). \tag{5.86}$$

The qualities are then independent. If we transfer the origin to the means of the two variates (5.83) becomes

$$\begin{aligned} \log\phi(t) &= n\log(p_{00}+p_{01}e^{\theta_1}+p_{10}e^{\theta_2}+p_{11}e^{\theta_1+\theta_2})-np\theta_1-np'\theta_2 \\ &= \tfrac{1}{2}n(pq\theta_1^2+2p_{(11)}\theta_1\theta_2+p'q'\theta_2^2)+o(n). \end{aligned} \tag{5.87}$$

If we then transform to standard measure, the variances of the two variates being npq and $np'q'$, and then let n tend to infinity, we find

$$\log\phi(t) = \tfrac{1}{2}(\theta_1^2+2\rho\theta_1\theta_2+\theta_2^2) \tag{5.88}$$

where

$$\rho = \frac{p_{(11)}}{(pqp'q')^{\frac{1}{2}}}. \tag{5.89}$$

This, as was seen, in effect, in Example 3.17, is the c.f. of the distribution

$$dF = \frac{1}{2\pi(1-\rho^2)^{\frac{1}{2}}} \exp\left\{\frac{-1}{2(1-\rho^2)}(x_1^2 - 2\rho x_1 x_2 + x_2^2)\right\} dx_1\, dx_2, \qquad -\infty \leqslant x_1, x_2 \leqslant \infty. \quad (5.90)$$

Thus the bivariate binomial distribution tends to the form (5.90) which may be regarded as the bivariate analogue of the normal distribution.

The variates x_1 and x_2 become independent if and only if $\rho = 0$, which implies that $p_{(11)} = 0$, or $p_{11} = pp'$. This in turn means that the proportion of P's among the P' is the same as the proportion among the Q', for then $p_{10} = p - p_{11} = qp'$. The table (5.81) then assumes the form

	P'	Q'	TOTAL
P	pp'	pq'	p
Q	qp'	qq'	q
TOTAL	p'	q'	1

(5.91)

We shall have more to say about this type of situation in discussing the theory of attributes. It is enough at this stage to remark that " independence " as we have defined it agrees with the ordinary meaning which we should assign to it in speaking of qualities.

Wishart (1949) has given general formulae for the cumulants of multivariate multinomial distributions.

EXERCISES

5.1 Show that for the binomial distribution

$$\kappa_{r+1} = pq\frac{d\kappa_r}{dp}, \qquad r > 1.$$

Hence, writing $c = pq$, $g = q - p$, that the cumulants are given by

$$\kappa_2 = nc\,; \quad \kappa_3 = ncg\,; \quad \kappa_4 = n(c - 6c^2)\,; \quad \kappa_5 = ng(c - 12c^2)\,; \quad \kappa_6 = n(c - 30c^2 + 120c^3)\,;$$
$$\kappa_7 = ng(c - 60c^2 + 360c^3)\,; \quad \kappa_8 = n(c - 126c^2 + 1680c^3 - 5040c^4).$$

(Cf. Frisch (1926), and Haldane (1939) who gives formulae up to κ_{12}.)

5.2 Show that for the *incomplete* moments about the mean of the binomial, starting from some variate-value ρ,

$$\mu_r = \sum_{j=\rho}^{n} (j - np)^r \binom{n}{j} p^j q^{n-j},$$

equation (5.10) holds, i.e.

$$\mu_{r+1} = pq\left(nr\mu_{r-1} + \frac{d\mu_r}{dp}\right).$$

(Romanovsky, 1925)

5.3 Writing $T_j = \binom{n}{j} p^j q^{n-j}$, show that the incomplete moments of the binomial are given by

$$\mu_0 = \sum_{j=\rho}^{n} T_j$$
$$\mu_1 = \rho q T_\rho$$
$$\mu_2 = \rho q T_\rho \{\rho - (n+1)p\} + npq\mu_0$$
$$\mu_3 = \rho q T_\rho [\{\rho - (n+1)p\}^2 + pq(2n-1)] + npq(q-p)\mu_0$$

and generally

$$\mu_r = \rho q T_\rho (\rho - np)^{r-1} + npq \sum_{j=0}^{r-2} \binom{r-1}{j} \mu_j - p \sum_{j=0}^{r-2} \binom{r-1}{j} \mu_{j+1}.$$

(Frisch (1926). This is the generalization of equation (5.9) to incomplete moments.)

5.4 In the binomial distribution let $np + \zeta$ be the least integer exceeding np. Show that the mean deviation of the distribution is

$$2(np+\zeta)\binom{n}{np+\zeta} p^{np+\zeta} q^{nq-\zeta+1}$$

(Johnson, 1957)

5.5 From equation (5.56) derive the recurrence formula for the moments of the binomial distribution

$$\{(1+E)^r - E^r\}(npq\mu_0 - p\mu_1) = \mu_{r+1}$$

and that for the Poisson distribution

$$\{(1+E)^r - E^r\}\lambda\mu_0 = \mu_{r+1}.$$

(K. Pearson, 1924*a*)

5.6 If a distribution of type (5.24) is represented approximately by a binomial $(Q+P)^\nu$, show that

$$\nu P = n\bar{p}$$
$$\nu PQ = n\bar{p}\bar{q} - n\sigma^2_p$$

so that $P = \bar{p} + \sigma^2_p/\bar{p}$ and hence is positive; consequently that ν is positive.

If, however, the distribution is of type (5.28), then

$$P = \bar{p} - (n-1)\sigma^2_p/\bar{p}$$

so that P, and hence ν, may be negative.

("Student," 1919)

5.7 (Neyman's Type A contagious distribution.) The frequency function of a variate r defined at $r = 0, 1, \ldots$ is given by

$$f_r = \frac{\lambda_2^r}{r!} e^{-\lambda_1} \sum_{j=0}^{\infty} \frac{j^r}{j!} (\lambda_1 e^{-\lambda_2})^j, \qquad \lambda_1, \lambda_2 > 0.$$

Show that the factorial cumulant g.f. is $\lambda_1(e^{\lambda_2 t} - 1)$ and that $\mu'_1 = \lambda_1\lambda_2$; $\mu_2 = \lambda_1\lambda_2(1+\lambda_2)$. Hence write down the c.f. and invert it, using Exercise 4.22.

(Neyman, J. (1940), *Ann. Math. Statist.*, **10**, 35)

5.8 (The Pólya-Aeppli distribution.) The frequency function of a variate r defined at $r = 0, 1, \ldots$ is given by

$$f_0 = e^{-\lambda_1}$$
$$f_r = e^{-\lambda_1} \tau^r \sum_{j=1}^{r} \binom{r-1}{j-1} \frac{1}{j!} \left\{\frac{\lambda_1(1-\tau)}{\tau}\right\}^j, \qquad r \geqslant 1.$$

Show that the f.c.g.f. is $\lambda_1 t/(1-\tau-\tau t)$ and hence find the mean and variance. (Exercises 5.7 and 5.8 are examples of the distribution of numbers of individuals living in colonies whose size also has a distribution. In both, the individuals are distributed in the Poisson form ; in 5.7 the colony-size is also a Poisson, but in 5.8 it has a J-shaped distribution $(1-\tau)\,\tau^{\rho-1}$, $\rho = 1, 2, \ldots$).

5.9 Show that if the frequencies of two symmetrical binomial forms of degree n are superposed so that the rth term of one is added to the $(r+1)$th term of the other, the resultant frequencies are those of a symmetrical binomial of degree $(n+1)$. Deduce that if two normal distributions with the same variance and means differing by a small part of the variance are added together, the resultant distribution is nearly normal.

5.10 Show that if $y = \dfrac{1}{\sigma\sqrt{(2\pi)}} \exp(-\frac{1}{2}x^2/\sigma^2)$

$$\int_{-\infty}^{\infty} y^2\,dx = \frac{1}{2\sigma\sqrt{\pi}}.$$

Hence, if a normal distribution is grouped in intervals with total frequency N_1, and N_2 is the sum of the squares of frequencies, an estimate of σ is

$$\frac{N_1^2}{2N_2\sqrt{\pi}} = 0{\cdot}282{,}095\frac{N_1^2}{N_2}.$$

For the height data of Table 1.7 show that this gives an estimate of σ equal to 2·553, an error of about 1 per cent.

(Yule, 1938)

5.11 (The bivariate Poisson distribution.) Show that when p_{01}, p_{10} and p_{11} in equation (5.82) are small, but $np_{11} = \lambda_3$, $np_{10} = \lambda_2 - \lambda_3$ and $np_{01} = \lambda_1 - \lambda_3$ are finite, the distribution tends to the form whose general term is

$$\frac{\lambda_3^i(\lambda_1-\lambda_3)^j(\lambda_2-\lambda_3)^k}{i!\,j!\,k!}\,e^{-\lambda_1-\lambda_2+\lambda_3}.$$

5.12 Use the Laplace continued fraction (5.73) to show that

$$\frac{x}{x^2+1} < R(x) < \frac{x^2+2}{x^3+3x}.$$

5.13 Show that Mills' ratio

$$R(x) = x\int_{-\infty}^{\infty} \frac{1}{u^2+x^2}\,\frac{1}{\sqrt{(2\pi)}}\,e^{-\frac{1}{2}u^2}\,du.$$

Replacing the integral by the sum $h\sum_{j=-\infty}^{\infty} \dfrac{1}{\sqrt{(2\pi)}}\,\dfrac{e^{-\frac{1}{2}j^2h^2}}{x^2+j^2h^2}$,

show that for $h = \frac{1}{4}$, $x = 1$, the sum to 13 terms gives $R = 0{\cdot}158{,}655{,}24$. (Das, 1955. The true value is 0·158,655,25.)

5.14 For the Laplace continued fraction (5.73) for Mills' ratio, show that if the sth convergent c_s is denoted by a_s/b_s,

$$a_0 = 0,\ a_1 = 1,\ b_0 = 1,\ b_1 = x \text{ and}$$

$$c_s - c_{s-1} = \frac{(-1)^s(s-1)!}{b_{s-1}b_s}$$

$$c_s - c_{s-2} = \frac{(-1)^{s-1}x(s-2)!}{b_{s-2}b_s}.$$

Hence show that

$$c_0 < c_2 < c_4 < \ldots < R(x) < \ldots < c_5 < c_3 < c_1.$$

(Shenton, 1954)

5.15 In the previous exercise show that

$$R(x) = \int_0^\infty \exp(-\tfrac{1}{2}t^2 - tx)\,dt$$

and hence that

$$R(x)\,b_s - a_s = \frac{(-1)^s \int_x^\infty (t-x)^s e^{-\frac{1}{2}x^2}\,dx / \sqrt{(2\pi)}}{\frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2}}.$$

In the Schwarzian inequality

$$\begin{vmatrix} \int f_1^2\,dx & \int f_1 f_2\,dx \\ \int f_1 f_2\,dx & \int f_2^2\,dx \end{vmatrix} > 0$$

put $f_1 = (x-t)^{\frac{1}{2}s} e^{-\frac{1}{4}x^2}, \qquad f_2 = (x-t)^{\frac{1}{2}s+1} e^{-\frac{1}{4}x^2},$

and hence show that

$$R > \frac{\alpha_{2s} + (2s)!\sqrt{(x^2+8s+4)}}{2\beta_{2s}}$$

$$R < \frac{2\gamma_{2s+1}}{\alpha_{2s+1} + (2s+1)!\sqrt{(x^2+8s+8)}}$$

where $\alpha_s = c_s a_{s+2} + c_{s+2} a_s - 2c_{s+1} a_{s+1}$

$$\beta_s = b_s b_{s+2} - b_{s+1}^2, \quad \gamma_s = a_s a_{s+2} - a_{s+1}^2.$$

Hence derive Birnbaum's inequality

$$\tfrac{1}{2}\{-x + \sqrt{(x^2+4)}\} < R(x) < 4/\{3x + \sqrt{(x^2+8)}\}, \quad x > 0,$$

the positive sign of the radical being taken.

(Shenton, 1954)

5.16 If $v(x) = 1/R(x)$ and $\lambda(x) = dv/dx = v(v-x)$ show that $0 < \lambda < 1$. If $\pi(x)$ is defined by

$$\pi(x) = e^{-x^2} \Big/ \int_{-\infty}^{x} e^{-\frac{1}{2}u^2}\,du \int_x^\infty e^{-\frac{1}{2}u^2}\,du$$

show that

$$d\pi/dx = 2x\pi(x)\{\lambda(x') - 1\},$$

where $-|x| \leqslant x' \leqslant |x|$. Hence show that $\pi(x)$ is a decreasing function of x^2.

(Sampford, 1953)

5.17 Consider the integral

$$\frac{1}{2\pi} \int_{-a}^{a} \int_{-a}^{a} \exp\{-\tfrac{1}{2}(x^2+y^2)\}\,dx\,dy.$$

Show that it is less than the same integral taken over a circle centred at the origin and of equal area of integration $4a^2$. Hence show that

$$\frac{1}{\sqrt{(2\pi)}} \int_0^x e^{-\frac{1}{2}u^2}\,du < \tfrac{1}{2}\{1 - \exp(-2x^2/\pi)\}^{\frac{1}{2}}, \quad x > 0.$$

(Pólya, 1945)

5.18 Show that the function

$$\overline{R}(x) = y = e^{\frac{1}{2}x^2}\int_0^x e^{-\frac{1}{2}u^2}\,du$$

obeys the differential equation

$$\frac{dy}{dx} = xy+1.$$

Hence show that

$$y = \frac{x}{1}+\frac{x^3}{1.3}+\frac{x^5}{1.3.5}+\frac{x^7}{1.3.5.7}+\ldots$$

(Pólya, 1945)

5.19 For the multinomial (5.77) show that if $a_i = p_i/p_0$ then

$$\kappa(r_1, r_2, \ldots, r_i+1, \ldots, r_k) = a_i\frac{\partial}{\partial a_i}\kappa(r_1, r_2, \ldots, r_i, \ldots, r_k)$$

where the cumulant on the right is of order r_1 in p_1 etc. (Guldberg, 1935). Hence derive equations (5.80).

5.20 Verify equations (5.85) and the expression (5.86).

5.21 In a certain experiment clusters are observed, each cluster consisting of individuals. Show that if the number of clusters follows a Poisson distribution and the number of individuals in a cluster follows a logarithmic distribution, the total number of individuals observed follows a negative binomial distribution.

(Quenouille, M. H. (1949), *Biometrics*, **5**, 162)

5.22 z is the sum of n independent variates x_i, each with c.f. $\phi_1(t)$, where n is a random variable, taking the values 0, 1, 2, . . ., with c.f. $\phi_2(t)$. Show that the unconditional c.f. of z is $\phi_2\left(\dfrac{\log\phi_1(t)}{i}\right)$, and that there is non-zero probability that $z = 0$ even if the x_i are continuous.

If n is a Poisson variable with parameter λ, show that the c.f. of z is $\exp\{\lambda[\phi_1(t)-1]\}$. Derive the results of Exercises 5.7 and 5.21 from this.

52 3 Use the result of Exercise 4.16 to establish (cf. **5.13**) that a Poisson distribution with parameter λ itself distributed in a Type III distribution (5.31) has the negative binomial distribution (5.33).

5.24 x is conditionally normally distributed with mean μ and variance λ. μ and λ are random variables with joint c.f. $\phi(t,u) = E\{\exp(it\mu+iu\lambda)\}$. Show that the unconditional c.f. of x is $\phi(t, \frac{1}{2}it^2)$.

5.25 In Exercise 5.24, show that if μ and λ are distributed independently of each other with respective c.f.'s

$$\phi_1(t) = (1+t^2\sigma^2)^{-1}, \qquad \phi_2(t) = (1-2iu\sigma^2)^{-m}, \qquad m \text{ a positive integer,}$$

(cf. Exercises 4.3, 4.18 for these c.f.'s) then x is distributed as the sum of $(m+1)$ independent variates, each distributed exactly like μ.

5.26 Show that for the Poisson f.f. (5.18), $\mu'_{(r)} = \lambda^r$; and that for the hypergeometric f.f. (5.48), $\mu'_{(r)} = n^{(r)}(Np)^{(r)}/N^{(r)}$.

CHAPTER 6

STANDARD DISTRIBUTIONS—(2)

6.1 In this chapter we continue the account, begun in the last, of the standard distributions of statistical theory. From the variety of forms assumed by the frequency-distributions of experience, as exemplified in Chapter 1, it is evident that an elastic system would be required to describe them all in mathematical terms. Three approaches will be considered herein : the first, due to Karl Pearson, seeks to ascertain a *family* of curves which will satisfactorily represent practical distributions ; the second, due to Bruns, Gram, Charlier and Edgeworth, seeks to represent a given frequency function as a series of derivatives of the normal frequency function ; the third, due to Edgeworth and later writers, seeks a transformation of the variate which will throw the distribution at least approximately into a known form.

Pearson distributions

6.2 It was noted in **5.28** that in the limiting case the hypergeometric series can be expressed in the form

$$\frac{df}{dx} = \frac{(x-a)f}{b_0+b_1x+b_2x^2}. \tag{6.1}$$

This equation may be considered from a slightly different standpoint. The unimodal distributions of Chapter 1 suggest that it might be worth while examining the class of frequency functions which (a) have a single mode, so that df/dx vanishes at some point $x = a$; (b) have smooth contact with the x-axis at the extremities, so that df/dx vanishes when $f = 0$. Evidently these conditions are in general obeyed by any distribution of the family (6.1). In actual fact, as will be seen below, there are also solutions of (6.1) in particular cases which are J- or U-shaped.

The family of frequency functions defined by (6.1) are known as Pearson distributions. Before obtaining explicit solutions of the equation, we consider certain general results which are true of all members of the system. We have immediately

$$(b_0+b_1x+b_2x^2)\,df = (x-a)f\,dx$$

or

$$x^n(b_0+b_1x+b_2x^2)\frac{df}{dx}dx = x^n(x-a)f\,dx.$$

Integrating the left-hand side by parts over the range of the distribution, we find, assuming that the integrals exist,

$$\Big[x^n(b_0+b_1x+b_2x^2)f\Big]_{-\infty}^{\infty} - \int_{-\infty}^{\infty}\{nb_0x^{n-1}+(n+1)b_1x^n+(n+2)b_2x^{n+1}\}f\,dx$$
$$= \int_{-\infty}^{\infty}x^{n+1}f\,dx - a\int_{-\infty}^{\infty}x^nf\,dx. \tag{6.2}$$

Let us assume that the expression in square brackets vanishes at the extremities of

the distribution, i.e. that $\lim_{x\to\pm\infty} x^{n+2}f \to 0$ if the range is infinite. We then have, substituting moments for integrals in (6.2) :—

$$-nb_0\mu'_{n-1}-(n+1)b_1\mu'_n-(n+2)b_2\mu'_{n+1} = \mu'_{n+1}-a\mu'_n$$

or
$$nb_0\mu'_{n-1}+\{(n+1)b_1-a\}\mu'_n+\{(n+2)b_2+1\}\mu'_{n+1} = 0. \tag{6.3}$$

This equation permits of the determination of any moment from those of lower orders. In fact, all moments can be expressed in terms of a, b_0, b_1 and b_2 and the moments μ_0 $(= 1)$ and μ'_1. Conversely we can express these four constants in terms of the moments μ'_1 to μ'_4, or the three moments about the mean μ_2 to μ_4. Putting $n = 0, 1, 2, 3$, successively in (6.3), we find equations for a, b_0, b_1, b_2 which result in

$$\left.\begin{aligned} a &= -\frac{\mu_3(\mu_4+3\mu_2^2)}{A} &&= -\frac{\sqrt{\mu_2}\sqrt{\beta_1}(\beta_2+3)}{A'} \\ b_0 &= -\frac{\mu_2(4\mu_2\mu_4-3\mu_3^2)}{A} &&= -\frac{\mu_2(4\beta_2-3\beta_1)}{A'} \\ b_1 &= -\frac{\mu_3(\mu_4+3\mu_2^2)}{A} &&= -\frac{\sqrt{\mu_2}\sqrt{\beta_1}(\beta_2+3)}{A'} \\ b_2 &= -\frac{(2\mu_2\mu_4-3\mu_3^2-6\mu_2^3)}{A} &&= -\frac{(2\beta_2-3\beta_1-6)}{A'} \end{aligned}\right\} \tag{6.4}$$

where
$$\left.\begin{aligned} A &= 10\mu_4\mu_2-18\mu_2^3-12\mu_3^2 \\ A' &= 10\beta_2-18-12\beta_1. \end{aligned}\right\} \tag{6.5}$$

For these values the origin is at the mean. In general, it follows that a curve of the family (6.1) is completely determined by its first four moments, μ'_1 to μ'_4.

6.3 In equation (6.1) the mode is evidently at the point $x = a$. From (6.4), we have for the Pearson measure of skewness **(3.31)**

$$\mathrm{Sk} = \frac{-a}{\sqrt{\mu_2}} = \frac{\sqrt{\beta_1}(\beta_2+3)}{10\beta_2-12\beta_1-18}, \tag{6.6}$$

the form given at (3.87).

Further, if we take an origin at the mode so that $a = 0$ we find

$$\frac{d^2f}{dx^2} = \frac{d}{dx}\frac{xf}{b_0+b_1x+b_2x^2} = \frac{f}{(b_0+b_1x+b_2x^2)^2}(b_0-b_2\ x^2). \tag{6.7}$$

Thus any points of inflection in the frequency curve are given by

$$x^2 = \frac{b_0}{b_2}. \tag{6.8}$$

Hence there cannot be more than two of them, and if two exist, they are equidistant from the mode. It is not to be inferred that a curve of the family cannot have a single point of inflection, for one point corresponding to the solution of (6.8) may be outside the permissible range of x.

6.4 By a simple transformation of the origin to the mode, (6.1) may be written

$$\left.\begin{aligned} \frac{d}{dx}(\log f) &= \frac{x-a}{B_0+B_1(x-a)+B_2(x-a)^2} \\ \text{or}\qquad \frac{d}{dX}(\log f) &= \frac{X}{B_0+B_1X+B_2X^2}. \end{aligned}\right\} \tag{6.9}$$

The explicit expression of the frequency function f is thus a matter of integrating the right-hand side of (6.9).

We may distinguish two main types according as the denominator on the right in (6.9) has real roots or imaginary roots.

TYPE I (*Beta-distribution*)

6.5 Let

$$B_0+B_1X+B_2X^2 = B_2(X+\alpha_1)(X-\alpha_2), \qquad \alpha_1,\ \alpha_2>0.$$

Then
$$\frac{d}{dX}(\log f) = \frac{X}{B_2(X+\alpha_1)(X-\alpha_2)}$$
$$= \frac{\alpha_1}{B_2(\alpha_1+\alpha_2)}\cdot\frac{1}{(X+\alpha_1)} + \frac{\alpha_2}{B_2(\alpha_1+\alpha_2)}\cdot\frac{1}{(X-\alpha_2)},$$

giving
$$f = k(X+\alpha_1)^{\frac{\alpha_1}{B_2(\alpha_1+\alpha_2)}}(X-\alpha_2)^{\frac{\alpha_2}{B_2(\alpha_1+\alpha_2)}}. \tag{6.10}$$

This may be written in the form

$$f = k\left(1+\frac{x}{a_1}\right)^{m_1}\left(1-\frac{x}{a_2}\right)^{m_2} \tag{6.11}$$

where
$$\frac{m_1}{a_1} = \frac{m_2}{a_2}.$$

The range of the curve is from $-a_1$ to a_2, and by integrating between these values we find

$$1 = k\int_{-a_1}^{a_2}\left(1+\frac{x}{a_1}\right)^{m_1}\left(1-\frac{x}{a_2}\right)^{m_2}dx,$$

which, on putting $x = (a_1+a_2)y-a_1$, reduces to

$$1 = k\int_0^1 y^{m_1}(1-y)^{m_2}\cdot\frac{(a_1+a_2)^{m_1+m_2+1}}{a_1^{m_1}a_2^{m_2}}\,dy$$
$$= \frac{k(a_1+a_2)^{m_1+m_2+1}}{a_1^{m_1}a_2^{m_2}}B(m_1+1,\ m_2+1).$$

This determines k and we have

$$f = \frac{a_1^{m_1}a_2^{m_2}}{(a_1+a_2)^{m_1+m_2+1}B(m_1+1,m_2+1)}\left(1+\frac{x}{a_1}\right)^{m_1}\left(1-\frac{x}{a_2}\right)^{m_2}. \tag{6.12}$$

The origin here is the mode. Taking an origin at the start of the curve we have

$$f = \frac{a_2^{m_2}}{(a_1+a_2)^{m_1+m_2+1}B(m_1+1,m_2+1)}x^{m_1}\left(1-\frac{x-a_1}{a_2}\right)^{m_2},$$

or again, measuring in units (a_1+a_2) times the original,

$$f = \frac{1}{B(m_1+1,m_2+1)}x^{m_1}(1-x)^{m_2}. \tag{6.13}$$

6.6 This is usually written in a form which has become almost standard :—

$$f = \frac{1}{B(p,q)} x^{p-1}(1-x)^{q-1}, \qquad p,\ q>0. \tag{6.14}$$

The distribution in this form ranges from 0 to 1. Its distribution function is the Incomplete Beta-function and it is sometimes known as the Beta-distribution. If both p and q exceed unity the curve has a unique mode at $(p-1)/(p+q-2)$ and is zero at its extremities. If either p or q is between 0 and 1 the ordinate at the corresponding terminal is infinite, and the distribution J-shaped.

The condition that $B_0+B_1X+B_2X^2$ has real roots of opposite sign, which has been assumed in deriving (6.14), is that B_0 and B_2 are of opposite sign, which is equivalent to $B_1^2/(4B_0B_2)<0$. In terms of β_1 and β_2 this becomes, on substitution from (6.4),

$$\frac{\beta_1(\beta_2+3)^2}{4(2\beta_2-3\beta_1-6)(4\beta_2-3\beta_1)}<0. \tag{6.15}$$

The quantity on the left is denoted by κ and is used as a criterion to distinguish the main types of Pearson distributions.

6.7 If the roots of $B_0+B_1X+B_2X^2$ are real and of the same sign it is easy to see by a similar argument that the frequency function may be written in the form

$$f = kx^{-q_1}(x-a)^{q_1}, \qquad q_1>q_2-1, \tag{6.16}$$

where $a\leqslant x\leqslant\infty$ if a is positive and $0\leqslant x\leqslant\infty$ if a is negative. This was called Type VI by Pearson, but since the simple transformation $y = a/x$ reduces it to the Type I form a separate study is unnecessary. The criterion κ of (6.15) is greater than unity.

In the form

$$f = \frac{kx^{p-1}}{(1+x)^{p+q}}, \qquad 0\leqslant x\leqslant\infty, \tag{6.17}$$

the distribution is sometimes known as a Beta-distribution of the second kind. Here $1/k = B(p,q)$ as in the distribution of the first kind. We shall refer to (6.14) and (6.17) respectively as Beta-distributions of the first and second kinds with parameters p and q.

TYPE IV

6.8 If the roots of $B_0+B_1X+B_2X^2$ are imaginary we have

$$\frac{d}{dX}(\log f) = \frac{X}{B_2\left\{\left(X+\frac{B_1}{2B_2}\right)^2+\frac{B_0}{B_2}-\frac{B_1^2}{4B_2^2}\right\}}$$

$$= \frac{X}{B_2\{(X+\gamma)^2+\delta^2\}}, \text{ say,}$$

giving

$$\log f = \log k+\frac{1}{2B_2}\log\{(X+\gamma)^2+\delta^2\}-\frac{\gamma}{B_2\delta}\arctan\frac{X+\gamma}{\delta},$$

$$f = k\{(X+\gamma)^2+\delta^2\}^{\frac{1}{2B_2}}\exp\left\{-\frac{\gamma}{B_2\delta}\arctan\frac{X+\gamma}{\delta}\right\}.$$

This is Pearson's Type IV and is usually written

$$f = k\left(1+\frac{x^2}{a^2}\right)^{-m} \exp\{-\nu \arctan(x/a)\}. \tag{6.18}$$

The distribution has unlimited range in both directions and is unimodal. It is very difficult to handle comfortably in practice, notwithstanding the existence of certain special tables, owing to the impossibility of expressing the distribution function in terms of ordinary functions. When the distribution function is required it is most easily derived by quadrature from the frequency function. The criterion κ of (6.15) lies between 0 and 1 in this case.

TYPE III (*Gamma-distribution*)

6.9 Pearson distinguished nine other types, some entirely trivial, some no longer of interest. We need mention only one, which has extensive theoretical applications. Some particulars of the others will be found in the exercises at the end of the chapter.

If, in (6.9), $B_2 = 0$ the distribution becomes

$$f = k\left(1+\frac{x}{a}\right)^{p} e^{-px/a}, \qquad -a \leqslant x \leqslant \infty. \tag{6.19}$$

Here the origin is at the mode. If we transfer to the start of the distribution and take a convenient scale we arrive at a form in which we have already encountered it (Examples 3.6, 3.12),

$$f = \frac{1}{\Gamma(\lambda)} x^{\lambda-1} e^{-x}, \qquad \lambda > 0,\ 0 \leqslant x \leqslant \infty. \tag{6.20}$$

In this case the criterion κ of (6.15) is infinite. The curve is unimodal except for values of λ less than or equal to unity, when it is J-shaped. It is known as Type III or the Gamma-distribution, the latter name being due to the fact that its distribution function is an Incomplete Γ-function.

6.10 It has been found that Pearson distributions sometimes give a good representation of observational data. Apart from this, they (and in particular Types I and III) are sometimes useful in approximating to theoretical distributions from their known moments, as we shall see in later chapters.

A systematic account of the technique of fitting is given by Elderton (1938*a*). We will here merely indicate the general principles and give one example of fitting in what is, perhaps, the most troublesome case.

6.11 All the Pearson distributions are determined by the first four moments, μ_1 to μ_4' inclusive, except some of the degenerate types which are determined by fewer than four moments. Pearson's method of fitting consists of

(1) determining the numerical values of the first four moments of the observed distribution ;

(2) calculating the numerical values of β_1, β_2, and the criterion κ of (6.15), and hence determining the type to which the distribution belongs ;

(3) equating the observed moments to the moments of the appropriate distribution expressed in terms of its parameters; and
(4) solving the resulting equations for those parameters, whereupon the distribution is determined.

The following example will illustrate the process :—

Example 6.1

In Table 1.15 there are shown, in the column totals, a distribution of 9440 beans according to length. The figures are repeated in Table 6.1 on page 160. Required to fit a Pearson distribution to these data.

For the moments it is found that, with Sheppard's corrections,

$$\begin{aligned} \mu_1' \text{ (origin at } 14{\cdot}5) &= -0{\cdot}190{,}783{,}898 \\ \mu_2 &= 3{\cdot}238{,}424{,}951 \\ \mu_3 &= -5{\cdot}306{,}566{,}352 \\ \mu_4 &= 50{\cdot}999{,}624{,}044 \end{aligned}$$

$$\beta_1 = 0{\cdot}829{,}135{,}838, \qquad \sqrt{\beta_1} = -0{\cdot}910{,}569$$
$$\beta_2 = 4{\cdot}862{,}944{,}362.$$

First of all, as to type. For the criterion of (6.15) we have

$$\kappa = \frac{\beta_1(\beta_2+3)^2}{4(4\beta_2-3\beta_1)(2\beta_2-3\beta_1-6)}$$
$$= \frac{51{\cdot}262}{84{\cdot}040}.$$

This lies between 0 and 1 and hence the appropriate curve is Type IV. Writing $\tan\theta = x/a$ and $2m-2 = r$ in equation (6.18) we find

$$\mu_n' = k\int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} a^{n+1}\cos^{r-n}\theta\sin^n\theta\, e^{-\nu\theta}\,d\theta,$$

whence, integrating by parts with $\cos^{r-n}\theta\sin\theta$ as one part,

$$\mu_n' = \frac{a}{r-n+1}\{(n-1)a\mu_{n-2}' - \nu\mu_{n-1}'\},$$

a particular case of (6.3). Hence, in terms of moments about the mean,

$$\mu_1' = -\frac{a\nu}{r},$$

$$\mu_2 = \frac{a^2}{r^2(r-1)}(r^2+\nu^2),$$

$$\mu_3 = -\frac{4a^3\nu(r^2+\nu^2)}{r^3(r-1)(r-2)},$$

$$\mu_4 = \frac{3a^4(r^2+\nu^2)\{(r+6)(r^2+\nu^2)-8\nu^2\}}{r^4(r-1)(r-2)(r-3)},$$

whence it is found that

$$r = \frac{6(\beta_2-\beta_1-1)}{2\beta_2-3\beta_1-6},$$

$$v = \frac{r(r-2)\sqrt{\beta_1}}{\sqrt{\{16(r-1)-\beta_1(r-2)^2\}}},$$

$$a = \sqrt{\left[\frac{\mu_2}{16}\{16(r-1)-\beta_1(r-2)^2\}\right]}.$$

Substituting for β_2, β_1 and μ_2 we find

$$r = 14{\cdot}697{,}72, \qquad m = 8{\cdot}348{,}86$$
$$v = 18{\cdot}380{,}43, \qquad a = 4{\cdot}159{,}49.$$

The signs here want a little watching. r and m present no difficulty; but a is to be taken positive and v positive since $\sqrt{\beta_1}$ is to be considered negative.

So far the process is straightforward, but to evaluate the constant k we require tables of the integral of the frequency function. Such tables were provided by Pearson.(*) On evaluation of k we find

$$f = 0{\cdot}395{,}121\left(1+\frac{x^2}{17{\cdot}301{,}34}\right)^{-8{\cdot}348{,}86}\exp\left(-18{\cdot}380{,}43\arctan\frac{x}{4{\cdot}159{,}49}\right).$$

With sufficient arithmetical patience the ordinates of the frequency-curve and the areas in given ranges can now be calculated. Table 6.1, column (3), gives the results for comparison with the observed frequencies.

6.12 The fitting of Pearson curves by the method of moments as above may be regarded from two rather different standpoints. If our object (for instance in actuarial work) is to obtain a mathematical expression which will satisfactorily represent observation and allow of accurate graduation and interpolation, fitting by moments is generally satisfactory. The method has, however, been criticized when the observed data are samples from a population, and it is desired to find a mathematical representation *of that population.* In such cases the moments calculated from observation are only *estimates* of population-moments and they do not, in general, lead to the most efficient estimates of the population parameters. We shall have to defer a fuller discussion of this point until Chapter 17, Volume 2.

6.13 Other systems of curves have been studied, with a view to representing frequency functions by expansions in series. It is well known in mathematical and physical work that functions can often be usefully expressed as a series of terms such as powers of the variable (Taylor's series) or trigonometrical functions (Fourier's series). Neither of these forms is very suitable for frequency functions, but we proceed to consider another set of functions with more promising possibilities.

(*) The tables were given in the first two editions of *Tables for Statisticians*, Part I, but have been omitted from later editions.

Tchebycheff-Hermite polynomials

6.14 Writing as before (5.21)

$$\alpha(x) = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2}$$

and

$$D = \frac{d}{dx},$$

consider successive derivatives of $\alpha(x)$ with respect to x. We have

$$\begin{aligned} D\alpha(x) &= -x\alpha(x) \\ D^2\alpha(x) &= (x^2-1)\alpha(x) \\ D^3\alpha(x) &= (3x-x^3)\alpha(x), \end{aligned}$$

and so on. The result will obviously be, in general, a polynomial in x multiplied by $\alpha(x)$. We then define the Tchebycheff-Hermite polynomial $H_r(x)$ by the identity

$$(-D)^r\alpha(x) = H_r(x)\alpha(x). \tag{6.21}$$

Evidently $H_r(x)$ is of degree r in x and the coefficient of x^r is unity. By convention $H_0 = 1$. We have

$$\alpha(x-t) = \frac{1}{\sqrt{(2\pi)}} \exp\left(-\tfrac{1}{2}x^2 + tx - \tfrac{1}{2}t^2\right) = \alpha(x)\exp\left(tx - \tfrac{1}{2}t^2\right)$$

and also, by Taylor's theorem,

$$\alpha(x-t) = \sum_{j=0}^{\infty} \frac{(-1)^j}{j!} t^j D^j\alpha(x) = \sum_{j=0}^{\infty} \frac{t^j}{j!} H_j(x)\alpha(x).$$

Consequently $H_r(x)$ is the coefficient of $\frac{t^r}{r!}$ in $\exp(tx - \frac{1}{2}t^2)$. It follows that

$$H_r(x) = x^r - \frac{r^{[2]}}{2.1!}x^{r-2} + \frac{r^{[4]}}{2^2.2!}x^{r-4} - \frac{r^{[6]}}{2^3.3!}x^{r-6} + \ldots \tag{6.22}$$

The first ten polynomials are

$$\left.\begin{aligned} H_0 &= 1 \\ H_1 &= x \\ H_2 &= x^2-1 \\ H_3 &= x^3-3x \\ H_4 &= x^4-6x^2+3 \\ H_5 &= x^5-10x^3+15x \\ H_6 &= x^6-15x^4+45x^2-15 \\ H_7 &= x^7-21x^5+105x^3-105x \\ H_8 &= x^8-28x^6+210x^4-420x^2+105 \\ H_9 &= x^9-36x^7+378x^5-1260x^3+945x \\ H_{10} &= x^{10}-45x^8+630x^6-3150x^4+4725x^2-945 \end{aligned}\right\} \tag{6.23}$$

6.15 The polynomials have a number of interesting properties. Differentiating the identity

$$\exp\left(tx - \tfrac{1}{2}t^2\right) = \sum_{j=0}^{\infty} \frac{t^j H_j(x)}{j!}$$

with respect to x and identifying coefficients in t^r we have

$$\frac{d}{dx}H_r(x) = rH_{r-1}(x) \tag{6.24}$$

and generally

$$D^j H_r(x) = r^{[j]} H_{r-j}(x). \tag{6.25}$$

Differentiating the identity with respect to t and identifying coefficients in t^{r-1} we have

$$H_r(x) - xH_{r-1}(x) + (r-1)H_{r-2}(x) = 0. \tag{6.26}$$

From (6.24) and (6.26) together we find

$$\frac{d^2 H_r(x)}{dx^2} - x\frac{dH_r(x)}{dx} + rH_r(x) = 0. \tag{6.27}$$

It is also known that the equation in x, $H_r(x) = 0$, has r real roots, each not greater in absolute value than $\sqrt{\{\frac{1}{2}r(r-1)\}}$. (Cf. Charlier, 1931.)

Tables of the values of the first six polynomials proceeding by $x = 0\ (0{\cdot}01)\ 4$ have been given by Jørgensen (1916).

6.16 The polynomials have an important orthogonal property, namely, that

$$\left.\begin{aligned} \int_{-\infty}^{\infty} H_m(x)H_n(x)\alpha(x)\,dx &= 0, \qquad m \neq n \\ &= n!, \qquad m = n. \end{aligned}\right\} \tag{6.28}$$

In fact, integrating by parts, we have, if $m \leqslant n$,

$$\begin{aligned} \int_{-\infty}^{\infty} H_m H_n \alpha\,dx &= (-1)^n \int_{-\infty}^{\infty} H_m D^n \alpha\,dx \\ &= (-1)^n \Big[H_m D^{n-1}\alpha\Big]_{-\infty}^{\infty} + (-1)^{n-1}\int_{-\infty}^{\infty} \frac{dH_m}{dx} D^{n-1}\alpha\,dx. \end{aligned}$$

The term in square brackets vanishes and, in virtue of (6.24), the integral becomes

$$m(-1)^{n-1}\int_{-\infty}^{\infty} H_{m-1} D^{n-1}\alpha\,dx.$$

Continuing the process, we find either zero, if m is not equal to n, or $m!$ if $m = n$.

The Gram-Charlier series of Type A

6.17 Suppose now that a frequency function can be expanded formally in a series of derivatives of $\alpha(x)$. (We shall discuss in **6.22** the conditions under which such an expansion is valid.) We have then

$$f(x) = \sum_{j=0}^{\infty} c_j H_j(x)\alpha(x).$$

Multiplying by $H_r(x)$ and integrating from $-\infty$ to ∞ we have, in virtue of the orthogonality relationship (6.28),

$$c_r = \frac{1}{r!}\int_{-\infty}^{\infty} f(x)H_r(x)\,dx. \tag{6.29}$$

The reader familiar with harmonic analysis will recognize the resemblance between this procedure and the evaluation of constants in a Fourier series.

Substituting in (6.29) the explicit value of $H_r(x)$ given in (6.22) we find

$$c_r = \frac{1}{r!}\left\{\mu'_r - \frac{r^{[2]}}{2.1!}\mu'_{r-2} + \frac{r^{[4]}}{2^2.2!}\mu'_{r-4} - \dots\right\}. \tag{6.30}$$

In particular, for moments about the mean,

$$\left.\begin{aligned} c_0 &= 1 \\ c_1 &= 0 \\ c_2 &= \tfrac{1}{2}(\mu_2 - 1) \\ c_3 &= \tfrac{1}{6}\mu_3 \\ c_4 &= \tfrac{1}{24}(\mu_4 - 6\mu_2 + 3) \\ c_5 &= \tfrac{1}{120}(\mu_5 - 10\mu_3) \\ c_6 &= \tfrac{1}{720}(\mu_6 - 15\mu_4 + 45\mu_2 - 15) \\ c_7 &= \tfrac{1}{5040}(\mu_7 - 21\mu_5 + 105\mu_3) \\ c_8 &= \tfrac{1}{40320}(\mu_8 - 28\mu_6 + 210\mu_4 - 420\mu_2 + 105) \end{aligned}\right\} \tag{6.31}$$

Thus we find the formal expansion

$$f(x) = \alpha(x)\{1 + \tfrac{1}{2}(\mu_2 - 1)H_2 + \tfrac{1}{6}\mu_3 H_3 + \tfrac{1}{24}(\mu_4 - 6\mu_2 + 3)H_4 + \dots\}. \tag{6.32}$$

If $f(x)$ is in standard measure the series becomes

$$f(x) = \alpha(x)\{1 + \tfrac{1}{6}\mu_3 H_3 + \tfrac{1}{24}(\mu_4 - 3)H_4 + \dots\}. \tag{6.33}$$

This is the so-called Gram-Charlier series of Type A.

Edgeworth's form of the Type A series

6.18 Consider the Fourier transformation of a term $H_r(x)\alpha(x)$.

Since $$\sqrt{(2\pi)}\,\alpha(t) = e^{-\frac{1}{2}t^2} = \int_{-\infty}^{\infty} e^{itx}\frac{1}{\sqrt{(2\pi)}}e^{-\frac{1}{2}x^2}dx$$

we have $$\sqrt{(2\pi)}\frac{d^r}{dt^r}\alpha(t) = (-1)^r\sqrt{(2\pi)}H_r(t)\alpha(t) = \int_{-\infty}^{\infty} i^r x^r \frac{e^{itx}}{\sqrt{(2\pi)}}e^{-\frac{1}{2}x^2}dx$$

and thus the transform of $x^r\alpha(x)$ is $i^r\sqrt{(2\pi)}H_r(t)\alpha(t)$. Conversely

$$x^r\alpha(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}i^r\sqrt{(2\pi)}H_r(t)\alpha(t)\,dt.$$

Interchanging x and t, we find

$$\sqrt{(2\pi)}(-i)^r t^r\alpha(t) = \int_{-\infty}^{\infty} e^{-ixt}H_r(x)\alpha(x)\,dx$$

and hence, changing the sign of t, that the transform of $H_r(x)\alpha(x)$ is $\sqrt{(2\pi)}i^r t^r\alpha(t)$.

Consider now the expression

$$\{\exp(\kappa_r D^r)\}\alpha(x). \tag{6.34}$$

Its characteristic function is

$$\begin{aligned}\int_{-\infty}^{\infty} e^{itx}\exp(\kappa_r D^r)\,\alpha(x)\,dx &= \int_{-\infty}^{\infty} e^{itx}\Sigma\left(\frac{\kappa_r^j D^{rj}}{j!}\right)\alpha(x)\,dx \\ &= \Sigma\frac{\kappa_r^j}{j!}\int_{-\infty}^{\infty} e^{itx}D^{rj}\alpha(x)\,dx \\ &= \Sigma\frac{\kappa_r^j}{j!}\int_{-\infty}^{\infty} e^{itx}(-1)^{rj}H_{rj}(x)\,\alpha(x)\,dx \\ &= \Sigma\frac{\kappa_r^j}{j!}\sqrt{(2\pi)}\,(-i)^{rj}t^{rj}\alpha(t) \\ &= \sqrt{(2\pi)}\,\alpha(t)\exp\{\kappa_r(-it)^r\}. \qquad (6.35)\end{aligned}$$

In a similar way it will be seen that the characteristic function of

$$\exp\left\{-\frac{\kappa_1-a}{1!}D+\frac{\kappa_2-b}{2!}D^2-\frac{\kappa_3}{3!}D^3+\frac{\kappa_4}{4!}D^4\ldots\right\}\alpha(x) \qquad (6.36)$$

is equal to

$$\sqrt{(2\pi)}\,\alpha(t)\exp\left\{\frac{\kappa_1-a}{1!}it+\frac{\kappa_2-b}{2!}(it)^2+\frac{\kappa_3}{3!}(it)^3+\frac{\kappa_4}{4!}(it)^4+\ldots\right\} \qquad (6.37)$$

More generally, if we replace $\alpha(x)$ by the unstandardized

$$\beta(x)=\frac{1}{\sigma\sqrt{(2\pi)}}e^{-\frac{1}{2}\frac{(x-m)^2}{\sigma^2}}=\alpha\{(x-m)/\sigma\}/\sigma,$$

the characteristic function of

$$\exp\left\{-\frac{\kappa_1-a}{1!}D+\frac{\kappa_2-b}{2!}D^2-\frac{\kappa_3}{3!}D^3+\frac{\kappa_4}{4!}D^4\ldots\right\}\beta(x) \qquad (6.38)$$

is equal to

$$\sqrt{(2\pi)}\,\alpha(t\sigma)\,e^{imt}\exp\left\{\frac{\kappa_1-a}{1!}it+\frac{\kappa_2-b}{2!}(it)^2+\frac{\kappa_3}{3!}(it)^3+\frac{\kappa_4}{4!}(it)^4\ldots\right\} \qquad (6.39)$$

as may be seen by the same line of argument.

Now suppose that (6.38) represents a frequency function. Its cumulant-generating function is then the logarithm of (6.39), i.e. is equal to

$$\frac{(\kappa_1-a+m)\,it}{1!}+\frac{\kappa_2-b+\sigma^2}{2!}(it)^2+\frac{\kappa_3}{3!}(it)^3+\frac{\kappa_4}{4!}(it)^4+\ldots$$

and hence its cumulants are κ_1-a+m, $\kappa_2-b+\sigma^2$, κ_3, $\kappa_4\ldots$, $\kappa_r\ldots$, etc. We may take $a=m$ and $b=\sigma^2$ and thus we obtain a distribution whose cumulants are κ_1, κ_2, ... etc. Now if these are in fact the cumulants of a distribution the series (6.38) must be equal to that distribution, provided that (1) the series converges to a frequency function, and (2) it is uniquely determined by its moments.

If we take the frequency function to be expressed in standard measure, then $\kappa_1=0$, $\kappa_2=1$ and (6.38) becomes

$$\exp\left\{-\kappa_3\frac{D^3}{3!}+\kappa_4\frac{D^4}{4!}\ldots\right\}\alpha(x)=f(x) \qquad (6.40)$$

where we have written $\alpha(x)$ for $\beta(x)$ because now m vanishes and $\sigma^2=1$.

A series of this kind was derived by Edgeworth (1904), though from an entirely different approach through the theory of elementary errors. Equation (6.40) is formally identical with (6.33), and the reader who consults the original memoirs on this subject may be puzzled by the fact that Edgeworth claimed his series to be different from the Type A series and better as a representation of frequency functions. The explanation is that for practical purposes it is necessary to take only a finite number of terms in the series and to neglect the remainder. If we take the first k terms in (6.33) the result is in general different from that obtained by taking the first $(k-1)$ terms of the operator in the exponential of (6.40). The argument centred on the fact (cf. Example 6.3 below) that the terms in (6.33) do not tend regularly to zero from the point of view of elementary errors, so that in general no term is negligible compared with a preceding term.

6.19 In standard measure the relations (6.31) become, in terms of cumulants,

$$\left.\begin{aligned} c_0 &= 1, \quad c_1 = c_2 = 0 \\ c_3 &= \frac{\kappa_3}{6} \\ c_4 &= \frac{\kappa_4}{24} \\ c_5 &= \frac{\kappa_5}{120} \\ c_6 &= \frac{1}{720}(\kappa_6 + 10\kappa_3^2) \\ c_7 &= \frac{1}{5040}(\kappa_7 + 35\kappa_4\kappa_3) \\ c_8 &= \frac{1}{40320}(\kappa_8 + 56\kappa_5\kappa_3 + 35\kappa_4^2) \end{aligned}\right\} \tag{6.41}$$

6.20 In the practical representation of frequency functions by the Type A series only the first few terms can be taken into account. The term in $H_r(x)$ has a coefficient dependent on μ_r and for $r>4$ this is unreliable owing to sampling fluctuations. When sampling effects are not in question the series may be taken to more terms, usually not higher than the term in H_6. We should then have to investigate how far the observed distribution can be represented by the series

$$\alpha(x)\left(1+\frac{\kappa_3}{6}H_3+\frac{\kappa_4}{24}H_4+\frac{\kappa_5}{120}H_5+\frac{\kappa_6+10\kappa_3^2}{720}H_6\right) \tag{6.42}$$

in the hope that the remainder after these terms could be neglected in comparison.

It may be noted in passing that, in contrast to the Pearson system, the distribution function of such a series is easy to obtain. If

$$f(x) = \Sigma a_r H_r(x)\alpha(x),$$

then

$$\begin{aligned} \int_{-\infty}^{x} f(x)dx &= \Sigma a_r \int_{-\infty}^{x} H_r(x)\alpha(x)\,dx \\ &= -\Sigma a_r H_{r-1}(x)\alpha(x). \end{aligned} \tag{6.43}$$

Table 6.1—Fitting of certain frequency curves to the distribution of length of beans (Table 1.15) (*from Pretorius, 1930, Johnson, 1949a*)

Length of beans (mm) (1)	Observed frequency (2)	Pearson Type IV (3)	Gram-Charlier Type A (three terms) (4)	Gram-Charlier Type A (four terms) (5)	Gram-Charlier Type A (five terms) (6)	Lognormal (7)	Johnson Type S_u (8)
>17·25	—	(	16·3	−15·2	2·0	—	0·1
17·0	6	(1·4	12·8	13·7	−35·3	(	2·0
16·5	55	28·5	25·6	116·6	22·3	(10·1	32·2
16·0	275	299·3	241·7	370·4	438·1	280·5	290·1
15·5	1129	1181·6	1012·7	926·2	1214·0	1255·2	1151·5
15·0	2082	2132·6	2155·4	1833·0	1866·9	2163·8	2130·3
14·5	2294	2229·8	2593·0	2506·4	2112·8	2179·6	2240·6
14·0	1787	1638·9	1788·4	2082·6	1916·7	1598·5	1642·5
13·5	929	968·9	713·4	921·3	1183·4	965·0	970·6
13·0	437	503·6	280·7	199·0	371·2	515·3	508·7
12·5	199	243·7	258·7	132·1	66·9	254·5	249·3
12·0	115	113·8	206·2	178·1	101·2	119·5	118·0
11·5	70	52·5	98·7	117·0	107·1	54·4	55·2
11·0	36	24·2	29·6	43·5	54·0	24·3	25·7
10·5	18	11·3	5·9	10·0	15·4	10·7	12·1
10·0	7	5·4	(	(	(	4·8	5·8
9·5	1	2·6	(·9	(1·7	(3·3	2·1	2·7
<9·25	—	1·9	—	—	—	1·7	2·6
TOTALS	9440	9440	9440	9440	9440	9440	9440

(The brackets mean that the frequencies shown are rounded up and include some small frequency in blank rows covered by the brackets.)

Example 6.2

Consider the fitting of a Type A series to the bean data of Example 6.1.

We have already found the first four moments. In standard measure we have

$$\mu_3 = -0{\cdot}910{,}569$$
$$\mu_4 = 4{\cdot}862{,}944$$

and we also find

$$\mu_5 = -12{\cdot}574{,}125$$
$$\mu_6 = 53{\cdot}221{,}083.$$

Hence the series is

$$9440\alpha(x)\{1-0{\cdot}151{,}762\,H_3+0{\cdot}077{,}622{,}7\,H_4-0{\cdot}028{,}903{,}6\,H_5+0{\cdot}014{,}273{,}5\,H_6\}.$$

Columns (4) to (6) of Table 6.1 above show the frequencies given by taking the first three, the first four and the first five terms of this series. A glance at the figures will show that the four- and five-term series are, if anything, rather worse than the three-term. Furthermore, the five-term series gives negative frequencies at one terminal and a mode at 12 mm., which is contrary to the data. The representation is clearly not very satisfactory and no better than that given by the Pearson Type IV curve.

Tetrachoric functions

6.21 The numerical evaluation of terms like $H_r(x)\alpha(x)$ may be carried out (a) directly by finding the polynomial and multiplying by $\alpha(x)$; (b) from tables of $H(x)$ and $\alpha(x)$; (c) from certain tables of the so-called Hermitian probability functions

with negative index (*British Association Mathematical Tables*, vol. 1, 1946). The latter give $Hh_n(x) \equiv (-D)^{n-1} e^{-\frac{1}{2}x^2}$ to 10 decimal places for $n = 0$ to 7, $x = -7$ (0·1) 6·5. Note that the factor $1/\sqrt{(2\pi)}$ does not appear. A similar function was tabulated by K. Pearson in the form

$$\tau_r(x) = \frac{H_{r-1}(x)\alpha(x)}{(r!)^{\frac{1}{2}}}. \tag{6.44}$$

This was known as a tetrachoric function, for reasons which will become apparent in the second volume when we discuss the estimation of correlation coefficients from a 2×2 table.

6.22 Up to this point it has been assumed that a frequency function possesses a convergent Type A series. We shall not here enter into a discussion of the conditions under which this is so, except to warn the reader that mistakes have been made on the subject and to quote some theorems without proof.

(1) Cramér (1926). If $f(x)$ is a function which has a continuous derivative such that

$$\int_{-\infty}^{\infty} \left(\frac{df}{dx}\right)^2 e^{\frac{1}{2}x^2}\, dx$$

converges and if $f(x)$ tends to zero as $|x|$ tends to infinity, then $f(x)$ may be developed in the series

$$f(x) = \sum_{j=0}^{\infty} \frac{c_j}{j!} D^j \alpha(x), \tag{6.45}$$

where c is given by

$$c_j = \int_{-\infty}^{\infty} f(x) H_j(x)\, dx.$$

This series is absolutely and uniformly convergent for $-\infty \leqslant x \leqslant \infty$.

(2) A theorem by Cramér (1926) based on one by Galbrun. If $f(x)$ is of bounded variation in every finite interval and if

$$\int_{-\infty}^{\infty} |f(x)|\, e^{\frac{1}{2}x^2}\, dx$$

exists, then the expansion of $f(x)$ in the series (6.45) converges everywhere to the sum $\frac{1}{2}\{f(x+0)+f(x-0)\}$. The convergence is uniform in every finite interval of continuity.

Cramér has also shown that this last theorem cannot be substantially improved upon as regards the behaviour of $f(x)$ at infinity. Consider in fact the function $f(x) = e^{-\lambda x^2}$. We have, in virtue of (6.26) and (6.24),

$$\begin{aligned}
\int_{-\infty}^{\infty} e^{-\lambda x^2} H_r(x)\, dx &= \int_{-\infty}^{\infty} e^{-\lambda x^2}.xH_{r-1}\, dx - (r-1)\int_{-\infty}^{\infty} e^{-\lambda x^2}.H_{r-2}\, dx \\
&= \left[\frac{e^{-\lambda x^2}}{-2\lambda} H_{r-1}\right]_{-\infty}^{\infty} + \frac{r-1}{2\lambda}\int_{-\infty}^{\infty} e^{-\lambda x^2} H_{r-2}\, dx - (r-1)\int_{-\infty}^{\infty} e^{-\lambda x^2} H_{r-2}\, dx \\
&= (r-1)\left(\frac{1}{2\lambda} - 1\right)\int_{-\infty}^{\infty} e^{-\lambda x^2} H_{r-2}(x)\, dx.
\end{aligned}$$

If r is odd the integral vanishes because H_r is an odd function of x. If r is even, say $2r$, the integral becomes

$$(2r-1)(2r-3)\ldots 1\left(\frac{1}{2\lambda}-1\right)^r\cdot\sqrt{\frac{\pi}{\lambda}}=\sqrt{\frac{\pi}{\lambda}}\frac{(2r)!}{2^r r!}\left(\frac{1}{2\lambda}-1\right)^r.$$

The appropriate coefficient of H_{2r} in the Type A series is then $\left(\frac{1}{2\lambda}-1\right)^r/(2^r r!)$. Now when $x=0$, $H_{2r}=(-1)^r(2r)!/(2^r r!)$. The series then becomes

$$\sum_{r=0}^{\infty}\sqrt{\frac{\pi}{\lambda}}\frac{(2r)!}{2^{2r}(r!)^2}\left(1-\frac{1}{2\lambda}\right)^r.$$

In virtue of the Stirling approximation to the factorial, the rth term of this, say u^r, becomes in the limit

$$u_r\sim\left(1-\frac{1}{2\lambda}\right)^r\frac{1}{\sqrt{(\lambda r)}}$$

so that

$$u_r^{1/r}\sim 1-\frac{1}{2\lambda}.$$

Hence, for $\lambda<\frac{1}{4}$ the series is divergent.

6.23 From the statistical viewpoint, however, the important question is not whether an *infinite* series can represent a frequency function, but whether a finite number of terms can do so to a satisfactory approximation. Two things seem clear :—

(a) The sum of a finite number of terms of the series may give negative frequencies, particularly near the tails (as, for instance, in Example 6.2).

(b) The series in the Charlier form (6.33) may behave irregularly in the sense that the sum of k terms may give a worse fit than the sum of $(k-1)$ terms.

Barton and Dennis (1952, *Biom.*, **39**, 425) show that if cumulants above the fourth are neglected, (6.33) and more particularly (6.42) do not give a unimodal f.f., or even non-negative frequencies, unless the skewness coefficient β_1 is close enough to zero. For example, (6.42) is certainly non-unimodal if $|\beta_1|\geqslant 0{\cdot}25$, and certainly has negative frequencies if $|\beta_1|\geqslant 0{\cdot}5$; (6.33) has a wider range of acceptability, but is certainly unimodal if $|\beta_1|\geqslant 0{\cdot}7$.

The finite series is successful only in cases of moderate skewness, and in such cases a Pearson distribution, or a transformed curve of the type we consider later, may be just as good. For many statistical purposes we are most interested in the tails of a distribution, and it is here that the inadequacies of this approach are most serious.

Example 6.3

As an illustration of the irregular behaviour of terms in the Type A series, consider the Gamma distribution

$$dF=\frac{1}{\Gamma(\lambda)}x^{\lambda-1}e^{-x}dx,\qquad 0\leqslant x\leqslant\infty;\quad \lambda>0.$$

Its characteristic function is (cf. Example 3.6)

$$\phi(t)=\frac{1}{(1-it)^{\lambda}}$$

and thus $\kappa_r = \lambda(r-1)!$ or, in standard measure,

$$\kappa_r = \frac{(r-1)!}{\lambda^{\frac{1}{2}r-1}}.$$

From the manner of the formation of terms in (6.41) it is evident that the coefficient c_r is the sum of terms κ_r, $\kappa_{r-3}\kappa_3$, ... $(\kappa_{q_1}\kappa_{q_2}\ldots\kappa_{q_m})$, where $(q_1 \ldots q_m)$ is a partition of r such that no q is less than 3. It will then be clear that, since κ_q is of order $\lambda^{1-\frac{1}{2}q}$, the term of greatest order in λ is that with the greatest number of parts in $(\kappa_{q_1} \ldots \kappa_{q_m})$. For example, if $r = 9$ it is (3^3), if $r = 8$ it is (4^2), and so on.

From these considerations we can find the order in λ of the terms in the Type A series. They are

Term	c_0	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	c_{12}
Order in λ	0	$-\frac{1}{2}$	-1	$-1\frac{1}{2}$	-1	$-1\frac{1}{2}$	-2	$-1\frac{1}{2}$	-2	$-2\frac{1}{2}$	-2

The terms decrease in order of λ, but not at all regularly, and it is clear that in general no coefficient will be negligible compared with a preceding one if λ is large. The asymptotic qualities of such series obviously require careful investigation in particular cases.

The Type B series

6.24 Charlier also proposed a Type B series based on derivatives of the Poisson frequency $e^{-\lambda}\lambda^x/x!$ with respect to λ, or equivalently, on first differences of that frequency with respect to x. This had certain mathematical attractions and has been used to graduate a few experiences of discontinuous variation resembling the Poisson distribution. It has not come into general use. A further form (Type C) proposed by Charlier to avoid negative frequencies has also failed to stand the test of experience.(*)

The normalization of frequency functions

6.25 Several of the important theoretical distributions occurring in statistics depend on some variable n in such a way that as n tends to infinity the distribution tends to normality. For large n it is often a sufficient approximation to assume the distribution normal, but for small or moderate n this may be hardly exact enough. In such a case we are nevertheless able to use the normal integral by seeking for a variate transformation

$$\xi = b_0 + b_1 x + b_2 x^2 + b_3 x^3 + \ldots \tag{6.46}$$

where the b's are of order $n^{-\frac{1}{2}}$ or smaller. By choosing the b's appropriately we can bring the distribution of ξ much nearer to normality than that of x and hence from (6.46) find the distribution function of x from that of ξ, assumed normal.

Consider in fact the Edgeworth form of the Type A expansion (6.38)

$$\exp\left\{-\frac{\kappa_1 - m}{1!}D + \frac{\kappa_2 - \sigma^2}{2!}D^2 - \frac{\kappa_3}{3!}D^3 + \ldots\right\}\beta(x). \tag{6.47}$$

(*) Expansions can also be obtained in terms of derivatives of the Gamma-and Beta-distributions. The appropriate polynomials are those of Laguerre and Jacobi respectively. See G. Szegö, *Orthogonal Polynomials*, New York, 1939, and Durbin and Watson (1951), *Biom.*, **38**, 159, especially p. 172.

Wallace (1958, *Ann. Math. Statist.*, **29**, 635) discusses the validity of asymptotic expansions generally, including non-normal f.f.'s as generating distributions, and also the problem of normalization discussed in **6.25–6** below.

We have retained the terms in D and D^2 because the approximation may perhaps be slightly improved by taking m and σ^2 in the ξ-distribution not quite equal to the mean and variance of x.

We now assume that the cumulant κ_r is of order n^{1-r}, a case of fairly common occurrence; by choice of m and σ^2, we may write

$$\kappa_1 - m = l_1\sigma, \quad l_1 = O(n^{-\frac{1}{2}}),$$
$$\kappa_2 - \sigma^2 = l_2\sigma^2, \quad l_2 = O(n^{-1}).$$

Then σ^2 is of order κ_2, i.e. n^{-1}, and thus

$$\frac{\kappa_r}{\sigma^r} = O(n^{1-\frac{1}{2}r}).$$

Thus (6.47) may be written

$$\exp\{-l_1\sigma D + \tfrac{1}{2}l_2\sigma^2 D^2 - \tfrac{1}{6}l_3\sigma^3 D^3 + \tfrac{1}{24}l_4\sigma^4 D^4 - \tfrac{1}{120}l_5\sigma^5 D^5 + \tfrac{1}{720}l_6\sigma^6 D^6 - \ldots\}\beta(x), \tag{6.48}$$

where l_1 and l_3 are $O(n^{-\frac{1}{2}})$, l_5 is $O(n^{-\frac{3}{2}})$,
l_2 and l_4 are $O(n^{-1})$, l_6 is $O(n^{-2})$, etc.

Expanding the operator and retaining only terms up to and including $O(n^{-2})$ in the l's we find for the operator

$$\begin{aligned}
&1 - l_1\sigma D + \tfrac{1}{2}l_2\sigma^2 D^2 - \tfrac{1}{6}l_3\sigma^3 D^3 + \tfrac{1}{24}l_4\sigma^4 D^4 - \tfrac{1}{120}l_5\sigma^5 D^5 + \tfrac{1}{720}l_6\sigma^6 D^6 + \tfrac{1}{2}(l_1^2\sigma^2 D^2 \\
&+ \tfrac{1}{4}l_2^2\sigma^4 D^4 + \tfrac{1}{36}l_3^2\sigma^6 D^6 + \tfrac{1}{576}l_4^2\sigma^8 D^8 - l_1 l_2\sigma^3 D^3 + \tfrac{1}{3}l_1 l_3\sigma^4 D^4 - \tfrac{1}{12}l_1 l_4\sigma^5 D^5 \\
&+ \tfrac{1}{60}l_1 l_5\sigma^6 D^6 - \tfrac{1}{6}l_2 l_3\sigma^5 D^5 + \tfrac{1}{24}l_2 l_4\sigma^6 D^6 - \tfrac{1}{72}l_3 l_4\sigma^7 D^7 + \tfrac{1}{360}l_3 l_5\sigma^8 D^8) + \tfrac{1}{6}(-l_1^3\sigma^3 D^3 \\
&- \tfrac{1}{216}l_3^3\sigma^9 D^9 + \tfrac{3}{2}l_1^2 l_2\sigma^4 D^4 - \tfrac{1}{2}l_1^2 l_3\sigma^5 D^5 - \tfrac{1}{8}l_1^2 l_4\sigma^6 D^6 + \tfrac{1}{288}l_3^2 l_4\sigma^{10} D^{10} - \tfrac{1}{12}l_1 l_3^2\sigma^7 D^7 \\
&+ \tfrac{1}{24}l_2 l_3^2\sigma^8 D^8 + \tfrac{1}{2}l_1 l_2 l_3\sigma^6 D^6 + \tfrac{1}{24}l_1 l_3 l_4\sigma^8 D^8) + \tfrac{1}{24}(l_1^4\sigma^4 D^4 + \tfrac{1}{1296}l_3^4\sigma^{12} D^{12} \\
&+ \tfrac{2}{3}l_1^3 l_3\sigma^6 D^6 + \tfrac{1}{6}l_1^2 l_3^2\sigma^8 D^8 + \tfrac{1}{54}l_1 l_3^3\sigma^{10} D^{10}).
\end{aligned} \tag{6.49}$$

The result of this operation is a similar expression, which we will not bother to write out at length, with the operator $\sigma^r D^r$ replaced by $(-1)^r \mathrm{H}_r\left(\frac{x-m}{\sigma}\right)$ and multiplied by $\beta(x)$.

The distribution function is given by integrating this expression, and we then have for the frequency less than or equal to $m + \sigma x$ (arranging the terms in order of magnitude in n)

$$\begin{aligned}
&\int_{-\infty}^{x} \alpha(x)\,dx + \alpha(x)[-(l_1 + \tfrac{1}{6}l_3 H_2) - (\tfrac{1}{2}l_1^2 H_1 + \tfrac{1}{2}l_2 H_1 + \tfrac{1}{6}l_1 l_3 H_3 + \tfrac{1}{24}l_4 H_3 \\
&+ \tfrac{1}{72}l_3^2 H_5) - (\tfrac{1}{6}l_1^3 H_2 + \tfrac{1}{2}l_1 l_2 H_2 + \tfrac{1}{12}l_1^2 l_3 H_4 + \tfrac{1}{12}l_2 l_3 H_4 + \tfrac{1}{24}l_1 l_4 H_4 \\
&+ \tfrac{1}{120}l_5 H_4 + \tfrac{1}{72}l_1 l_3^2 H_6 + \tfrac{1}{144}l_3 l_4 H_6 + \tfrac{1}{1296}l_3^3 H_8) - (\tfrac{1}{24}l_1^4 H_3 + \tfrac{1}{8}l_2^2 H_3 + \tfrac{1}{4}l_1^2 l_2 H_3 \\
&+ \tfrac{1}{36}l_1^3 l_3 H_5 + \tfrac{1}{12}l_1 l_2 l_3 H_5 + \tfrac{1}{48}l_1^2 l_4 H_5 + \tfrac{1}{48}l_2 l_4 H_5 + \tfrac{1}{120}l_1 l_5 H_5 + \tfrac{1}{720}l_6 H_5 \\
&+ \tfrac{1}{144}l_1^2 l_3^2 H_7 + \tfrac{1}{144}l_2 l_3^2 H_7 + \tfrac{1}{1152}l_4^2 H_7 + \tfrac{1}{144}l_1 l_3 l_4 H_7 + \tfrac{1}{720}l_3 l_5 H_7 + \tfrac{1}{1296}l_1 l_3^3 H_9 \\
&+ \tfrac{1}{1728}l_3^2 l_4 H_9 + \tfrac{1}{31104}l_3^4 H_{11})].
\end{aligned} \tag{6.50}$$

6.26 Now let ξ be a normal variate. We will determine ξ in terms of x so that

$$G(\xi) = \int_{-\infty}^{\xi} \alpha(y)\,dy = F(x), \tag{6.51}$$

$F(x)$ being the distribution function given by the Type A expansion (6.50).

We have, by Taylor's theorem,

$$G(\xi) = G\{x+(\xi-x)\} = G(x)+\sum_{r=1}^{\infty}\frac{(\xi-x)^r}{r!}\frac{d^r}{dx^r}G(x)$$

and on using (6.21), this becomes

$$G(\xi) = G(x)-\sum_{r=1}^{\infty}\frac{(x-\xi)^r}{r!}H_{r-1}(x)\alpha(x) \tag{6.52}$$

and this is equal to (6.50).

The next step is to invert this series to give $(x-\xi)$ in powers of x. Assuming

$$x-\xi = a_0+a_1x+a_2x^2+\ldots, \tag{6.53}$$

which is (6.46) slightly re-arranged for the sake of convenience, we see that when $x = 0$, $\xi = -a_0$, $x-\xi$ is of order $n^{-\frac{1}{2}}$; and hence, to order n^{-2}, we have from (6.52) with $x = 0$, on using (6.23),

$$\xi-\frac{\xi^2}{2}(+0)+\frac{\xi^3}{6}(-1) = -\left(a_0-\frac{a_0^3}{6}\right)$$

and this is equal to the expression in square brackets in (6.50) with $x = 0$.

We then find

$$a_0 = l_1-\tfrac{1}{6}l_3-\tfrac{1}{2}l_1l_2+\tfrac{1}{6}l_1^2l_3+\tfrac{1}{4}l_2l_3+\tfrac{1}{8}l_1l_4+\tfrac{1}{40}l_5-\tfrac{7}{36}l_1l_3^2-\tfrac{15}{144}l_3l_4+\tfrac{52}{648}l_3^3.$$

We can now find a_1 in (6.50) by identifying coefficients in x, and so on. After some algebraic reduction we find, writing the terms in descending order in n,

$$\begin{aligned}
x-\xi = {} & l_1+\tfrac{1}{6}l_3(x^2-1)+\tfrac{1}{2}l_2x-\tfrac{1}{3}l_1l_3x+\tfrac{1}{24}l_4(x^3-3x)-\tfrac{1}{36}l_3^2(4x^3-7x)-\tfrac{1}{2}l_1l_2 \\
& +\tfrac{1}{6}l_1^2l_3-\tfrac{1}{12}l_2l_3(5x^2-3)-\tfrac{1}{8}l_1l_4(x^2-1)+\tfrac{1}{120}l_5(x^4-6x^2+3)+\tfrac{1}{36}l_1l_3^2(12x^2-7) \\
& -\tfrac{1}{144}l_3l_4(11x^4-42x^2+15)+\tfrac{1}{648}l_3^3(69x^4-187x^2+52)-\tfrac{3}{8}l_2^2x+\tfrac{5}{6}l_1l_2l_3x \\
& +\tfrac{1}{8}l_1^2l_4x-\tfrac{1}{48}l_2l_4(7x^3-15x)-\tfrac{1}{30}l_1l_5(x^3-3x)+\tfrac{1}{720}l_6(x^5-10x^3+15x) \\
& -\tfrac{1}{3}l_1^2l_3^2x+\tfrac{1}{72}l_2l_3^2(36x^3-49x)-\tfrac{1}{384}l_4^2(5x^5-32x^3+35x)+\tfrac{1}{36}l_1l_3l_4(11x^3-21x) \\
& -\tfrac{1}{360}l_3l_5(7x^5-48x^3+51x)-\tfrac{1}{324}l_1l_3^3(138x^3-187x) \\
& +\tfrac{1}{864}l_3^2l_4(111x^5-547x^3+456x)-\tfrac{1}{7776}l_3^4(948x^5-3628x^3+2473x).
\end{aligned} \tag{6.54}$$

This is our required expression of the variate ξ in terms of the variate x. To order n^{-2} at least, ξ will be normally distributed.

It is often more convenient to express x in terms of ξ. This may be done by noting that

$$\begin{aligned}
x-\xi &= g(x) = g(\xi+x-\xi) \\
&= g(\xi)+(x-\xi)g'(\xi)+\ldots \\
&= g(\xi)+g'(\xi)\{g(\xi)+(x-\xi)g'(\xi)\}+\ldots
\end{aligned}$$

and by continuing the process

$$x-\xi = g(\xi)+g(\xi)g'(\xi)+g(\xi)g'^2(\xi)+\tfrac{1}{2}g^2(\xi)g''(\xi)+g(\xi)g'^3(\xi)+\tfrac{3}{2}g^2(\xi)g'(\xi)g''(\xi) + \tfrac{1}{6}g^3(\xi)g'''(\xi)+\ldots \tag{6.55}$$

Hence, using the value of ξ given by (6.54), we find, after some reduction,

$$\begin{aligned} x-\xi = {} & l_1+\tfrac{1}{6}l_3(\xi^2-1)+\tfrac{1}{2}l_2\xi+\tfrac{1}{24}l_4(\xi^3-3\xi)-\tfrac{1}{36}l_3^2(2\xi^3-5\xi)-\tfrac{1}{6}l_2l_3(\xi^2-1) \\ & +\tfrac{1}{120}l_5(\xi^4-6\xi^2+3)-\tfrac{1}{24}l_3l_4(\xi^4-5\xi^2+2)+\tfrac{1}{324}l_3^3(12\xi^4-53\xi^2+17) \\ & -\tfrac{1}{8}l_2^2\xi-\tfrac{1}{16}l_2l_4(\xi^3-3\xi)+\tfrac{1}{720}l_6(\xi^5-10\xi^3+15\xi)+\tfrac{1}{72}l_2l_3^2(10\xi^3-25\xi) \\ & -\tfrac{1}{3840}l_4^2(3\xi^5-24\xi^3+29\xi)-\tfrac{1}{1080}l_3l_5(2\xi^5-17\xi^3+21\xi) \\ & +\tfrac{1}{2880}l_3^2l_4(14\xi^5-103\xi^3+107\xi)-\tfrac{1}{77760}l_3^4(252\xi^5-1688\xi^3+1511\xi). \end{aligned} \tag{6.56}$$

The approach of these sections is due to Cornish and Fisher (1937). The same authors (1960, *Technometrics*, **2**, 209) give higher-order expansions and tables to facilitate their use.

Example 6.4

Consider again the distribution of Example 6.3 :—

$$dF = \frac{1}{\Gamma(\lambda)}e^{-x}x^{\lambda-1}\,dx, \qquad 0\leqslant x\leqslant\infty.$$

We have already found that, in standard measure, this tends to the normal form, and that κ_r is of order $\lambda^{1-\frac{1}{2}r}$.

We will take l_1 and l_2 of (6.48) to be zero, which implies that our normal variate ξ is to have the same mean and variance as that of x. We have

$$l_3 = 2\lambda^{-\frac{1}{2}}, \quad l_4 = 6\lambda^{-1}, \quad l_5 = 24\lambda^{-3/2}, \quad l_6 = 120\lambda^{-2}.$$

(6.50) then becomes

$$\begin{aligned} G(x)-\alpha(x)\Big\{ & \frac{1}{3\lambda^{\frac{1}{2}}}H_2+\frac{1}{4\lambda}H_3+\frac{1}{18\lambda}H_5+\frac{1}{5\lambda^{\frac{3}{2}}}H_4+\frac{1}{12\lambda^{\frac{3}{2}}}H_6+\frac{1}{162\lambda^{\frac{3}{2}}}H_8 \\ & +\frac{1}{6\lambda^2}H_5+\frac{1}{32\lambda^2}H_7+\frac{1}{15\lambda^2}H_7+\frac{1}{72\lambda^2}H_9+\frac{1}{1944\lambda^2}H_{11}\Big\}. \end{aligned}$$

Let us, as a simple illustration, find the distribution function of x for $\lambda = 9$, $x = 12$. The mean of the distribution is then 9 and its variance 9, so that this corresponds to a deviation $(12-9)/\sqrt{9}$ in standard measure, equal to unity. It is found from (6.23) and an additional equation for H_{11} that

$$H_2 = 0,\ H_3 = -2,\ H_4 = -2,\ H_5 = 6,\ H_6 = 16,\ H_7 = -20,$$
$$H_8 = -132,\ H_9 = 28,\ H_{10} = 1216,\ H_{11} = 936.$$

We then find for the distribution function

$$\int_{-\infty}^{1}\alpha(x)\,dx+\alpha(x)(0{\cdot}015{,}163{,}5).$$

The values for the normal function are obtained from the tables and we get the value

$$0{\cdot}841{,}345+(0{\cdot}241{,}970{,}7)(0{\cdot}015{,}163{,}5) = 0{\cdot}8450,$$

which is exact to four places. The approximation is evidently fairly good, even for values of λ as low as 9.

We could have found the same result by using (6.54). Substituting $x = 1$ in that equation we find

$$\xi = 1{\cdot}015{,}386,$$

and the distribution function for the normal integral with deviate equal to this value of ξ is 0·8450 as before.

Suppose now that we wish to find the deviate x whose distribution function is $F(x) = 0{\cdot}99$ when $\lambda = 15$.

The normal deviate ξ corresponding to such a value is found from tables to be 2·326,348. We then have from (6.56)

$$x - \xi = \frac{1}{3\lambda^{\frac{1}{2}}}(\xi^2 - 1) + \frac{1}{4\lambda}(\xi^3 - 3\xi) + \ldots$$

which will be found to give

$$x = 2{\cdot}697{,}22.$$

This is the value in standard measure. The deviate in ordinary measure is

$$15 + x\sqrt{15} = 25{\cdot}45.$$

This is exact to two places of decimals.

The example shows that, notwithstanding the non-convergence of the infinite Type A series, a satisfactory approximation may be obtained from its first few terms, at least in certain cases. We may remark without proof that by an adaptation of a procedure given by Cramér (1928) it may be shown that an asymptotic expansion does in fact exist for the distribution of this example.

Johnson's transformations

6.27 The idea of transforming from one variate to another with a more convenient frequency function may be developed further. We need not confine ourselves to a polynomial expression of the type of (6.46); nor is that kind of representation necessarily the best. Several writers have studied more general types. Our treatment follows that of Johnson (1949*a*).

We shall consider a transformation of the type

$$\xi = \gamma + \delta g\left(\frac{x - \mu}{\lambda}\right) \tag{6.57}$$

where μ, λ, γ and δ are parameters at choice and g is some convenient function. For practical purposes it is desirable that g should not itself depend on variable parameters and that it should be a monotonic function; otherwise it might be troublesome to match the ranges of ξ and x. Without loss of generality we may suppose that g is monotone-increasing. If we write the transformation in the form

$$\frac{\xi - \gamma}{\delta} = g\left(\frac{x - \mu}{\lambda}\right) \tag{6.58}$$

the parameters μ, γ appear as location parameters and λ, δ as scale parameters. We may suppose the latter pair to be positive. We shall try to choose the parameters and the function g so that ξ is normal or approximately so.

6.28 We clearly lose no generality by introducing a standardized variate

$$y = \frac{x-\mu}{\lambda}. \tag{6.59}$$

If ξ is normal with frequency function $\alpha(\xi)$ we have for the frequency function of y

$$\begin{aligned} dF(y) &= \alpha(\xi)\left|\frac{d\xi}{dy}\right| dy \\ &= \frac{1}{\sqrt{(2\pi)}} \exp\left[-\tfrac{1}{2}\{\gamma+\delta g(y)\}^2\right] \delta g'(y)\, dy \end{aligned} \tag{6.60}$$

where $g'(y) = dg(y)/dy$. We shall examine three types of system :—

(1) The S_L or lognormal system : $g(y) = \log y$.
(2) The S_B system : $g(y) = \log\{y/(1-y)\}$.
(3) The S_U system : $g(y) = \sinh^{-1} y = \log\{y+\sqrt{(y^2+1)}\}$.

Other systems, of course, are possible. The variety of shapes given by these three types is, however, fully as great as that of the Pearson system.

The lognormal distribution

6.29 We have

$$\xi = \gamma + \delta \log y. \tag{6.61}$$

As ξ goes from $-\infty$ to ∞, y goes from 0 to ∞. It has infinitely high-order contact at each terminal of the range ; for the distribution, from (6.60), is

$$\begin{aligned} dF(y) &= \frac{\delta}{\sqrt{(2\pi)}\, y} \exp\left\{-\tfrac{1}{2}(\gamma+\delta\log y)^2\right\} dy, \qquad 0 \leqslant y \leqslant \infty \\ &= p(y)\, dy, \text{ say.} \end{aligned}$$

The limit for any positive r

$$\lim_{y\to\infty} y^r p(y) = \lim \frac{\delta e^{-\frac{1}{2}\gamma^2}}{\sqrt{(2\pi)}} \exp\left\{\tfrac{1}{2}(\delta\log y)^2\right\} y^{r-1-\gamma\delta} = 0,$$

and

$$\lim_{y\to 0} y^r p(y) = \lim \frac{\delta e^{-\frac{1}{2}\gamma^2}}{\sqrt{(2\pi)}} \cdot \frac{y^{r-1-\gamma\delta}}{y^{\frac{1}{2}\delta^2}\log y} = 0.$$

Also the distribution is unimodal, for

$$\frac{dp(y)}{dy} = \frac{\delta}{\sqrt{(2\pi)}} \exp\left\{-\tfrac{1}{2}(\gamma+\delta\log y)^2\right\}\left\{-\frac{1}{y^2} - \frac{\delta}{y^2}(\gamma+\delta\log y)\right\},$$

which, apart from the values $y = 0$ and $y = \infty$, vanishes where

$$1+\delta(\gamma+\delta\log y) = 0,$$

namely at one point in the range 0 to ∞.

6.30 For the moments of y about zero we have

$$\begin{aligned}\mu'_r(y) &= \int_0^\infty y^r g(y)\,dy \\ &= \frac{1}{\sqrt{(2\pi)}}\int_{-\infty}^{\infty} \exp\left\{\frac{r(\xi-\gamma)}{\delta}\right\}\exp(-\tfrac{1}{2}\xi^2)\,d\xi \\ &= \exp\left(\frac{1}{2}\frac{r^2}{\delta^2}-\frac{r\gamma}{\delta}\right).\end{aligned} \tag{6.62}$$

Thus we find, putting $\omega = \exp(1/2\delta^2)$, $\rho = \exp(-\gamma/\delta)$,

$$\left.\begin{aligned}\mu'_1 &= \omega\rho \\ \mu_2 &= \omega^2\rho^2(\omega^2-1) \\ \mu_3 &= \omega^3\rho^3(\omega^2-1)^2(\omega^2+2) \\ \mu_4 &= \omega^4\rho^4(\omega^2-1)^2(\omega^8+2\omega^6+3\omega^4-3),\end{aligned}\right\} \tag{6.63}$$

and

$$\beta_1 = (\omega^2-1)(\omega^2+2)^2 \tag{6.64}$$

$$\beta_2-3 = (\omega^2-1)(\omega^6+3\omega^4+6\omega^2+6). \tag{6.65}$$

The skewness is always positive and the distribution is always leptokurtic. In general the frequency curve of y will rise from zero to a maximum and then tail off more slowly as y tends to infinity.

6.31 The values of μ_2 and μ_3 determine ω and ρ, and hence δ and γ. Thus, apart from location and scale parameters, we may fit a lognormal distribution from the second and third moments of the x-distribution.

In point of fact, the lognormal distribution does not have the generality of the family (6.57). It depends on three, not four parameters. For

$$\xi = \gamma+\delta\log\{(x-\mu)/\lambda\}$$

may be written

$$\xi = (\gamma-\delta\log\lambda)+\delta\log(x-\mu)$$

and the term $\gamma-\delta\log\lambda$ can be regarded as a single parameter. It is for this reason that the distribution can be fitted from μ_2 and μ_3 without invoking μ_4. The simplest procedure, suggested by Wicksell (1917), is as follows:

Using (6.64) we find $t \equiv (\omega^2-1)^{\frac{1}{2}}$ as the positive root of

$$t^3+3t-\sqrt{\beta_1} = 0. \tag{6.66}$$

Then the location parameter μ is determined, by using (6.63), as

$$\mu = \mu'_1(x)-\frac{\sqrt{\mu_2}}{t}. \tag{6.67}$$

ρ is found from μ_2 and t and hence γ and δ are found from the definitions of ω and ρ. Yuan (1933) has given tables to facilitate the solution of (6.66).

For a more extensive study of the history, theory and applications of the distribution, reference may be made to J. Aitchison and J. A. C. Brown, *The Lognormal Distribution*, 1957, Cambridge University Press.

Example 6.5

For the purposes of comparison we will fit a lognormal curve to the bean data which have been used in previous examples in this chapter. As we have considered the bean distribution, from low to high values of the variate " length," it is skewed to the left. We will therefore fit a lognormal curve with the variate-range reversed, i.e. so that negative values of ξ correspond to the greater values of x. This does not, in fact, affect the arithmetic process of fitting except at the final stage.

From Example 6.1 we have $\sqrt{\beta_1} = -0{\cdot}910{,}569$. As we are reversing the variate range we change the sign of this, and (6.66) then becomes

$$t^3+3t+0{\cdot}910{,}569 = 0.$$

The positive root is $t = 0{\cdot}294{,}968$. Also we have (as given in Example 6.1) $\mu_2 = 3{\cdot}238{,}425$. Then from the second equation of (6.63) we have $\rho = 5{\cdot}8516$. We notice in passing that from (6.65) we shall then have, for ξ, $\beta_2 = 4{\cdot}510$ as against the β_2 found for x in Example 6.1 $= 4{\cdot}863$. Our transformation happens to bring the values of β_2 roughly into agreement but not completely so.

We then have

$$1/(2\delta^2) = \log_e \omega = \tfrac{1}{2}\log_e(1+t^2) = 0{\cdot}041{,}718$$
$$-\gamma/\delta = \log_e \rho = 1{\cdot}7667$$

giving

$$\delta = 3{\cdot}462, \quad \gamma = -6{\cdot}116.$$

The units (width of class-intervals) are $\frac{1}{2}$ millimetre and hence, from the data of Example 6.1, the variate x has a mean $14{\cdot}5-\frac{1}{2}(0{\cdot}190{,}78) = 14{\cdot}405$. The mean of ξ, from its start, is by (6.67) equal to

$$14{\cdot}405-\sqrt{\mu_2}/t = 17{\cdot}455.$$

Thus the curve " starts " at 17·455 millimetres.

Column (7) of Table 6.1 on page 160 gives the corresponding frequencies. The fit is not particularly good, being rather worse than that of Type IV but better than those of the Gram-Charlier series.

6.32 Curves of the S_B type with $g(y) = \log\{y/(1-y)\}$ have the variate-transformation

$$\xi = \gamma+\delta\log\frac{x-\mu}{\lambda+\mu-x}. \tag{6.68}$$

This may itself be regarded, of course, as a simple transformation of the lognormal $\xi = \gamma+\delta\log(x-\mu)$ with $y = 1-\mu/x$. We may also write

$$\xi = \gamma+\delta\log\{y/(1-y)\} = \gamma+2\delta\tanh^{-1}(2y-1), \tag{6.69}$$

the latter form enabling numerical transformations to be made with the aid of tables of inverse hyperbolic tangents. The frequency function of y, say $p(y)$, is given by

$$p(y) = \frac{\delta}{\sqrt{(2\pi)}}\frac{1}{y(1-y)}\exp\left\{-\tfrac{1}{2}\left(\gamma+\delta\log\frac{y}{1-y}\right)^2\right\}, \qquad 0\leqslant y\leqslant 1. \tag{6.70}$$

It may be shown without difficulty that this has contact of infinitely high order at both terminals of its range.

6.33 The frequency-curves may, however, be bimodal. For the modal values of y we find, on differentiating (6.70),

$$2y-1 = \delta\left(\gamma+\delta\log\frac{y}{1-y}\right).$$

Putting $y = \frac{1}{2}(y'+1)$ we find

$$y' - \gamma\delta = \delta^2\log\frac{1+y'}{1-y'}.$$

The straight line, in Cartesian co-ordinates (y', u), $u = y'-\gamma\delta$, meets the curve $u = \delta^2\log\{(1+y')/(1-y')\}$ in one or three points. In the first case the curve is unimodal, in the second bimodal. It may be shown (we omit the proof) that the necessary and sufficient conditions for bimodality are

$$\delta < 1/\sqrt{2}, \qquad |\gamma| < \delta^{-1}\sqrt{(1-2\delta^2)} - 2\delta\tanh^{-1}\sqrt{(1-2\delta^2)}. \tag{6.71}$$

6.34 The moments of the distribution are very difficult to determine in a manageable form. The simplest method of fitting appears to be to identify with observed quantities certain quantiles, as in the following example.

Example 6.6 (Johnson, 1949*a*)

Consider the data of Table 1.11 (cloudiness at Greenwich). On the assumption that degree of cloudiness stated as 0, 1, 2, . . . can be regarded as groupings −0·5 to 0·5, 0·5 to 1·5, etc., we may fix the location parameter μ at −0·5 and the scale parameter λ at 11. We then choose γ and δ so as to provide exact agreement in the two terminal classes.

Thus, y for the first class is 1/11 and $\log_e\{y/(1-y)\}$ becomes −2·302,585. The

Table 6.2—Frequency-curves of various types fitted to the data of Table 1.11

Degree of cloudiness	Observed frequencies	S_B (1)	S_B (2)	Pearson Type I
0	320	320·0	320·0	321·7
1	129	100·9	120·9	121·5
2	74	73·9	72·0	75·1
3	68	63·8	57·5	61·4
4	45	59·8	52·1	56·0
5	45	59·9	51·6	55·2
6	55	63·4	54·9	57·8
7	65	72·0	63·9	65·5
8	90	90·0	85·5	83·2
9	148	135·4	160·7	139·6
10	676	676·0	676·0	678·0
TOTALS	1715	1715·1	1714·9	1715·0

normal deviate corresponding to a frequency 320/1715 = 0·186,589 is −0·8965 and hence, from (6.69),

$$-0{\cdot}8965 = \gamma - 2{\cdot}3026\delta.$$

Likewise for the frequency at the other end, $y = 10/11$ and the deviate is 0·2684. Hence

$$0{\cdot}2684 = \gamma + 2{\cdot}3026\delta.$$

These two equations yield

$$\gamma = 0{\cdot}3110, \quad \delta = 0{\cdot}25166.$$

Table 6.2 shows the results of fitting this curve, referred to as S_B (1). The curve S_B (2) was fitted in the same way except that it was assumed that the groups were 0 to 0·5, 0·5 to 1·5, etc., and μ was taken as 0, λ as 10. The results of fitting a Pearson Type I curve are also shown. All three curves fail to give sufficiently small frequencies in the trough in the centre of the range.

6.35 The S_U family of curves are defined by

$$\xi = \gamma + \delta \sinh^{-1} y \tag{6.72}$$

and for the frequency function we have

$$p(y) = \frac{\delta}{\sqrt{(2\pi)}} \frac{1}{\sqrt{(y^2+1)}} \exp\left[-\tfrac{1}{2}[\gamma + \delta \log\{y + \sqrt{(y^2+1)}\}]^2\right] \quad -\infty \leqslant y \leqslant \infty. \tag{6.73}$$

Again there is high-order contact at the ends of the distribution. It may be verified without much difficulty that the curves are unimodal and that the mode lies between the median and zero. Thus the distribution is positively or negatively skewed according as γ is negative or positive.

For the moments we have

$$\mu_r' = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} \exp(-\tfrac{1}{2}\xi^2) \cdot \frac{1}{2^r} \left\{ \exp\left(\frac{\xi-\gamma}{\delta}\right) - \exp\left(-\frac{\xi-\gamma}{\delta}\right) \right\}^r d\xi,$$

leading to

$$\left.\begin{aligned} \mu_1' &= -\omega \sinh \Omega \\ \mu_2 &= \tfrac{1}{2}(\omega^2-1)(\omega^2 \cosh 2\Omega + 1) \\ \mu_3 &= -\tfrac{1}{4}\omega^2(\omega^2-1)^2\{\omega^2(\omega^2+2)\sinh 3\Omega + 3\sinh\Omega\} \\ \mu_4 &= \tfrac{1}{8}(\omega^2-1)^2\{\omega^4(\omega^8+2\omega^6+3\omega^4-3)\cosh 4\Omega \\ &\qquad + 4\omega^4(\omega^2+2)\cosh 2\Omega + 3(2\omega^2+1)\} \end{aligned}\right\} \tag{6.74}$$

where $\omega = \exp(1/2\delta^2)$, $\Omega = \gamma/\delta$. The method of fitting is illustrated in the following example.

Example 6.7

Consider once more the distribution of beans by length which has served in previous examples. For the observed distribution we have, in working units of half a millimetre,

$$\text{mean} = 14{\cdot}405 \text{ mm.}, \quad \beta_1 = 0{\cdot}829,$$
$$\mu_2 = (0{\cdot}9036)^2, \quad \beta_2 = 4{\cdot}863.$$

We could substitute in equations (6.74) and solve for ω and Ω. This is a tedious process and Johnson (1949*a*) constructed an abac for the purpose. He gives the values

$\delta = 2{\cdot}64$, $\gamma = 2{\cdot}38$. With these values we find, using the relations

$$\text{mean } \xi = \mathrm{E}\{(x-\mu)/\lambda\} = 1{\cdot}1029,$$
$$\text{var } (x/\lambda) = (0{\cdot}5948)^2,$$

the values of u and λ as

$$\lambda = 1{\cdot}5192, \text{ mean } \xi = 14{\cdot}399 + 1{\cdot}1029\lambda = 16{\cdot}0745.$$

Column (8) of Table 6.1 on page 160 shows the frequencies given by an S_U curve with these values of the constants. The fit seems about as good as the Pearson Type IV distribution.

6.36 The fitting of mathematical curves to observational data has a certain intrinsic interest which is apt to outrun its statistical usefulness. In order to keen this chapter within bounds we shall refer only briefly to a number of further topics which have been discussed in the literature, leaving the student who wishes to specialize his knowledge to follow the references.

Burr's distributions

6.37 The three systems we have considered in this chapter all attempt to fit curves to the *frequency* function ; and we have seen that in some cases, at least, this involves troublesome summation or integration when we require the *distribution* function, as we usually do, if only to calculate theoretical frequencies for comparison with observation. The question naturally arises whether we could not fit curves direct to the distribution function and obtain the frequency function, if we need it, by the relatively simple process of differentiation.

Such an approach has been considered by Burr (1942) and Hatke (1949). Using a generalization of the Pearson equation (6.1), we consider

$$dF = F(1-F)g(x)\,dx, \tag{6.75}$$

where $g(x)$ is some convenient function, which must be non-negative in $0 \leqslant F \leqslant 1$ and the range of x. The solution of (6.75) is obtained at once, since

$$dF\left(\frac{1}{F}+\frac{1}{1-F}\right) = g(x)\,dx$$

and is given by

$$F(x) = [1+\exp\{-G(x)\}]^{-1}, \tag{6.76}$$

where

$$G(x) = \int_{-\infty}^{x} g(t)\,dt.$$

For example, one convenient form is

$$F(x) = 1-\frac{1}{(1+x^{\alpha})^{\beta}}, \qquad 0 \leqslant x \leqslant \infty, \tag{6.77}$$

and many others can be chosen.

6.38 Burr's own method of fitting was to consider *cumulative moments* typified by

$$M_j(a) = \int_a^\infty (x-a)^j \{1-F(x)\}\,dx - \int_{-\infty}^a (x-a)^j F(x)\,dx \tag{6.78}$$

and to equate these quantities, calculated for the data, to the theoretical values expressed in terms of the parameters of the distribution ; in fact, to do with cumulative moments what Pearson did with ordinary moments. This can be tedious, although tables were provided by Burr and Hatke for the purpose. The simpler method of fitting in the manner of Example 6.6, by equating frequencies in certain ranges, does not seem to have been tried.

Fitting by frequency-moments

6.39 Fitting frequency-distributions by the method of moments has certain disadvantages, especially when the range is unlimited and observed data are only a sample from the possible observations ; outlying members and " tail " frequencies exert an undue effect on the calculated moments. It is possible to fit by identifying moments taken about the x axis, that is to say, by equating observed and theoretical values of the so-called *probability-moments* typified by

$$\omega_r = \int_{-\infty}^\infty \{f(x)\}^r\,dx. \tag{6.79}$$

This method has been examined at length by Sichel (1949), especially in one of the cases where the Pearson distributions are somewhat inadequate. The point concerning efficiency in estimation, made in **6.12** for Pearson distributions, applies here also. From the standpoint adopted in the present chapter, the appropriate method would be to equate the observed and theoretical values of (6.79). The probability-moments, however, require grouping corrections and the method has not been tried out extensively.

Moriguti (1952, *Ann. Math. Statist.*, **23,** 286) has shown that for any continuous distribution with variance σ^2

$$\sigma^{r-1}\,\omega_r \geqslant \frac{2r}{3r-1}\left[\left(\frac{r-1}{3r-1}\right)^{\frac{1}{2}} \Big/ B\left(\frac{r}{r-1}, \frac{1}{2}\right)\right]^{r-1}.$$

Truncated forms

6.40 We have referred briefly to the fact that some observed distributions may be fitted by mathematical expressions which extend over non-observed values of the variate ; or, to put the matter another way, in which the data may be fitted to a truncated theoretical distribution. Fitting in such cases may be very troublesome, because in general the point of truncation has to be estimated and the moments will depend upon it. Systematic developments of this topic have only recently begun and have tended to consider individual distributions separately.

Bivariate forms

6.41 No satisfactory method has yet been found of setting up families of (bivariate) frequency-surfaces in extension of the families of frequency-curves we have considered in this chapter. Some early efforts are referred to in Pretorius (1930) and Johnson (1949*b*). A bivariate Gram-Charlier series may also be developed—cf. Exercises 6.16 and 6.17. If this subject were thought to be worth pursuing, perhaps the best line of development would be to seek variate-transformations which transformed the original variates to approximate independence; univariate forms could then be fitted to the marginal frequencies.

EXERCISES

6.1 Show that the following are members of the Pearson system of curves and sketch them for some typical values of the constants:

(a) (the normal distribution)

$$f = \frac{1}{\sigma\sqrt{(2\pi)}} \exp\left(-\frac{x^2}{2\sigma^2}\right), \quad -\infty \leqslant x \leqslant \infty.$$

(b) (the Type II distribution)

$$f = \frac{1}{aB(\frac{1}{2}, m+1)}\left(1-\frac{x^2}{a^2}\right)^m, \quad -a \leqslant x \leqslant a.$$

(c) (the Type VII distribution)

$$f = \frac{1}{aB(\frac{1}{2}, m-1)}\left(1+\frac{x^2}{a^2}\right)^{-m}, \quad -\infty \leqslant x \leqslant \infty.$$

6.2 With a simple transformation of the variate when necessary, assign the following distributions to one of Pearson's types:—

$$dF = ke^{-\frac{1}{2}\chi^2}\chi^{\nu-1}\,d\chi, \qquad dF = k(1-r^2)^{\frac{1}{2}(n-4)}\,dr,$$

$$dF = \frac{k\,dt}{(1+t^2/\nu)^{\frac{1}{2}(\nu+1)}}, \qquad dF = k\eta^{p-2}(1-\eta^2)^{\frac{1}{2}(N-p-2)}\,d\eta.$$

(All these distributions are important in the theory of sampling.)

6.3 Show that for the Pearson distributions

$$\frac{d\log f}{dx} = \frac{x}{B_0+B_1x+B_2x^2}$$

the range is unlimited in both directions if $B_0+B_1x+B_2x^2$ has no real roots; limited in one direction if the roots are real and of the same sign; and limited in both directions if the roots are real and of opposite sign.

6.4 Show that the Pearson Type VI curve may be written

$$y = y_0\left(1-\frac{x^2}{a^2}\right)^{-m} \exp\{-\nu \tanh^{-1}(x/a)\}$$

and discuss the relationship with the Type IV curve.

6.5 Show that for a Pearson Type III distribution $2\beta_2-3\beta_1-6 = 0$ and hence that the criterion κ is infinite.

6.6 If the roots of $B_0+B_1X+B_2X^2$ are equal, show that the criterion $\kappa = 1$ and that the corresponding Pearson curve (Type V) may be written

$$f = \frac{\gamma^{p-1}}{\Gamma(p-1)} x^{-p} e^{-\gamma/x}, \qquad 0 \leqslant x \leqslant \infty.$$

6.7 Show that for a Pearson distribution $\frac{df}{f} = \frac{(a+x)\,dx}{b_0+b_1x+b_2x^2}$ the characteristic function obeys the relation

$$b_2\theta\frac{d^2\phi}{d\theta^2}+(1+2b_2+b_1\theta)\frac{d\phi}{d\theta}+(a+b_1+b_0\theta)\,\phi = 0,$$

where $\theta = it$. Deduce the recurrence relation between the moments.

Show also that the cumulant-generating function obeys the relation

$$b_2\theta\left\{\frac{d^2\psi}{d\theta^2}+\left(\frac{d\psi}{d\theta}\right)^2\right\}+(1+2b_2+b_1\theta)\frac{d\psi}{d\theta}+(a+b_1+b_0\theta)=0.$$

Hence show that the cumulants obey the recurrence relations

$$1+(r+2)\,b_2\}\,\kappa_{r+1}+rb_1\kappa_r+rb_2\left\{\kappa_1\,\kappa_r+\binom{r-1}{1}\kappa_2\kappa_{r-1}+\binom{r-1}{2}\kappa_3\kappa_{r-2}+\ldots\right.$$
$$\left.+\binom{r-1}{j}\kappa_{j+1}\kappa_{r-j}+\ldots+\binom{r-1}{1}\kappa_{r-1}\kappa_2+\kappa_r\,\kappa_1\right\} = 0.$$

6.8 (The Charlier Type B series.) Defining

$$\gamma(\lambda,x) = e^{-\lambda}\lambda^x/x!, \qquad \nabla\gamma(\lambda,x-1) = \gamma(\lambda,x)-\gamma(\lambda,x-1), \quad \text{and}$$
$$G_r(\lambda,x) = \left\{\frac{d^r}{d\lambda^r}\gamma(\lambda,x)\right\}\Big/\gamma(\lambda,x)$$

show that

$$G_r(\lambda,x) = \{(-1)^r\,\nabla^r\gamma(\lambda,x)\}/\gamma(\lambda,x).$$

$$\sum_{x=0}^{\infty}(G_rG_s\gamma) = 0, \quad r\neq s, \quad \text{and} \quad = r!/\lambda^r, \quad r = s.$$

Hence, if f can be expressed as a sum

$$f = \sum_j (b_j G_j\gamma)$$

show that

$$b_j = \frac{\lambda^j}{j!}\sum_{x=0}^{\infty}(fG_j).$$

6.9 Show that no distribution which is not completely determined by its moments can be expanded in a convergent Type A series.

6.10 Show that if y is a function of x which it is desired to represent approximately by the form

$$y = \sum_{j=0}^{r} c_j H_j(x)\,\alpha(x),$$

then the values of the c's appropriate to the expansion of y in this form are such as to minimize the sum

$$\int_{-\infty}^{\infty}\left\{\frac{y-\sum_{j=0}^{r}c_jH_j(x)\,\alpha(x)}{\sqrt{\alpha(x)}}\right\}^2 dx.$$

6.11 Show that if a function in standard measure is expanded in a Type A series the coefficients of the second and third terms depend respectively on β_1 and β_2 and thus provide measures of skewness and kurtosis.

6.12 If $f(x)$ and $g(x)$ are two differentiable frequency functions with cumulants denoted respectively by κ and κ' show that formally

$$f(x) = \exp\left\{\sum_{j=0}^{\infty} \frac{\kappa_j - \kappa_j'}{j!}\left(-\frac{d}{dx}\right)^j\right\} g(x).$$

6.13 If y is distributed in the form

$$dF = \frac{1}{\Gamma(\lambda)} y^{\lambda-1} e^{-y}\, dy, \qquad 0 \leqslant y \leqslant \infty,$$

and a new variate ξ is defined by

$$\xi = \gamma + \delta \log y,$$

show that for the cumulants of ξ we have

$$\kappa_1(\xi) = \gamma + \delta \frac{d}{d\lambda} \log \Gamma(\lambda)$$

$$\kappa_r(\xi) = \delta^r \frac{d^r}{d\lambda^r} \log \Gamma(\lambda), \qquad r > 1.$$

Hence show that

$$\beta_1(\xi) = \frac{1}{\lambda} + O(\lambda^{-2})$$

$$\beta_2(\xi) = 3 + \frac{2}{\lambda} + O(\lambda^{-2}).$$

Compare with the corresponding coefficients for y,

$$\beta_1(y) = \frac{4}{\lambda}$$

$$\beta_2(y) = 3 + \frac{6.}{\lambda}$$

(Johnson, 1949*a*)

6.14 If y is distributed in the form

$$dF = \frac{\Gamma(\tau)}{\Gamma(\tau-\nu)\,\Gamma(\nu)} y^{\nu-1}(1+y)^{-\tau}, \qquad 0 \leqslant y \leqslant \infty,$$

show that if $\xi = \gamma + \delta \log y$ we have, for large ν, $\tau - \nu$,

$$\beta_1(\xi) = \frac{1}{\nu} - \frac{4}{\tau} + \frac{1}{\tau-\nu}$$

$$\beta_2(\xi) = 3 + \frac{2}{\nu} - \frac{6}{\tau} + \frac{2}{\tau-\nu}.$$

Compare with

$$\beta_1(y) = \frac{4}{\nu} - \frac{4}{\tau} + \frac{16}{\tau-\nu}$$

$$\beta_2(y) = 3 + \frac{6}{\nu} - \frac{6}{\tau} + \frac{30}{\tau-\nu}.$$

(Johnson, 1949*a*)

6.15 If R is distributed in the form

$$dF = \frac{\{\Gamma(\frac{1}{2}n-1)\}^2}{\Gamma(n-2)} R^{\frac{1}{2}(n-4)}(1-R)^{\frac{1}{2}(n-4)}, \qquad 0 \leqslant R \leqslant 1,$$

show that for

$$\xi = \tfrac{1}{2}\sqrt{(n-3)}\log\{R/(1-R)\}$$

$$\beta_1(\xi) = 0$$

$$\beta_2(\xi) = 3 + \frac{\tfrac{1}{2}\left[\dfrac{d^2}{d\lambda^2}\log\Gamma(\lambda)\right]_{\lambda=\frac{1}{2}n-1}}{\left[\log\Gamma(\lambda)\right]^2_{\lambda=\frac{1}{2}n-1}}.$$

(Johnson, 1949*a*)

6.16 If x, y are jointly distributed in the normal form

$$f(x,y) = \frac{1}{2\pi(1-\rho^2)^{\frac{1}{2}}}\exp\left\{-\frac{1}{2(1-\rho^2)}(x^2-2\rho xy+y^2)\right\}, \qquad -\infty \leqslant x,\ y \leqslant \infty,$$

and $D_1 = d/dx$, $D_2 = d/dy$, show that the characteristic function of

$$\exp\left\{\kappa_{11}D_1D_2 + \sum_{r+s\geqslant 3}(-1)^{r+s}\kappa_{rs}\frac{D_1^r D_2^s}{r!\ s!}\right\}\alpha(x)\alpha(y) \tag{A}$$

is the exponential of

$$-\tfrac{1}{2}(t_1^2+2\rho t_1t_2+t_2^2) + \sum_{r+s\geqslant 3}\kappa_{rs}\frac{(it_1)^r}{r!}\frac{(it_2)^s}{s!}.$$

Hence show that the bivariate form of the Edgeworth series may be written either as at (A) or in the form

$$\exp\left\{\sum_{r+s\geqslant 3}(-1)^{r+s}\kappa_{rs}\frac{D_1^r D_2^s}{r!\ s!}\right\}f(x,y). \tag{B}$$

(Kendall, 1949*b*)

6.17 From (A) of the previous exercise show that formally

$$\frac{\partial f}{\partial \kappa_{rs}} = (-1)^{r+s}\frac{D_1^r D_2^s}{r!\ s!}f, \qquad r+s\geqslant 3, \text{ or } r=s=1,$$

and verify directly this formula for the bivariate normal distribution for $r = s = 1$, namely

$$\frac{\partial f}{\partial \rho} = \frac{\partial^2 f}{\partial x\,\partial y}.$$

(Kendall, 1949*b*)

6.18 Verify that the S_U curves of **6.35** are unimodal and show that they are positively or negatively skewed according as γ is negative or positive.

6.19 Show that the following distribution functions are admissible solutions of equation (6.76), the parameters in all cases being positive:—

$$F(x) = (1+e^{-x})^{-\alpha}, \qquad -\infty \leqslant x \leqslant \infty;$$

$$F(x) = (1+\alpha e^{-\tan x})^{-\alpha}, \qquad -\tfrac{1}{2}\pi \leqslant x \leqslant \tfrac{1}{2}\pi;$$

$$F(x) = \left(\frac{2}{\pi}\arctan e^{-x}\right)^{\alpha}, \qquad -\infty \leqslant x \leqslant \infty;$$

$$F(x) = \left(x - \frac{1}{2\pi}\sin 2\pi x\right)^{\alpha}, \qquad 0 \leqslant x \leqslant 1.$$

(Burr, 1942)

6.20 A variate has zero mean, unit variance and the rth cumulant κ_r is of order $n^{1-\frac{1}{2}r}$, $r \geqslant 2$. By considering an Edgeworth expansion show that the mode, to order n^{-1}, is given by $-\frac{1}{2}\kappa_3$ and the median by $-\frac{1}{6}\kappa_3$. Hence derive the approximate relationship

$$(\text{median} - \text{mode})/(\text{mean} - \text{mode}) = \tfrac{2}{3}.$$

(Haldane, 1942)

CHAPTER 7

THE CALCULUS OF PROBABILITIES

7.1 The previous six chapters have dealt with the theory of statistical distributions from the descriptive point of view. We have seen that the distributions encountered in practice exhibit certain regular features which permit of representation by mathematical forms; that they can be characterized by certain quantities such as moments and cumulants; and that certain general theorems about distribution- and frequency-functions can be deduced. We now begin a study of a different kind, namely the inquiry whether any meaningful and objective statements can be made about a population or a system generating a population when only a sample of the possible observations is available for scrutiny. This, in broad terms, is the problem of statistical inference.

7.2 Except in trivial cases it is not possible to make inferences from sample to population with the certainty of deductive logic. Either our statements must be somewhat vaguely phrased or, if precise, they are surrounded by a band of doubt. In ordinary speech we use various words and phrases to denote the attitude of mind towards propositions of whose truth we are uncertain. We speak of something as being more or less *probable*, or more or less *likely*; we say that we have a certain degree of *confidence* in a proposition or that the *chances* are against it; we refer to it as being surrounded by or open to *doubt* or requiring *confirmation*. Many of these expressions have acquired a technical meaning in statistics which we shall indicate in due course. Fundamentally they all express aspects of the same thing, the uncertainty with which we are compelled by circumstances to regard particular propositions.

7.3 In the next chapter we shall discuss some of these ideas in relation to statistical inference. Before we do so, however, we must give a brief account of a calculus which is required in order to treat them objectively and systematically: the calculus of probabilities.

Some writers use the expression " the theory of probability " to cover all uncertain inference. Others attempt to relate it only to sets of repeatable events. The first approach takes " probability " to be an undefined idea, like the straight line of Euclidean geometry, and builds up the theory from certain axioms. The second usually seeks to define probability in terms of the relative frequency of events and thus to refer the theory to the pure mathematics of abstract sets or to the limiting properties of sequences. Our own opinion (cf. Kendall, 1949*a*) is that neither approach is sufficient by itself for the statistician, whatever may be its value to the mathematician. It seems, however, that every man must choose for himself and that his psychological make-up, his experience and his fields of interest all determine the kind of axiomatization which he prefers. In statistics it is a mark of immaturity to argue overmuch about the fundamentals of probability theory.

7.4 We are not, in this chapter, concerned with the theory of probability in the broad sense. That is a branch of scientific methodology. We are reviewing the calculus of probabilities, which is a branch of mathematics. The English language is to some extent responsible for a confusion between the two. In medieval times and at the Renaissance there was a distinction between *probabilitas*, which related to the intensity of degrees of belief, and the *doctrine of chances*, which calculated the number of ways in which certain classes of event could occur. In the seventeenth and eighteenth centuries the doctrine of chances became assimilated to " probability," so that nowadays the same word is used to denote both. A useful distinction has been lost. We may well admit " probability " as including chances ; but we shall continue to distinguish between the *theory of probability*, which is concerned with the use of probabilities in making statements about the external world, and the *calculus of probabilities*, which is concerned with the mathematical development of the consequences of certain axioms and postulates. It is the latter which we proceed to describe.

7.5 We shall enunciate the basic rules of the calculus without attempting to derive them from more primitive propositions. Fortunately, whatever the disagreements about fundamentals, there is little or no disagreement about the rules themselves.

In the first place it is assumed that probability is measurable on a continuous scale, so that any probability can be expressed as a real number. We shall, in fact, say that a probability *is* x, a real number. This assumption implies, among other things, that any two probabilities may be compared ; for if they are measured by the numbers x and y we may say that the probability of the first is greater than, equal to, or less than, that of the second according as $x>y$, $x = y$, or $x<y$.

7.6 The probability of a proposition q on data p is written $\mathrm{P}(q|p)$. We have then *Rule 1* :

$$\text{If } p \text{ entails } q, \qquad \mathrm{P}(q|p) = 1. \tag{7.1}$$

$$\text{If } p \text{ entails not-}q, \quad \mathrm{P}(q|p) = 0. \tag{7.2}$$

This rule defines the end-points of our scale of probability. Certainty that a proposition is not true is represented by zero, certainty that it is true by unity. Any probability lies in the range 0 to 1.

7.7 *Rule 2.* If the probabilities of n mutually exclusive propositions $q_1, \ldots, q_n$ on data p are $\mathrm{P}_1, \ldots, \mathrm{P}_n$, then the probability on data p that one of them is true is $\mathrm{P}_1+\mathrm{P}_2 + \ldots +\mathrm{P}_n$.

This is generally known as the " theorem of the addition of probabilities." In the language of the textbooks, the probability that one of n mutually exclusive events will happen is the sum of their separate probabilities.

7.8 It follows that if $q_1, \ldots, q_n$ are mutually exclusive and equally probable on data p, and if Q is a subset of m of them, and if one of the q's must hold, the probability that it is one of the subset Q is

$$P(Q|p) = m/n. \tag{7.3}$$

This proposition is the starting point of the frequency theory of probability. It is usually stated in some such form as : if, of a set of n mutually exclusive and equally probable events, m are distinguished by some characteristic A, the probability of an event bearing A is m/n. As a *definition* of probability it is open to the logical objection of being circular, for it contains the phrase " equally probable."

7.9 *Rule 3.* The probability of two propositions q and r on data p is the product of the probability of q given p and that of r given q and p. Symbolically,

$$\mathrm{P}(qr\,|\,p) = \mathrm{P}(q\,|\,p)\,\mathrm{P}(r\,|\,qp). \tag{7.4}$$

Since q and r appear symmetrically we also have

$$\mathrm{P}(qr\,|\,p) = \mathrm{P}(r\,|\,p)\,\mathrm{P}(q\,|\,rp). \tag{7.5}$$

From the frequency standpoint this rule is almost self-evident. If of a set n, (a) bear the characteristic A, (b) the characteristic B, and (ab) both characteristics, then the rule states that

$$\frac{(ab)}{n} = \frac{(a)}{n}\frac{(ab)}{(a)} = \frac{(b)}{n}\frac{(ab)}{(b)},$$

a simple arithmetical proposition.

More generally we have

$$\mathrm{P}(q_1 q_2 \ldots q_k\,|\,p) = \mathrm{P}(q_1\,|\,p)\,\mathrm{P}(q_2\,|\,q_1,p)\,\mathrm{P}(q_3\,|\,q_2,q_1,p)\ldots\mathrm{P}(q_k\,|\,q_{k-1},\ldots,q_1,p), \tag{7.6}$$

a result which follows from the repeated application of Rule 3.

If, as a particular case,

$$\mathrm{P}(qr\,|\,p) = \mathrm{P}(q\,|\,p)\,\mathrm{P}(r\,|\,p) \tag{7.7}$$

we have, in virtue of (7.4),

$$\mathrm{P}(r\,|\,p) = \mathrm{P}(r\,|\,q,p) \tag{7.8}$$

and q is then said to be irrelevant to r, given p. A knowledge of q does not affect the probability of r on data p.

7.10 The above three rules and various elaborations of them form the basis of the direct calculus of probability ; everything further is pure mathematics. We assume that the reader has some acquaintance with the elements of the subject and give a few examples mainly for the purpose of emphasizing the importance of keeping strictly to the rules.

Example 7.1 Importance of specifying the basic reference-set of probabilities

Three pennies are tossed. What is the probability that they fall either all heads or all tails ?

We assume that the probability of a head with any penny is $\frac{1}{2}$ and that the result with one penny is independent of that with the others. Then there are eight possible and equiprobable cases, *HHH*, *HHT*, *HTH*, *HTT*, *THH*, *THT*, *TTH*, *TTT*. Two of these give us all heads or all tails and hence the required probability is $\frac{1}{4}$.

Now consider this argument: there are two possibilities, either the three coins

all fall alike or two of them are alike and the other different. Of these two possibilities one is of the type required and therefore the probability is $\frac{1}{2}$.

Consider also this argument: there are four possibilities, three heads, two heads and a tail, two tails and a head, three tails. Two of these four are of the type required and therefore the probability is $\frac{1}{2}$.

Finally, consider this argument: of the three coins, two must fall alike. The other must either be the same as these two or different. Thus there are two possibilities and again the chance is $\frac{1}{2}$.

These three arguments are fallacious. They assume equiprobability among events which are not equiprobable and the application of Rule 2 is not legitimate. For example, in the first case, it is true that there are two possibilities, but they are not equally probable under our assumptions. The reader may care to examine why this is so and how the other two arguments break down on the same point.

Example 7.2

In a hand at bridge, what is the probability that a player has at least two aces in his hand? The question as posed is not unambiguous, and there are in fact four distinct probabilities, in any of which we may be interested:

(a) The probability that a specified player (say, South) has at least two aces, irrespective of the other players' hands.

(b) The probability that some one player of the four has at least two aces, irrespective of the other three players' hands.

(c) The probability that South and no other player has at least two aces.

(d) The probability that exactly one player of the four has at least two aces.

We proceed to evaluate these in turn.

(a) South may have 0, 1, 2, 3 or 4 aces. The pack of 52 cards may be regarded as consisting of a pack of 4 aces and a pack of 48 other cards. The probability that South gets no ace is

$$p_0 = \frac{\binom{4}{0}\binom{48}{13}}{\binom{52}{13}}, \tag{7.9}$$

and the probability of his getting exactly one ace is

$$p_1 = \frac{\binom{4}{1}\binom{48}{12}}{\binom{52}{13}}.$$

The probability we require is

$$\begin{aligned} p_a &= 1-(p_0+p_1) \\ &= 1-\frac{38.37.11}{49.25.17} = 0{\cdot}257. \end{aligned} \tag{7.10}$$

(b) It is tempting to imagine that the second of our probabilities is simply four times the first, but the four players' probabilities are not mutually exclusive. This is

otherwise clear from the fact that $4p_a > 1$. In fact, the probability we now seek is simply the complement of the probability that each player has one ace, and is

$$p_b = 1 - 4! \frac{\binom{48}{12}\binom{36}{12}\binom{24}{12}\binom{12}{12}}{\binom{52}{13}\binom{39}{13}\binom{26}{13}\binom{13}{13}},$$

the factor 4! arising from the possible permutation of the four aces among the players. We find

$$p_b = 1 - \frac{13^3}{49.25.17} = 0{\cdot}895. \tag{7.11}$$

(c) The third probability is less than p_a by the probabilities that (South, North), (South, East) or (South, West) each have two aces. These three events are mutually exclusive, and each has probability equal to

$$p_{2,2} = \binom{4}{2} \frac{\binom{48}{11}\binom{37}{11}\binom{26}{13}\binom{13}{13}}{\binom{52}{13}\binom{39}{13}\binom{26}{13}\binom{13}{13}} = 0{\cdot}02248,$$

the factor $\binom{4}{2}$ arising from the possible distributions of the four aces into two pairs. The probability we require is

$$p_c = p_a - 3p_{2,2} = 0{\cdot}190. \tag{7.12}$$

(d) The fourth probability is the sum of the probabilities of four mutually exclusive events, each with probability p_c. Thus

$$p_d = 4p_c = 0{\cdot}759. \tag{7.13}$$

Example 7.3 Inductive methods

Our first two examples enumerate the basic possibilities directly. A large class of examples for which direct enumeration may be difficult can be solved by inductive or semi-inductive methods. We will illustrate one such method on a very old problem.

n letters, to each of which corresponds an envelope, are placed in the envelopes at random. What is the probability that no letter is placed in the right envelope?

The condition that the letters are put in the envelopes " at random " is to be interpreted as meaning that every possible way of assigning the letters to envelopes is equally probable. The question, under Rule 2, then reduces to the purely algebraic one: in what proportion of the possible cases does no letter get into the right envelope?

Suppose that u_n is the number of ways in which all the letters can go wrong. Consider any particular letter. If this occupies another's envelope and vice versa, which can happen in $n-1$ ways, the number of ways in which the remaining $n-2$ letters can go wrong is u_{n-2}. But if the letter occupies another's place and not vice versa, which can happen in $n-1$ ways, there are u_{n-1} ways in which others can go wrong.

Hence we have the difference equation

$$u_n = (n-1)(u_{n-1}+u_{n-2}).$$

We may re-write this

$$u_n - nu_{n-1} = -\{u_{n-1}-(n-1)u_{n-2}\}$$

and putting $v_n = u_n - nu_{n-1}$,

we find $v_n = -v_{n-1}$,

$$= (-1)^{n-2}v_2.$$

Thus $u_n - nu_{n-1} = (-1)^{n-2}(u_2 - 2u_1)$.

But $u_1 = 0$ and $u_2 = 1$ and thus

$$u_n - nu_{n-1} = (-1)^n$$

whence $$u_n = n!\left\{\frac{1}{2!}-\frac{1}{3!}+\ldots+\frac{(-1)^n}{n!}\right\}.$$

The total number of possible ways of arranging the letters is $n!$ and hence the probability required is

$$\frac{1}{2!}-\frac{1}{3!}+\ldots+\frac{(-1)^n}{n!}, \tag{7.14}$$

i.e. (7.14) is the first $n+1$ terms of e^{-1}, the first two cancelling each other and not appearing.

Probability in a continuum

7.11 Up to this point we have considered only probabilities of finite and discrete events; but we may also ask whether any meaning can be attached to probabilities in a continuum. For example, if a square is inscribed in a circle, what is the probability that a point taken at random in the circle is also inside the square? If a line is divided into three segments, what is the probability that they can form a triangle? What is the probability that $x < x_0$ where x is a positive real number less than y_0? And so on.

All probabilities of this kind must be considered as limits. Consider the first example, that of the square inscribed in the circle. Imagine the whole figure divided into small cells of area ε by a rectangular mesh. If we assume that the occurrence of a point in a cell is equally probable for all cells, the probability that a point falls inside both circle and square is the ratio of the number of cells in the latter to those in the former, neglecting the cells at the edges which become of diminishing importance as $\varepsilon \to 0$. In fact, the required probability can be made as near the ratio of the area of the square to that of the circle as we please by taking ε small enough. We may say that the probability *is* that ratio, which is easily seen to be $2/\pi$, an incommensurable number.

We should get the same limiting form of probability if we took other meshes which adequately represented areas; but it is most important to specify the method of proceeding to the limit in speaking of probabilities in a continuum. Otherwise the result has no meaning. The following example will illustrate the point.

Example 7.4

Consider a straight line OA bisected at B. What is the probability that a point chosen at random on the line falls into the segment OB?

Let us suppose in the first place that the line is divided into n equal segments of length OA/n. If we interpret the choosing of a point at random to mean the choice of one of these intervals, the probability is obviously $\frac{1}{2}$ as $n \to \infty$, for there will be half the intervals in the segment OB.

Now let OP be drawn perpendicular to OA and equal to it in length, and imagine a star of $n+1$ lines drawn through P, including OP and PA, so as to divide the angle OPA $(= \pi/4)$ into equal angles $\pi/(4n)$. These lines cut off segments on OA, and we may, if we regard equal angles as having equal probability, assign to these segments an equal probability, for they subtend equal angles at P. If we make this convention it is evident that as $n \to \infty$ the probability of a point falling into any segment on OA is proportional to the angle subtended at P. For example, the probability that a point falls in the segment OB is $(\arctan \frac{1}{2})/(\frac{1}{4}\pi)$.

Now this is not the same answer that we got by assuming all small segments of OA equally probable. There is nothing contradictory in this—the two answers are different because the two limiting processes were different. On a little reflection it will be clear that by moving the point P on the perpendicular to OA and taking a star of lines as before we can make the probability of obtaining a point in OB have any value we like. It is thus abundantly clear that the concept of probability in a continuum depends on the limiting process by which that continuum is reached from a finite subdivision of equiprobable intervals.

The concept of the random variable

7.12 Suppose we have a statistical population, finite and discontinuous, distributed according to a variate x. If we take a member at random from this population, the probability that it bears an assigned variate-value x_0 is the frequency function $f(x_0)$, for this is the proportion of members bearing that value. Further, the probability that it bears a value less than or equal to x_0 is the distribution function $F(x_0)$, as follows at once from Rule 2 and the definition of the distribution function.

This is the essential link between probabilities and distributions. The distribution function gives the probability that a member of the population chosen at random will bear a specified value of the variate or less. We must, however, consider whether this statement can still be regarded as true for populations which are infinite or continuous.

Suppose in the first instance that the population is infinite and discontinuous. In such a case we cannot select a member at random, but we may imagine a selection from a finite population which tends to the infinite form under consideration. In this finite population the proportion of members with values less than or equal to some x_0 will be $F(x_0)$ and thus, with due regard to the nature of the limiting process, we may still say that in the infinite population the probability of a value less than or equal to x_0 is $F(x_0)$.

Similarly for a continuous distribution. In Chapter 1 we considered the continuous form as a limiting expression of

$$\Delta F = f(x)\,\Delta x.$$

If a member is chosen at random from this population, the probability that it falls in the range Δx is $f(x)\Delta x$, for the total frequency is unity and $f(x)\Delta x$ is that proportion of it falling in the range Δx. In the limit we may say that the probability of obtaining a value less than or equal to x_0 in taking a member at random from a continuous population is $\int_{-\infty}^{x_0} dF = F(x_0)$. It must, however, be remembered that the nature of the process to the limit should be specified.

Hereafter, in speaking of selecting a member at random from a population $dF = f(x)\,dx$ we shall assume that what is meant is a selection random in the limit for intervals dx.

7.13 We have spoken above of the selection of objects "at random." In the mathematical theory of probability it is customary to define randomness in terms of probability itself. A member of a population is said to be chosen at random if it is chosen by a random method; and a random method is one which makes it equally probable that each member of the population will be chosen. Randomness is extremely important in the theory of sampling and we shall consider it at some length in Chapter 9. At this point it is sufficient to note that when we speak of random choice we really mean a method of selection which gives to certain propositions an equal probability and hence allows us to apply the calculus of probability *a priori*. The justification for this is, in the ultimate analysis, empirical. It is found in practice that there exist selective processes which educe members of a population in such a way that the constituent events may be regarded as equiprobable; and the theory of sampling is largely concerned with samples generated by such processes.

7.14 The idea of a variable which can appear with varying degrees of probability has been elevated into a distinct concept, that of a *random variable*. In ordinary analysis no such idea appears. We write "a variable" meaning that we are considering propositions about numbers which may be any of a certain range; there is no thought that one of these values is to be considered more frequently than others or that it will occur more frequently in practice. The random variable, on the other hand, is to be regarded as defined by a distribution function. It may take any values in a given range, but the values are distinguished by an associated function.

7.15 Let us consider what is meant by the addition of random variables. In ordinary analysis, given two variables ξ and η, we may define a third variable

$$\zeta = \xi + \eta,$$

which merely means that when $\xi = x$ and $\eta = y$, ζ will be $x+y$. If ξ and η are random variables, can we attach any useful meaning to ζ?

If the joint distribution function of ξ and η is F_{12}, the frequency of $\xi \leqslant x$ and $\eta \leqslant y$ is $F_{12}(x,y)$. Consider some value z. We may then determine from F_{12} the frequency such that $x+y \leqslant z$ which will, in fact, be the integral

$$\iint dF_{12}(x,y)$$

taken over the region for which $x+y \leqslant z$.

This integral defines a function of z which is in fact a distribution function, for it is zero at $-\infty$, non-decreasing, and unity at $+\infty$. We may then define this as the distribution function of the random variable and say that ζ is the sum of the random variables ξ and η.

7.16 More generally, suppose that we have random variables $\xi_1, \xi_2, \ldots, \xi_n$ with corresponding variates $x_1, x_2, \ldots, x_n$. We may define a random variable ζ, with variate z, by the equation

$$z = z(x_1, x_2, \ldots, x_n). \tag{7.15}$$

The random variable ζ is such that the distribution function of z is the integral of $dF(x_1, x_2, \ldots, x_n)$ over all values of $x_1, x_2, \ldots, x_n$ such that $z(x_1, x_2, \ldots, x_n) \leqslant z$.

In particular we may define the sum of the random variables by

$$z = x_1 + x_2 + \ldots + x_n. \tag{7.16}$$

We may even write

$$\zeta = \xi_1 + \xi_2 + \ldots + \xi_n. \tag{7.17}$$

This notation is justified by the fact that the commutative and distributive laws of algebra are valid for such expressions; e.g. it will be evident on examination that $\xi_1 + \xi_2 = \xi_2 + \xi_1$, and $\xi_1 + (\xi_2 + \xi_3) = (\xi_1 + \xi_2) + \xi_3 = \xi_1 + \xi_2 + \xi_3$.

7.17 The notation embodied in (7.17) is not, however, without its dangers. Some writers prefer to avoid the use of the word "sum" and to speak of the *convolution* of random variables. This might be preferable but has not come into general use among English writers.

We have not yet given any meaning to an expression such as $-\eta$. The most obvious interpretation, which we shall adopt, is that $-\eta$ is a random variable whose variate-values are the negatives of those of a variable η. Our notation has the disadvantage of seeming to imply that if a random variable ζ is the sum of two independent random variables ξ, η, and we subtract a random variable ξ from ζ, we shall necessarily be left with a random variable η. This is not true. As Example 7.5 makes clear, $\xi + \eta_1$ may have the same distribution as $\xi + \eta_2$ (where $\eta_2 \neq \eta_1$) even if the variables are independent.

7.18 We may notice one very simple but powerful result concerning the characteristic function of independent random variables. If we have n such variables distributed as $dF_1, \ldots, dF_n$ the element of frequency of their sum $\zeta = \xi_1 + \xi_2 + \ldots + \xi_n$ is the integral of $dF_1 \ldots dF_n$ through the element of volume between z and $z + dz$. Thus the characteristic function of their sum, being the integral of e^{itz} through the range of z, is equal to

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{itz}\, dF_1 \ldots dF_n$$

$$= \int_{-\infty}^{\infty} e^{itx_1}\,dF_1 \int_{-\infty}^{\infty} e^{itx_2}\,dF_2 \ldots \int_{-\infty}^{\infty} e^{itx_n}\,dF_n$$
$$= \phi_1\phi_2\ldots\phi_n. \tag{7.18}$$

Thus for independent random variables the c.f. of a sum is the product of the c.f.'s; and the cumulant-generating function of a sum is the sum of the individual cumulant-generating functions.

The essential point in Example 7.5, which now follows, is that two independent random variables, ξ, η, may have characteristic functions which coincide in some interval of t containing the origin. If the c.f. of ξ is zero outside this interval, (7.18) ensures that the sum of two independent ξ's has the same distribution as $\xi+\eta$.

Example 7.5

Starting from the integral

$$\int_0^{\infty} \frac{\sin\alpha x}{x}\,dx = \tfrac{1}{2}\pi, \qquad \alpha>0,$$

we have, on integration with respect to α,

$$\int_0^{\infty} \frac{\cos\beta x-\cos\alpha x}{x^2}\,dx = \tfrac{1}{2}\pi(\alpha-\beta), \qquad \alpha>\beta.$$

It is then easy to show, by the double use of identities such as

$$\sin\theta\cos m\theta = \tfrac{1}{2}\{\sin(m+1)\theta-\sin(m-1)\theta\},$$

that

$$\frac{2}{\pi}\int_0^{\infty} \frac{1-\cos x}{x^2}\cos tx\,dx = 1-|t| \text{ for } |t|\leqslant 1$$
$$= 0 \qquad \text{for } |t|>1.$$

Thus the distribution

$$dF = \frac{1}{\pi}\,\frac{1-\cos x}{x^2}\,dx, \qquad -\infty\leqslant x\leqslant\infty,$$

has the c.f.

$$\left.\begin{aligned}\phi_1(t) &= 1-|t|, \quad |t|\leqslant 1\\ &= 0, \quad |t|>1\end{aligned}\right\} \tag{7.19}$$

for the integral

$$\frac{1}{\pi}\int_{-\infty}^{\infty} \frac{1-\cos x}{x^2}\sin tx\,dx = 0$$

in virtue of the oddness of the function $\sin tx$. Moreover, from the Fourier series, which is easily obtained,

$$\cos\theta+\frac{1}{3^2}\cos 3\theta+\frac{1}{5^2}\cos 5\theta+\ldots = \tfrac{1}{8}(\pi^2-2\pi\theta), \qquad 0\leqslant\theta\leqslant\pi,$$
$$= \tfrac{1}{8}(\pi^2+2\pi\theta), \qquad -\pi\leqslant\theta\leqslant 0,$$

we may show that the discontinuous distribution

$$\left.\begin{aligned}f(0) &= 1-\tfrac{1}{2}\pi,\\ f(x) &= \frac{2}{\pi x^2}, \quad x = \text{a positive or negative odd integer},\end{aligned}\right\} \tag{7.20}$$

has the c.f.

$$\phi_2(t) = 1-|t|, \qquad |t| \leqslant 1. \tag{7.21}$$

For $|t|>1$ the c.f. consists of periodic repetitions of the values inside the range $|t| \leqslant 1$.

It follows from (7.19) and (7.21) that *for all* t, $\phi_1(t)\phi_2(t) = \phi_1(t)\phi_1(t)$. Thus the sum of the first and second variables is distributed in the same form as the sum of two random variables distributed like the first. Thus it is not true, in general, that if $\xi+\eta_1 = \xi+\eta_2$ then $\eta_1 = \eta_2$.

Example 7.6

The converse of the proposition of **7.18** is untrue. Although it is true that, if ξ and η are independent, their joint frequency-distribution is the product of their individual frequency-distributions *and conversely*, it is not true that if the c.f. of their sum is the product of their individual c.f.'s they are independent.

Consider the distribution

$$dF = \left\{\frac{1}{2\pi}x^{-\frac{1}{2}}y^{-\frac{1}{2}}e^{-\frac{1}{2}(x+y)} + \varepsilon(x-y)(xy-x-y+2)e^{-(x+y)}\right\}dx\,dy, \qquad 0 \leqslant x, y \leqslant \infty. \tag{7.22}$$

Since $x^{\frac{1}{2}}y^{\frac{1}{2}}(x-y)(xy-x-y+2)e^{-\frac{1}{2}(x+y)}$ has a lower bound there will be some value of ε, perhaps small, for which the term in braces on the right in (7.22) is non-negative throughout the range of x and y. That is to say, (7.22) can genuinely represent a frequency function. Moreover, if we interchange x and y in the term involving ε we change its sign and hence the integral of this term over the range $0 \leqslant x, y \leqslant \infty$ is zero. The total frequency is then the integral of the first term, which is clearly $\{2^{\frac{1}{2}}\Gamma(\frac{1}{2})\}^2/(2\pi) = 1$.

The c.f. of x and y is obtained by integrating $\exp(it_1x+it_2y)$ over (7.22) and is found to reduce to

$$\phi(t_1, t_2) = \frac{1}{(1-2it_1)^{\frac{1}{2}}(1-2it_2)^{\frac{1}{2}}} + \frac{2\varepsilon t_1 t_2(t_1-t_2)}{(1-it_1)^3(1-it_2)^3}. \tag{7.23}$$

To obtain the c.f. of x we put $t_2 = 0$, obtaining

$$\phi_1(t_1) = (1-2it_1)^{-\frac{1}{2}}.$$

Similarly for the c.f. of y we put $t_1 = 0$, obtaining

$$\phi_2(t_2) = (1-2it_2)^{-\frac{1}{2}}.$$

For the c.f. of $x+y$ we put $t_1 = t_2$, obtaining $(1-2it)^{-1}$ which is the product of $\phi_1(t)$ and $\phi_2(t)$. The relationship between the c.f.'s is, in fact, the same as if ε were zero, which is due to the fact that in (7.23) the term containing ε has factors t_1, t_2 and (t_1-t_2).

7.19 Notwithstanding the nomenclatural dangers, we shall often in the sequel pass over the distinction between " random variable " and " variate." When we use the expression " variable " alone, an ordinary mathematical variable is to be understood unless the context makes it clear that a random variable is meant. Where there is possible confusion we shall speak of random variable. To avoid drawing distinctions between the random variable and its variate-values, typified respectively by ξ and x, we shall often use roman letters to specify random variables and shall, where the context permits, refer to them as variates.

Sampling distributions

7.20 We have noted that if a member of a population is chosen at random the probability that it will bear a variate-value not greater than x is the distribution function $F(x)$. Similarly, if we choose a member from a multivariate population, the probability that it will bear a value of the first variate not greater than x_1, of the second not greater than $x_2, \ldots$, of the nth not greater than x_n, is the multivariate distribution function $G(x_1, x_2, \ldots, x_n)$. Further, if the variates are independent, the rth variate having the distribution function $F_r(x_r)$, this probability is equal to

$$F_1(x_1)\,F_2(x_2)\ldots F_n(x_n).$$

Now suppose that we have a selective process, which we will call sampling, applied to a univariate population in such a way that it abstracts a group of n members. If this process is repeated it will generate a multivariate distribution, each sample exhibiting n values $x_1, x_2, \ldots, x_n$. The nature of this multivariate distribution depends on the sampling process as well as the population. If the distribution is $G(x_1, x_2, \ldots, x_n)$, then this function represents the probability that a random sample will result in n values, the first not greater than x_1, the second not greater than x_2, and so on. We may regard the x's as corresponding to n random variables $\xi_1, \xi_2, \ldots, \xi_n$.

There is one type of sampling process of outstanding importance in statistical theory, namely that in which the distribution $G(x_1, x_2, \ldots, x_n)$ is the product of factors $F_1(x_1)\,F_2(x_2)\ldots F_n(x_n)$. In such a case the sampling is said to be simple. The distributions of the values $x_1, x_2, \ldots, x_n$ are independent one of another, and we may thus say that the selection of any member is independent of that of any other. Moreover, if the sampling is random, every $F_r(x)$ will be equal to $F(x)$, the distribution function of the population from which the sample has been drawn. Thus in this case we have, for the distribution of the variate-values in samples of size n obtained by a simple random method,

$$\begin{aligned} dG(x_1, x_2, \ldots, x_n) &= dF(x_1)\,dF(x_2)\ldots dF(x_n) \\ &= f(x_1)\,f(x_2)\ldots f(x_n)\,dx_1\,dx_2\ldots dx_n, \end{aligned} \qquad (7.24)$$

and $F(x_1)F(x_2)\ldots F(x_n)$ is the probability that in such a sample the first value will not exceed x_1, and so on. Moreover, since the x's appear symmetrically in (7.24) their order is not material.

7.21 If we have a sample of size n, characterized by the variates $x_1, x_2, \ldots, x_n$, we may, as in (7.15), consider the distribution of some function $z = z(x_1, x_2, \ldots, x_n)$ which might, for instance, be their mean or variance. The corresponding random variable ζ will have a distribution which is known as the sampling distribution of ζ (or of z). The probability that a value of ζ less than or equal to z be obtained on random sampling is the distribution function of z. In the case of simple random sampling it is given by integrating the element of (7.24) over the range of x's for which $z(x_1, x_2, \ldots, x_n) \leqslant z$.

Example 7.7

Suppose we draw a simple random sample of size two from the normal population

$$dF = \frac{1}{\sigma\sqrt{(2\pi)}} e^{-x^2/2\sigma^2}, \qquad -\infty \leqslant x \leqslant \infty.$$

In virtue of (7.24) the distribution of the two values, say x_1 and x_2, is given by

$$dF = \frac{1}{2\pi\sigma^2} \exp\{-(x_1^2+x_2^2)/2\sigma^2\}\, dx_1\, dx_2. \tag{7.25}$$

We will integrate this over values $z \leqslant \frac{1}{2}(x_1+x_2)$. This will give us the distribution of the mean of the two values.

Put $$u = \tfrac{1}{2}(x_1+x_2), \qquad v = \tfrac{1}{2}(x_1-x_2).$$

The distribution becomes

$$dF = \frac{1}{\pi\sigma^2} \exp\{-(u^2+v^2)/\sigma^2\}\, du\, dv. \tag{7.26}$$

It so happens that u and v are, in this case, independent. Integrating out for v we find for u the distribution

$$dF = \frac{1}{\sigma\sqrt{\pi}} \exp(-u^2/\sigma^2)\, du. \tag{7.27}$$

Thus the probability that $\frac{1}{2}(x_1+x_2)$ is not greater than some z is simply

$$\frac{1}{\sigma\sqrt{\pi}} \int_{-\infty}^{z} e^{-u^2/\sigma^2}\, du, \tag{7.28}$$

a result which we may express by saying that u is distributed normally with variance $\frac{1}{2}\sigma^2$.

7.22 In the remainder of this chapter we shall be concerned with some statistically important theorems concerning the limit of the sum of n random variables as n tends to infinity. We first of all prove a simple but effective result due to Tchebycheff, previously stated at **3.97**.

Let x be a random variable and $g(x)$ a non-negative function of it. Then for every $k > 0$,

$$\mathrm{P}\{g(x) \geqslant k\} \leqslant \mathrm{E}\{g(x)\}/k. \tag{7.29}$$

In fact, if R is the region for which $g(x) \geqslant k$ we have

$$\mathrm{E}\{g(x)\} = \int_{-\infty}^{\infty} g(x)\, dF \geqslant \int_R g(x)\, dF \geqslant k \int_R dF \geqslant k\mathrm{P}\{g(x) \geqslant k\}.$$

In the particular case when $g(x) = \{x - \mathrm{E}(x)\}^2$ this reduces to the Bienaymé-Tchebycheff inequality (3.95).

The Weak Law of Large Numbers

7.23 Let $\xi_1, \xi_2, \ldots$ be a set of independent random variables distributed in the same form with mean μ. Let $\bar{\xi}_n$ be the mean of the first n,

$$\bar{\xi}_n = \frac{1}{n} \sum_{i=1}^{n} \xi_i.$$

Then $\bar{\xi}_n$ also has mean μ, whatever the value of n. The (Weak) Law of Large Numbers states that, in effect, $\bar{\xi}_n$ becomes more and more narrowly dispersed about μ as n increases. More precisely, given any positive ε,

$$\lim_{n\to\infty} \mathrm{P}\{|\bar{\xi}_n-\mu|>\varepsilon\} = 0. \tag{7.30}$$

If we may assume that the variance of any ξ exists, and is equal to σ^2, say, the proof is very easy. For $\bar{\xi}_n$ is a random variable with variance σ^2/n and, in virtue of the Bienaymé-Tchebycheff inequality (3.95),

$$\mathrm{P}\{|\bar{\xi}_n-\mu|>\varepsilon\} \leqslant \sigma^2/n\varepsilon^2.$$

For given ε the probability can therefore be made as small as we please by increasing n.

7.24 The theorem remains true even if we discard the requirement that σ^2 exists. The foregoing proof is evidently easily adaptable to the case when $E(|\xi|^\alpha)$ exists for $\alpha>1$ but for the case $\alpha = 1$ a different type of proof is required. (Cf. Exercise 7.20.) It is also easily adapted to the case where the ξ's have different means.

The Strong Law of Large Numbers

7.25 The Weak Law of Large Numbers states a limiting property of sums of random variables. The so-called "strong" law states something about the behaviour of the sequence $\Sigma_n \equiv \sum_{i=1}^{n} \xi_i$ for all values of n; something, as it were, about its properties on the way to the limit.

In fact, given any positive numbers ε and δ there is an N such that for every $M>0$

$$\mathrm{P}\{|\bar{\xi}_n-\mu|\geqslant\varepsilon\}\leqslant\delta, \quad n = N, N+1,\ldots,N+M. \tag{7.31}$$

The weak law states that $|\bar{\xi}_n-\mu|$ is ultimately small but not that every value is small; it might be that for some n it was large, although such cases could only occur infrequently. The strong law says that the probability of such an event is extremely small. The law is true for variables which are identically distributed under the sole weak condition that μ exists; in other cases further conditions must be added.

The Central Limit theorem

7.26 A much more precise theorem is available in cases where the variances exist. In fact, the mean $\bar{\xi}_n$ tends to be distributed normally about μ with variance σ^2/n. More precisely, if $\alpha(x)$ represents the normal frequency function then, for every t_1, t_2,

$$\lim \mathrm{P}\left\{t_1 \leqslant \frac{\bar{\xi}_n-\mu}{\sigma/\sqrt{n}} \leqslant t_2\right\} = \int_{t_1}^{t_2} \alpha(t)\,dt. \tag{7.32}$$

This theorem is also true if the ξ's have different distributions, μ then being replaced by the mean of individual μ's and $\sigma/\sqrt{n}$ being replaced by the variance of $\bar{\xi}_n$.

Let us suppose that each ξ has a finite absolute third moment. Then as in **3.15** we have, for the c.f. of the rth variate (with first and second moments, say, of μ'_{1r} and μ'_{2r}),

$$\phi_r(t) = 1+\mu'_{1r}(it)+\tfrac{1}{2}\mu'_{2r}(it)^2+R \tag{7.33}$$

where R is not greater than $\frac{1}{6}\nu'_{3r}|t|^3$. Likewise, for the cumulative function we have

$$\psi_r(t) = \mu'_{1r}(it)+\tfrac{1}{2}\mu_{2r}(it)^2+S$$

where S is some constant times $|t|^3$, say $k|t|^3$. Thus, for the c.f. of all the n variables

o

when summed we have

$$\Psi(t) = \sum_{r=1}^{n} \mu'_{1r}(it) + \tfrac{1}{2} \sum_{r=1}^{n} \mu_{2r}(it)^2 + \Sigma S.$$

Let us now take the mean of the sum as origin and transfer to standard measure by putting $u = \Sigma \xi / (\Sigma \mu_{2r})^{\frac{1}{2}}$. This gives us for u

$$\psi(t) = -\tfrac{1}{2}t^2 + \Sigma S / (\Sigma \mu_{2r})^{3/2}.$$

If each ν_{3r} is finite the numerator of the remainder term is bounded by $nK|t|^3$, where K is the largest value of k. The denominator is not less than $n^{3/2}M_{2r}$, where M_2 is the smallest variance. Thus the remainder term is of order $n^{-\frac{1}{2}}$ and tends to zero. Thus

$$\psi(t) \longrightarrow -\tfrac{1}{2}t^2$$

and consequently the distribution to which it relates tends to normality.

7.27 The theorem may be proved under conditions which do not require the existence of third moments. In fact, it is a necessary and sufficient condition that

$$\lim_{n\to\infty} \frac{1}{M_n} \sum_{r=1}^{n} \int_{|x| > \varepsilon\sqrt{M_n}} x^2 \, dF_r = 0, \tag{7.34}$$

where $M_n = \Sigma \mu_{2r}$.

This condition, due to Lindeberg and Cramér, implies that the total variance M_n tends to infinity and that every μ_{2r}/M_n tends to zero; in fact that no random variable dominates the others. The theorem may fail to hold for variables which do not possess a second moment; for example, we shall see in Chapter 11 that the mean of n variates, each distributed in the Cauchy form

$$dF = \frac{1}{\pi} \frac{dx}{1+x^2}, \qquad -\infty \leqslant x \leqslant \infty, \tag{7.35}$$

is distributed in precisely the same form. This is easily deduced from the c.f., given in Example 4.2.

7.28 The Central Limit theorem occupies an important place in statistical theory. We have already met instances of its operation in the tendency of the binomial and the Poisson distributions to normality. We shall meet many more in the sequel. Central Limit theorems are also obtainable for sequences of random variables which are not independent. A general account of these limit theorems will be found in the book by Feller (1950).

7.29 The foregoing methods can also be extended to sets of multivariate random variables. For example, if ξ, η have a joint distribution, a set of n individuals drawn at random from that distribution under simple conditions yields a set of n pairs of variate-values $(x_1, y_1), (x_2, y_2) \ldots, (x_n, y_n)$. We might then enquire after the sampling distribution of some quantity such as their joint product-moment $z = \sum_{i=1}^{n} x_i y_i / n$. This would be obtained by integrating

$$dF(x_1, y_1)\, dF(x_2, y_2) \ldots dF(x_n, y_n)$$

over a region such that $\Sigma x_i y_i / n \leqslant z$. Generalizations to more than two random variables are immediate.

7.30 It is also worth recording that Central Limit theorems can be proved for multivariate random variables. In particular, corresponding to the theorem of **7.26**, the means of n x's and y's tend to joint bivariate normality as n increases under similar very general conditions.

EXERCISES

7.1 Let $E_1, E_2, \ldots E_n$ be n compatible events, i.e. such that any number of them from 0 to n may be simultaneously realized. Let p_i be the probability that E_i is realized (whether the others are realized or not), p_{ij} the probability that E_i and E_j are realized (whether the others are realized or not); and so on. Show that the probability that no event is realized is

$$1-\Sigma p_i+\Sigma p_{ij}-\Sigma p_{ijk}+\ldots+(-1)^n p_{12\ldots n},$$

where summations take place over all possible different values of the suffixes.

7.2 Show that the probability that, of n compatible events, exactly r unspecified events are realized is

$$\Sigma p_{(r)}-\binom{r+1}{1}\Sigma p_{(r+1)}+\binom{r+2}{2}\Sigma p_{(r+2)}+\ldots+(-1)^{n-r}\binom{n}{n-r}p_{(n)},$$

where $p_{(j)}$ denotes a p with j suffixes.

7.3 In the previous exercise show that the probability that at least r unspecified events are realized is

$$\Sigma p_{(r)}-\binom{r}{1}\Sigma p_{(r+1)}+\binom{r+1}{2}\Sigma p_{(r+2)}-\ldots+(-1)^{n-r+1}\binom{n-1}{n-r}p_{(n)}.$$

7.4 Use Exercise 7.2 to solve the envelope problem of Example 7.3. Show also that the probability that exactly r unspecified letters get into the right envelopes is

$$\frac{1}{n!}\left\{\binom{n}{r}(n-r)!-\binom{r+1}{1}\binom{n}{r+1}(n-r-1)!+\text{etc.}\right\}=\frac{1}{r!}\left\{1-1+\frac{1}{2!}-\ldots+\frac{(-1)^{n-r}}{(n-r)!}\right\}.$$

7.5 Given n independent but compatible events $E_1 \ldots E_n$, and that the probability that E_i alone happens is α_i, show that the probability that E_i happens, whether the others happen or not, p_i, is given by

$$p_i = \alpha_i/(\alpha_i+t)$$

where t is a root of

$$\prod_{i=1}^{n}(\alpha_i + t) = t^{n-1}.$$

Show that in general there are two values of p_i satisfying the conditions.

7.6 From a heap of counters of unknown number N a player takes a handful of n at random. Examine this argument: it is equally probable that N is odd or even. If it is odd, the probability that n is odd is greater than $\frac{1}{2}$, whereas if it is even the probability that n is odd is $\frac{1}{2}$. Thus the probability that n is odd is greater than $\frac{1}{2}$, and the player should bet on getting an odd number.

7.7 An event happens at random on an average once in time t. Regarding occurrences in equal small intervals as equiprobable, show that the probability that it does not happen in a specified interval T is $\exp(-T/t)$.

7.8 If a is equiprobably distributed in the interval $0 \leqslant a \leqslant 6$ and b equiprobably in the interval $0 \leqslant b \leqslant 9$, show that the probability that $x^2-ax+b = 0$ has two real roots is $\frac{1}{3}$.

7.9 A straight line in a plane is represented by the pedal equation

$$x\cos\theta + y\sin\theta = p.$$

A random line is drawn in such a way that elements $d\theta\, dp$ are equally probable. If it intersects a closed convex curve of length l_1, show that the probability that it will also intersect a convex curve of length l_2, lying inside the first, is l_2/l_1.

Deduce that if a random line intersects a circle, the probability that it also intersects a fixed diameter of the circle is $2/\pi$.

7.10 If, in the previous exercise, the co-ordinates are changed by a translation of the origin and a rotation about it, show that $dp'\, d\theta' = dp\, d\theta$ where the primes relate to the new co-ordinates. Hence show that the " randomness " of lines drawn in the foregoing manner is independent of the co-ordinate system.

7.11 Three points are taken at random on a circle. Show that the probability that they lie on the same semicircle is $\frac{3}{4}$. (Assume that in the limit elementary intervals of arc are equally probable.)

Explain the fallacy in the following argument: one pair of points *must* lie on a semicircle terminating at one of them. The probability that the third point lies on this semicircle is $\frac{1}{2}$, which is therefore the required answer.

Consider also this argument: it does not matter where the first point is chosen. Imagine the circle cut at that point and unrolled into a horizontal straight line. The probability that the second and third points fall into the left half of the line is $\frac{1}{2}\times\frac{1}{2} = \frac{1}{4}$; similarly for the right half; and hence the probability required is $\frac{1}{2}$.

7.12 If x and y are independent random variables with characteristic functions ϕ_1 and ϕ_2 respectively, show that the c.f. of $x-y$ is $\phi_1(t)\,\phi_2(-t)$. Hence show that if $x+y$ is distributed in the same form as $x-y$, then y is symmetrically distributed about zero.

7.13 x and y are distributed in the bivariate normal form of (5.90). Find the c.f. of $x+y$ and hence show that $x+y$ is normally distributed with variance $2(1+\rho)$. Deduce that $|\rho| \leqslant 1$.

7.14 Show that the sum of a number of independent random variables, each distributed in the Poisson form with possibly different parameters, is also distributed in the Poisson form.

7.15 x is a standardized normal variate and $y = x^2$. Show that the joint c.f. of x and y is

$$\phi_{x,y}(t_1, t_2) = (1-2it_2)^{-\frac{1}{2}}.\exp\left\{-\frac{1}{2}\frac{t_1^2}{1-2it_2}\right\}$$

and deduce the c.f.'s of x and of y alone. Show that the covariance of x and y is zero.

7.16 A random sample of n values $x_1, x_2, \ldots, x_n$ is chosen from a population which has mean μ and takes only positive values. Show that the probability that $\sum_{i=1}^{n} x_i$ exceeds some constant λ is not greater than $n\mu/\lambda$.

7.17 Show that the mean (say u) of two variates, distributed independently in the form

$$dF = \tfrac{3}{2}x^2\,dx, \qquad -1 \leqslant x \leqslant 1,$$

is distributed in the form

$$dF = \tfrac{3}{5}(3-15u+20u^2-8u^5).$$

7.18 Find the distribution of the sum of k independent random variables, each distributed in the form

$$dF = e^{-x}\,dx, \qquad 0 \leqslant x \leqslant \infty.$$

Graph the distribution for some values of k and observe its tendency to normality.

7.19 The frequency diagram of a distribution is defined by a series of isosceles triangles with bases two units wide and height $A/|x|$ for $x = \pm (p+1)^2, p = 1, 2, \ldots$, together with the remaining part of the x-axis. Show that if A is suitably chosen this gives a properly defined d.f.

Show that the expectation of $|x|$ does not exist. Show also that if an expectation be defined as the principal value $\lim_{n\to\infty} \sum_{-n}^{n} xf(x)$, it does exist. By taking a different origin show that the principal value about some other point does not exist. Hence, by taking one random variable to be a constant, show that the expectation of the sum of two independent random variables is not necessarily equal to the sum of their expectations if the principal value be admitted as an expectation.

(Fréchet, 1937)

7.20 By expanding their c.f.'s as

$$\phi_j(t) = 1 + \mu it + o(t),$$

show that the mean $\bar{\xi}_n$ of n identical independent variates ξ_j has a c.g.f. for which

$$\lim_{n\to\infty} \psi(t) = \mu it,$$

and hence that $\bar{\xi}_n$ has the unit distribution of Example 4.3 in the limit. This establishes the Weak Law of Large Numbers of **7.30.**

CHAPTER 8

PROBABILITY AND STATISTICAL INFERENCE

8.1 The calculus of probabilities, as outlined in the previous chapter, leads us from the given probabilities of primary events to the probabilities of more complex events based upon them. In practice we usually require to make inferences in the reverse direction ; that is to say, given the observations, we require to know something about the parent population from which they emanated or the generating mechanism by which they were produced. In the second volume we shall begin a systematic study of the various methods and inferential processes which are employed in statistics for this purpose. At this stage we shall merely attempt an introductory account, in very broad terms, with the object of giving some point to the topics considered and discussed later in this volume.

Bayes' Theorem

8.2 Let $q_1, q_2, \ldots, q_n$ be a set of exclusive propositions and let H be the information available. Let p represent some further proposition. Then by the multiplicative rule of probabilities (our Rule 3) we have

$$\begin{aligned} \mathrm{P}(q_r, p \mid H) &= \mathrm{P}(p \mid H)\,\mathrm{P}(q_r \mid p, H) \\ &= \mathrm{P}(q_r \mid H)\,\mathrm{P}(p \mid q_r, H). \end{aligned}$$

Thus we have

$$\mathrm{P}(q_r \mid p, H) = \frac{\mathrm{P}(q_r \mid H)\,\mathrm{P}(p \mid q_r, H)}{\mathrm{P}(p \mid H)}. \tag{8.1}$$

If, as we now suppose, the q's are exhaustive so that one of them must hold, we have, on summing (8.1) over the q's,

$$1 = \sum_r \frac{\mathrm{P}(q_r \mid H)\,\mathrm{P}(p \mid q_r, H)}{\mathrm{P}(p \mid H)}. \tag{8.2}$$

On substituting for $\mathrm{P}(p \mid H)$ from (8.2) into (8.1), we then find

$$\mathrm{P}(q_r \mid p, H) = \frac{\mathrm{P}(q_r \mid H)\,\mathrm{P}(p \mid q_r, H)}{\sum_r \{\mathrm{P}(q_r \mid H)\,\mathrm{P}(p \mid q_r, H)\}} = \frac{\mathrm{P}(q_r, p \mid H)}{\sum_r \mathrm{P}(q_r, p \mid H)}. \tag{8.3}$$

This is known as Bayes' Theorem. It states that the probability of q_r on data p and H is proportional to the probability of q_r on H multiplied by the probability of p on q_r and H.

8.3 Now suppose that an event can be explained on the mutually exclusive hypotheses represented by $q_1, q_2, \ldots, q_n$. These possibilities have certain probabilities $\mathrm{P}(q_r \mid H)$ of being true. Each of them can give rise to p, but with differing probabilities $\mathrm{P}(p \mid q_r, H)$. The theorem gives us the probabilities of the various q's *when p is known to have happened.* The quantities $\mathrm{P}(q_r \mid H)$ are called *prior* probabilities, those of

type $P(q_r \mid pH)$ are called *posterior* probabilities and $P(p \mid q_r, H)$ may be called the *likelihood.* Bayes' theorem may then be re-stated in the form : the posterior probability varies as the prior probability multiplied by the likelihood.

Bayes' Postulate

8.4 In this form the theorem is seen to be a simple logical consequence of the multiplication rule of probabilities and is not subject to criticism. What has given rise to criticism in the past has been the use to which the theorem has been put. There is an implied principle that, if we have to choose one of the q's, we take the one with the greatest posterior probability. This is equivalent to choosing the hypothesis which maximizes the *joint probability* of q and p as is seen at once from the extreme right of equation (8.3). The difficulty arises from the fact that to calculate the posterior probabilities we require to know the prior probabilities. These are, in general, unknown, and Bayes suggested that where this is so they should be assumed to be equal ; or rather, that they should be assumed equal where nothing was known to the contrary. This assumption, known variously as Bayes' Postulate, the Principle of the Equidistribution of Ignorance and one or two other names, furnishes one of the most contentious points in the theory of statistical inference.(*) Before we discuss the point it may be useful to give one or two examples.

Example 8.1

An urn contains four balls, which are known to be either (a) all white, or (b) two white and two black. A ball is drawn at random and found to be white. What is the probability that all the balls are white ?

We have here two hypotheses, q_1 and q_2. On q_1 the probability of getting a white ball is 1, on q_2 it is $\frac{1}{2}$. From (8.1) we have

$$P(q_1 \mid p, H) = \frac{P(q_1 \mid H)}{P(q_1 \mid H) + \frac{1}{2}P(q_2 \mid H)}$$

$$P(q_2 \mid p, H) = \frac{\frac{1}{2}P(q_2 \mid H)}{P(q_1 \mid H) + \frac{1}{2}P(q_2 \mid H)}.$$

Now, in accordance with Bayes' postulate we assume

$$P(q_1 \mid H) = P(q_2 \mid H) = \tfrac{1}{2}$$

and find

$$P(q_1 \mid p, H) = \tfrac{2}{3}$$

$$P(q_2 \mid p, H) = \tfrac{1}{3}.$$

If we had to choose between the two possibilities (a) and (b) we should select the one with the greater posterior probability, i.e. we proceed on the assumption that the balls are all white.

Now suppose that we replace the ball and draw another at random. If it is found to be black the hypothesis (a) is decisively rejected. But if it turns out to be white we can calculate new posterior probabilities in which our former posterior probabilities

(*) The modern versions of the postulate, as, for example, that given by Jeffreys, sometimes assume prior distributions which are not equiprobable. This requires some special pleading, and we defer a discussion until Volume 2.

become prior. We now have $P(q_1|H) = \frac{2}{3}$, $P(q_2|H) = \frac{1}{3}$, where H includes p, and a renewed application of (8.1) gives us for the posterior probabilities of the new event, say p',

$$P(q_1|p', H) = \tfrac{2}{3}/(\tfrac{2}{3}+\tfrac{1}{2}.\tfrac{1}{3}) = \tfrac{4}{5}$$
$$P(q_2|p', H) = \tfrac{1}{2}.\tfrac{1}{3}/(\tfrac{2}{3}+\tfrac{1}{2}.\tfrac{1}{3}) = \tfrac{1}{5}.$$

It will be clear that if we repeat the process and again get a white ball the new posterior probability of (a) will be still higher. This is in agreement with the requirements of common sense ; the longer we go on sampling (with replacement) without producing a black ball, the more probable it is that there are no black balls present.

Example 8.2

We may look at the same problem from a slightly different viewpoint. Suppose we draw balls one at a time, replacing them after each drawing, and obtain n white balls in succession. The probability of this event on hypothesis (a) is unity ; that on hypothesis (b) is $1/2^n$. From (8.1) we then have, (p referring to the observation of all n balls as white),

$$P(q_1|p, H) = \frac{\frac{1}{2}}{\frac{1}{2}+\frac{1}{2}.1/2^n}$$
$$= 2^n/(2^n+1).$$
$$P(q_2|p, H) = 1/(2^n+1).$$

As n becomes larger $P(q_1|p, H)$ tends to unity and $P(q_2|p, H)$ to zero.

Moreover, this will be true whatever the original prior probabilities may have been. In fact, if that of hypothesis (a) is t and that of (b) is $1-t$, we find

$$P(q_1|p, H) = \frac{2^n t}{2^n t+(1-t)}$$

which tends to unity for any non-zero t. This also agrees with common sense. Whatever the original probabilities, the new evidence is so strong as to outweigh them.

Example 8.3

From an urn full of balls of unknown colour a ball is drawn at random and replaced. The process is continued m times and a black ball is drawn each time. What is the probability that if a further ball is drawn it will be black ?

The question as framed does not admit of a definite answer, for, there being an infinite number of possible colours and combinations of colours, we do not know which are the hypotheses to be compared. Let us suppose that the balls are either black or white, and thus consider the hypotheses (1) that all are black, (2) that all but one are black, (3) that all but two are black, and so on. The problem still lacks precision, for the number of balls is not specified. Suppose there are N balls. We shall later let N tend to infinity to get the limiting case.

Consider the hypothesis q_R that there are R black balls and $N-R$ white ones. The probability of choosing a black ball is R/N and that of doing so m times in succession

is $(R/N)^m$. If the q's have equal prior probabilities we have, from (8.1),

$$P(q_R|p,H) = \frac{(R/N)^m}{\sum_{R=0}^{N}(R/N)^m}.$$

Now the probability of getting a further black ball on hypothesis q_R is R/N. Since the hypotheses q_R are mutually exclusive, the probability of getting a further black ball is

$$\sum_{R=0}^{N}\frac{R}{N}P(q_R|p,H) = \frac{\Sigma(R/N)^{m+1}}{\Sigma(R/N)^m}. \tag{8.4}$$

This is the answer to the limited form of the question. As $N\to\infty$ this tends to the quotient of definite integrals

$$\int_0^1 x^{m+1}\,dx \Big/ \int_0^1 x^m\,dx = \frac{m+1}{m+2}. \tag{8.5}$$

This is a particular case of the so-called Succession Rule of Laplace. Enthusiasts have applied it indiscriminately in some such unconditioned form as the statement that if an event is observed to happen m times in succession the chances are $m+1$ to 1 that it will happen again. This is clearly unjustified.

8.5 The principal difficulties arising out of Bayes' postulate appear from the standpoint of the frequency theory of probability. If we adopt the approach in which probability is a measure of attitudes of mind, it is reasonable to take prior probabilities to be equal when nothing is known to the contrary, for the mind holds them in equal doubt. The frequency theory, however, would require the states of events corresponding to the various q's to be distributed with equal frequency in some population from which the actual q has emanated, if Bayes' postulate is to be applied. This has appeared to some statisticians, though not to all, to be asking too much of the universe. The postulate is one of the crucial points in the theory of probability. Many adherents of the subjective school accept it. Many of those of the frequency school explicitly reject it.

There is still so much disagreement on this subject that one cannot put forward any set of viewpoints as orthodox. One thing, however, is clear—anyone who rejects Bayes' postulate must put something in its place. The problem which Bayes attempted to solve is supremely important in scientific inference and it scarcely seems possible to have any scientific thought at all without some solution, however intuitive and however empirical, to the problem. We are constantly compelled to assess the degree of credence to be accorded to hypotheses on given data; the struggle for existence, in Thiele's phrase, compels us to consult the oracles.

Maximum likelihood

8.6 Various substitutes for Bayes' postulate have been proposed. Some of these are put forward as solutions to the problem in particular classes of case; such are the principles of least squares and minimum chi-squared which we shall meet later. There is one principle, however, of general application, that of maximum likelihood.

Reverting to (8.3) we may write Bayes' theorem in the form

$$\mathrm{P}(q_r \mid p, H) \propto \mathrm{P}(q_r \mid H)\mathrm{L}(p \mid q_r, H) \tag{8.6}$$

where we now write $\mathrm{L}(p \mid q_r, H)$ for the likelihood. The Principle of Maximum Likelihood states that, when confronted with a choice of hypotheses q_r we choose that one (if any) which maximizes L. In other words, we are to choose that hypothesis which gives the greatest probability to the observed event. Whereas Bayes' theorem enjoins the maximization of the joint probability of q_r and p, maximum likelihood requires the maximization of the conditional probability of p given q_r.

8.7 It is to be particularly noted that this is not the same thing as choosing the hypothesis with the greatest probability. Some advocates of the maximum-likelihood principle explicitly deny any meaning to such expressions as "the probability of a hypothesis." We shall see later that in practice the differences between results obtained from maximum likelihood and Bayes' postulate are not so large as might be expected. There is, however, an important conceptual difference involved.

There is, in fact, a shift of emphasis in the way we regard the likelihood function, reflected by our writing it with an L instead of a P. The ordinary probability function gives the probabilities of p on data q_r and H; p varies, q_r and H are given. From the point of view of likelihood we consider various values of q_r for the observed p and given H; q_r varies, p and H are given. It is this variation of the function for differing values of the q's that we have in mind in speaking of likelihood.

8.8 Suppose (as is nearly always the case in statistical work) that the hypotheses with which we are concerned assert something about the numerical value of a parameter θ. For instance, the hypotheses might be $q_1 \equiv \theta < 0$, $q_2 \equiv \theta \geqslant 0$, in which case there are two alternatives. Or we might have $q_1 \equiv \theta = 1$, $q_2 \equiv \theta = 2$, and so on, in which case there is a denumerable infinity of hypotheses.

If now θ can have only discontinuous values, we may, confronted with an observed event p, require to estimate θ, or to ask what is the "best" value of θ to take on the evidence p. The method of Bayes would state that the "best" value was the most probable value. In (8.3) we should seek for that q_r which made $\mathrm{P}(q_r \mid p, H)$ a maximum. If we know nothing of the prior probabilities $\mathrm{P}(q_r \mid H)$ we should, in accordance with Bayes' postulate, assume all such probabilities equal. We then merely have to find that q_r which maximizes $\mathrm{L}(p \mid q_r, H)$. In other words, the postulate of Bayes and the principle of maximum likelihood result in the same answer and are equivalent.

8.9 This position apparently does not hold if the permissible values of θ are continuous. We must now replace such expressions as

$$\mathrm{P}(q_r \mid H)$$

by

$$\mathrm{P}(\theta_0 - \tfrac{1}{2}d\theta_0 \leqslant \theta \leqslant \theta_0 + \tfrac{1}{2}d\theta_0 \mid H)$$

and in place of (8.6) we get

$$\mathrm{P}(\theta_0 - \tfrac{1}{2}d\theta_0 \leqslant \theta \leqslant \theta_0 + \tfrac{1}{2}d\theta_0 \mid p, H) \propto \mathrm{P}(\theta_0 - \tfrac{1}{2}d\theta_0 \leqslant \theta \leqslant \theta_0 + \tfrac{1}{2}d\theta_0 \mid H) \times \mathrm{L}(p \mid \theta_0 - \tfrac{1}{2}d\theta_0 \leqslant \theta \leqslant \theta_0 + \tfrac{1}{2}d\theta_0, H). \tag{8.7}$$

If we now require the " best " value of θ, we should, in accordance with Bayes' postulate, take the prior probability to be a constant and once again we should have to maximize L for variations of θ.

We might, however, have chosen to represent our hypotheses, not by θ, but by some quantity ϕ functionally related to θ, e.g. the standard deviation instead of the variance. In this case we should have reached equation (8.7) with ϕ written everywhere instead of θ ; we should have taken the prior probability as constant ; and we should have arrived at the conclusion that we should maximize L for variations of ϕ.

But are we being consistent in so doing ? If we assume that the elementary intervals of θ are equiprobable we cannot assume the same of ϕ, and thus the use of Bayes' postulate appears to involve self-contradiction. The principle of maximum likelihood is free from this difficulty, for if $L(\theta)$ is to be maximized for variations of θ it will, at the same time, be maximized for variation of ϕ, since

$$\frac{\partial L}{\partial \theta} = \frac{\partial L}{\partial \phi}\frac{\partial \phi}{\partial \theta} \tag{8.8}$$

and the two sides of this equation vanish together.

8.10 This is one of the grounds on which adherents of the frequency school have rejected Bayes' postulate in favour of the principle of maximum likelihood ; but in our view the matter has been misunderstood. It would seem that Bayes' postulate and the principle give the same answer in the continuous case as well as in the discontinuous case when proper regard is had to the limiting processes involved. We saw in **7.11** that in speaking of probability in a continuum it was essential to specify the nature of the process to the limit. If we regard θ (from the frequency viewpoint) as having emanated from a population by a process random in the limit for intervals $d\theta$ then Bayes' postulate applied to this process will clearly give a different answer from that obtained by supposing that θ emanated from a process random in the limit for $d\phi$. The two are different just as the probabilities in Example 7.4 are different, and for the same reason. Thus the apparent inconsistency is not an inconsistency at all, but a difficulty introduced by ignoring the limiting process in continuous populations.

It remains true, of course, that for many practical purposes we do not know how the actual value of θ arose. If we require a theory of inference which is unaffected by our ignorance on such points, the objection to Bayes' postulate remains and does not apply to the principle of maximum likelihood. On the other hand there is still the difficulty of adducing convincing reasons why we should adopt the principle of maximum likelihood as a principle of scientific inference. It has many properties, as we shall see in Volume 2, which give it a kind of posterior justification.

Example 8.4

Let us consider Example 8.2 from the likelihood viewpoint. The probability of n white balls on hypothesis (a) is unity ; that on hypothesis (b) is 2^{-n}. The likelihood of (a) is then always greater than that of (b) and hence we choose (a). This is true for any n. In particular, as n increases, the ratio of the likelihoods tends to the ratio of posterior probabilities.

Example 8.5

Consider an independent sample of size n from the normal distribution

$$dF = \frac{1}{\sigma\sqrt{(2\pi)}}\exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}dx.$$

If the observations are $x_1, x_2, \ldots, x_n$ the likelihood function may be written

$$\mathrm{L} = \frac{1}{\sigma^n(2\pi)^{\frac{1}{2}n}}\exp\left\{-\frac{1}{2}\sum_{j=1}^{n}\left(\frac{x_j-\mu}{\sigma}\right)^2\right\}.$$

We consider a range of possible values of μ which could have generated these observations. To estimate μ we take that value which maximizes L, i.e. find the largest value of L satisfying

$$\frac{\partial \mathrm{L}}{\partial \mu} = 0, \quad \frac{\partial^2 \mathrm{L}}{\partial \mu^2} < 0.$$

Since L is positive this is, in general, the same as maximizing log L, sometimes (as here) a more convenient procedure. We then have

$$\frac{\partial \log \mathrm{L}}{\partial \mu} = +\Sigma\left(\frac{x_j-\mu}{\sigma}\right) = 0, \tag{8.9}$$

and thus the estimator of μ, say $\hat{\mu}$, is given by

$$\Sigma x_j = \Sigma\hat{\mu} = n\hat{\mu}$$

or

$$\hat{\mu} = \bar{x}, \text{ the mean of the } x\text{'s}. \tag{8.10}$$

This is easily seen to be a unique maximum, and is therefore the maximum-likelihood solution.

Had we wished to estimate both μ and σ we should find, in addition to (8.9),

$$\frac{\partial \log \mathrm{L}}{\partial \sigma} = -\frac{n}{\sigma} + \Sigma\frac{(x_j-\mu)^2}{\sigma^3} = 0 \tag{8.11}$$

giving

$$\hat{\sigma}^2 = \frac{1}{n}\Sigma(x-\mu)^2. \tag{8.12}$$

Whereas $\hat{\mu}$ does not depend on σ, $\hat{\sigma}$ does depend on μ. We choose those estimators which maximize the likelihood for simultaneous variations in μ and σ, i.e. solve (8.9) and (8.11) together. This gives us

$$\hat{\sigma}^2 = \frac{1}{n}\Sigma(x-\bar{x})^2, \tag{8.13}$$

and (8.10) and (8.13) jointly maximize the likelihood.

EXERCISES

8.1 An event of constant probability ϖ is observed r times in n independent trials. Show that by Bayes' theorem or the method of maximum likelihood the value indicated for ϖ is r/n. (For the application of Bayes' theorem it may be assumed that ϖ is distributed equally frequently from 0 to 1.)

8.2 In the previous Exercise, trials are repeated until r events are observed, the number of trials necessary being n. Show that Bayes' theorem and the method of maximum likelihood give r/n as the appropriate estimate of ϖ.

8.3 Show that r/n, averaged over all samples, equals ϖ in Exercise 8.1, but is not equal to ϖ in Exercise 8.2.

8.4 On the assumption of the previous Exercises show that if n trials are conducted and r is the number of events observed, the probability that the event occurs on the next trial is $(r+1)/(n+2)$. (Use a more general form of the rule of succession.) How do you reconcile this with the result of Exercise 8.1 ?

8.5 An urn is known to contain balls of different colours but the number of colours M is unknown. A sample of n balls is drawn with replacement and is found to contain balls of $m(\leqslant n)$ different colours. Show that the estimator of M indicated by the method of maximum likelihood is m.

8.6 In the previous exercise, use Bayes' theorem to estimate M if the prior probability distribution of M is proportional to 2^M $(M = 1, \ldots, n)$.

CHAPTER 9

RANDOM SAMPLING

The sampling problem

9.1 In the previous chapter we have referred incidentally to the sampling problem, which can be stated quite simply: given a sample from a population, to determine from it some or all of the properties of that population. We noted that only in exceptional cases is it possible to make assertions about the population with complete certainty, and that consequently it is necessary to fall back on statements of a less categorical kind expressible in terms of probability.

9.2 In order to be able to apply the theory of probability to this problem, it is necessary that the sampling should be random. By a random sample, we mean a sample which has been selected in such a manner that every possible sample has a calculable chance of selection. Two remarks are necessary concerning this definition. First, it contains the word " calculable " rather than " calculated," because in practice the chances of selection often need not be exhaustively calculated for every sample—the specification and control of the sampling procedure is all that is required for the application of probability theory. Second, the definition does not imply that every possible sample must have an *equal* chance of selection. If this is in fact the case, and successive drawings are independent, the sampling procedure is called *simple random sampling*, in accordance with the usage of **7.20**. For many theoretical purposes, simple random sampling is the selection process considered, but we shall see in the third volume of this book that the search for efficiency in the design of sampling inquiries often leads to the deliberate abandonment of simple random sampling. The point here is that such abandonment does not impair the application of the theory of probability, so long as the (unequal) chances of selection are uniquely determined by the sampling process, and are therefore calculable.

9.3 In actual practice, we often meet with samples which are not random, having been chosen purposively in some way. In such circumstances, it is not possible to make precise probability statements about the parent population, and where a decision has to be taken one is forced to rely on subjective judgments of an unsatisfactory kind. It is for this reason that random sampling is of primary importance in sampling investigations of a population.

Henceforth, we shall discuss only random samples, and to avoid constant repetition we shall understand the unqualified use of the words " sample " or " sampling distribution " to refer to simple random sampling conditions.

9.4 It is useful to begin a discussion of random sampling by considering the types of parent population from which samples can be chosen.

(a) In the first place, the population may be finite and existent, e.g. the population of human beings in Europe at a fixed point of time, or the population of apples on a given tree. A sampling process which extracts members one at a time from this population will evidently eventually exhaust the supply of members if continued long enough. Thus the sampling, though random, is not simple, for successive drawings are not independent.

We may, however, reduce this process to one of simple sampling by replacing the members after withdrawal. The population then remains the same at each trial. The two cases are sometimes distinguished as " sampling without replacement " and " sampling with replacement."

Furthermore, we may also in many cases regard the sampling as simple to an adequate approximation even when there is no replacement. If the population is large compared with the size of the sample, the abstraction of relatively few members will not materially affect the constitution of the remaining population, which may thus be regarded as approximately the same for subsequent samplings.

(b) Sampling with replacement from a finite population may, in fact, be regarded as sampling from an infinite population, for the process will never exhaust the supply. We may, however, have to deal with a population which is infinite in rather a different sense, namely, that of a limiting form. We may, for example, wish to consider the probability of a sample from the positive integers or the real numbers from 0 to 1. The latter case presents itself in sampling from a continuous frequency-distribution which we must necessarily regard as infinite.

Thus, if we replace an observational distribution by a conceptual continuous mathematical distribution, we replace at the same time a finite population by an infinite population. The drawing of random samples from such a population is attended by the circumstances referred to in **7.11**, namely, that the process to the limit must be taken into account.

(c) Thirdly, the population may be purely hypothetical. Consider, for example, the throws of a die. We may picture the continual throwing as a sampling process drawing existent members from some non-existent population. In such cases what we are really doing is constructing by mental fiction an imaginary population round the sample.

The concept of the hypothetical population is necessitated by ideas of frequency in probability. It is not required (and indeed has been explicitly rejected by Jeffreys) in the approach which takes probability as an undefinable measurement of attitudes of doubt. But if we take probability as a relative frequency, then to speak of the probability of a sample such as that given by throwing a die or growing wheat on a plot of soil, we must consider the sample against the background of a population. There are obvious logical difficulties in regarding such a sample as a selection—it is a selection without a choice—and still greater difficulties about supposing the selection to be random ; for to do so we must try to imagine that all the other members of the population, themselves imaginary, had an equal probability of assuming the mantle of reality, and that in some way the actual event was chosen to do so. This is a rather baffling conception. At the same time, it has to be admitted that certain events such as dice-throwing do happen as if the constituents were chosen at random from an

existent population, and it accordingly seems that the concept of the hypothetical population can be justified empirically.

Randomness in sampling practice

9.5 In its colloquial use the word " random " is applied to any method of choice which lacks aim or purpose. We speak of drawing names at random out of a hat, choosing plants at random from a field of corn, selecting family budgets at random from the population, meaning thereby that the selection is completely haphazard.

Now it is found in practice that choice by a human being is not random in the stricter sense that it produces equally frequently events which we are entitled to expect to have equal prior probabilities. Some examples will make this clear.

Example 9.1

In the course of certain work at the Rothamsted Experimental Station, sets of eight wheat plants were chosen for measurement. Six of these were chosen by approved methods, referred to below, and may be taken to be truly random. The other two were chosen haphazardly by eye. If, in any set, the eight plants were ranged in order of magnitude, the two selected by eye could have any number from one to eight; and if they, in common with the other six, were chosen at random, they should occupy these places with approximately equal frequency in a large number of sets. Table 9.1 shows what actually occurred on two different occasions (a) on May 31st, before the ears of wheat had formed, and (b) on June 28th, after the ears had formed.

Table 9.1—Distribution of plants chosen haphazardly in ranks 1 to 8 (*Yates*, 1935)

Date	Observation	Numbers bearing specified rank								Total
		1	2	3	4	5	6	7	8	
May 31st	Shoot height . .	9	7	11	8	11	18	21	31	116
June 28th	Ear height . . .	9	19	27	23	15	10	5	4	112

The divergence of actual from expected results is quite striking. On May 31st, before the ears had formed, the observer was strongly biassed towards the taller shoots; whereas in June he was biassed strongly towards the central plants and avoided short and tall plants.

Thus it is seen that bias can appear even in a trained observer, and that the bias need not be consistent in over- or under-estimation in different circumstances.

Example 9.2

Table 9.2 shows the frequencies of final digits in a number of measurements made by four different observers.

It is hard to suppose that there was any genuine difference which would lead to the appearance of certain digits at the expense of others, and we may confidently

Table 9.2—Bias in scale reading : distribution of final digits in measurements by four observers (*Yule*, 1927)

Final digit	Frequency of final digit per 1000			
	A	B	C	D
0	158	122	251	358
1	97	98	37	49
2	125	98	80	90
3	73	90	72	63
4	76	100	55	37
5	71	112	222	211
6	90	98	71	62
7	56	99	75	70
8	126	101	72	44
9	129	81	65	16
TOTAL	1001	999	1000	1000

suppose that the deviations from approximate equality indicate bias on the part of the observer.

Observer A had decided preference for 0, 2, 8 and 9, avoiding the centre of the scale. Observer B is quite good, his deviations from expected values being small, though he also showed some preference for 0. Observer C was poor, rounding off one measurement in two to the whole or half unit. Observer D was obviously very bad indeed, nearly 57 per cent. of his measurements being rounded off to the whole or half unit.

The observations were all made by reading a scale, those under A being on drawings to the nearest tenth of a millimetre, those under B, C and D being measurements on the heads of living subjects to the nearest millimetre. We may conclude from this that different observers may exhibit different degrees of bias even under comparable circumstances, and that even those who are aware of the existence of the possibility of bias and the necessity for taking great care (as observer A was) may nevertheless fail to avoid it.

Example 9.3

An observer was placed before a machine consisting of a circular disc divided into ten equal sections in which were inscribed the digits 0 to 9. The disc rotated at high speed and every now and then a flash occurred from a nearby electric lamp of such short duration that the disc appeared at rest. The observer had to watch the disc and write down the number occurring in the division indicated by a fixed pointer.

This was a machine designed to produce random numbers, and had been found by another observer to do so. But this particular observer produced a definite bias. The frequencies of digits in 10,000 run off by him are shown in Table 9.3.

Table 9.3—Distribution of digits obtained by an observer in using a randomizing machine (*Kendall and Babington Smith*, 1939)

Digit . . .	0	1	2	3	4	5	6	7	8	9	TOTAL
Frequency . .	1083	865	1053	884	1057	1007	1081	997	1025	948	10,000

P

If the observer was unbiassed the digits should appear in approximately equal numbers; but there is a bias in favour of all the even numbers and against the odd numbers 1, 3 and 9. The cause of this bias is obscure, for the observer did not have to estimate (as in the previous example) but merely to write down something which he saw, or thought he saw. The explanation seemed to be that he had a strong number-preference, i.e. that he actually mis-saw the numbers, or that his brain controlled his ocular impressions and censored them. We have here to deal with one of the deadliest forms of bias in psychology.

Example 9.4

Every year a number of crop reporters in England and Wales estimate the prospective yields of certain crops, forecasts being obtained at different periods of the year and final estimates when the crop is harvested. Table 9.4 shows the average estimated yield of potatoes at the various times for the years 1929–1936.

Table 9.4—Bias in crop forecasting : forecasts of yields of potatoes in England and Wales (tons per acre) (*from the official agricultural statistics*)

Year	Sept. 1st		Oct. 1st		Nov. 1st		Final estimate
	Yield	% difference from final	Yield	% difference from final	Yield	% difference from final	
1929	5·7	−17·4	6·2	−10·1	6·5	−5·8	6·9
1930	6·0	− 7·7	6·1	− 6·2	6·1	−6·2	6·5
1931	5·5	0·0	5·3	− 3·6	5·3	−3·6	5·5
1932	6·4	− 3·0	6·2	− 6·1	6·3	−4·5	6·6
1933	6·4	− 4·5	6·2	− 7·5	6·4	−4·5	6·7
1934	6·0	−15·5	6·3	−11·3	6·7	−5·6	7·1
1935	5·6	− 9·7	5·7	− 8·1	6·0	−3·2	6·2
1936	6·0	− 3·2	5·9	− 4·8	5·8	−6·5	6·2

This table exhibits very clearly an effect which has shown itself in nearly all the English crop reports (and appears also in other countries), namely, the chronic pessimism of crop forecasts. In every case but one in the table, the forecasts are below the final estimate. Nor do crop reporters seem able to learn by experience that they are underestimating. Nothing in this table indicates that the differences between forecast and final estimate diminished during the period concerned.

It should also be noticed that these estimates are the weighted average of a large number of independent observations. One of the commoner misunderstandings in this type of work is based on the supposition that, though individuals may make mistakes, their errors will cancel out in the aggregate. Our present example shows this to be untrue in general. There can appear a systematic bias affecting all the individuals making estimates.

9.6 The foregoing examples are enough to indicate that human bias is very prevalent. Trained observers may be biassed even when conscious of their own imperfec-

tions ; different observers may be biassed in different ways in similar circumstances ; and the same observer may be biassed in different ways in different circumstances. It is abundantly clear that we must look for true randomness elsewhere than in mere lack of purpose on the part of human observers. There may be persons whose psychological processes are so finely balanced that they can deliberately select random samples, but few statisticians who have experimented in this interesting field would regard themselves as so gifted.

9.7 In Chapter 7 we saw that the primary function of randomness in probability was that it ensured that certain primitive events were equally probable. We may say that a method of selection is random for a population U if, when applied to U, it gives all members an equal probability of being chosen ; or, in the language of frequency, if, when continually applied to U, it educes the members approximately equally frequently.

But this is not enough. Suppose we had a population of two members A and B, and sampled with replacement. Then a method which chooses A and B alternately and produces the series $ABAB\ldots$ educes each member approximately equally frequently ; but it is not what we customarily mean by a random method. What we require of a random method is that in such circumstances it should produce a series in which no systematic arrangement is evident. Not only single characteristics, but all possible groups of characteristics should appear equally frequently.

9.8 A further point is to be noted. We may, in drawing the sample, be interested in one particular variate exhibited by the members, and it is possible that a method may give a satisfactory random sample *so far as this variate is concerned* without doing so for other variates. Suppose, for example, we are anxious to take a random sample from the inhabitants of a particular area. If we are concerned with a variate such as eye-colour it might be sufficient to choose a house every so often, say every tenth house, and select one inhabitant of that house as part of the sample. Such a method would not give every inhabitant of the area an equal chance of being chosen ; but if we look back to the time when the inhabitants took up residence, we may imagine that the colour of their eyes did not influence their geographical distribution, and thus if we consider the distribution of the inhabitants into houses as independent of eye-colour, we may suppose that so far as eye-colour is concerned the sample is random. But the matter would stand differently if we were sampling for income, for we may reasonably expect poorer people to live in more crowded conditions, and a sample of one person from each house chosen would therefore under-represent the poor. Thus our sample would not be random with respect to income.

Thus a method which may properly be deemed to be random for one variate may not be so for another.

The technique of random sampling

9.9 Suppose, then, that we are given a population and a variate is specified. How are we to draw a random sample, i.e. how can we find a method which is random for that population and that variate ? The answer lies partly in theory and partly in practice.

(a) In the first place we must require that there is no obvious connection between the method of selection and the properties under consideration. The method and the properties must be independent so far as our prior knowledge is concerned. In sampling a field of wheat for shoot height, for example, we must not use a method which could be influenced by that height, such as skimming a hoop over the field and selecting the plants round which it fell (for the hoop might tend to catch on the taller plants). Again, in sampling the inhabitants of a town by choosing names from a telephone directory, we should undoubtedly tend to get the more well-to-do classes and hence, if the variate under consideration is wealth or any related characteristic such as number of children, political opinion, standard of education and so on, the sample would not be random. If we were concerned with characteristics such as height, hair colour, or blood group the sample might be random, though it is not difficult in many similar cases to think of reasons why the variate might be linked with wealth.

If this matter is viewed from the standpoint of the subjective theory of probability the absence of knowledge about relationship between the method of selection and the characteristic under consideration may be sufficient to ensure randomness, for the probabilities of elementary propositions then become equal—the probabilities being measures of prior attitudes of mind.(*) But if the frequency viewpoint is adopted, it is not enough that there should be absence of knowledge of this kind, for unknown to the observer there may be relations which will prevent the elementary propositions from being true in approximately equal proportions. The presumption is that if we make as great an effort as possible to ascertain whether any relationship exists and fail to find it, there is no relationship ; and hence we can assume the randomness of the method with more or less confidence. But in this approach the assumption of randomness is ultimately part of the general uncertainty of the inference from sample to population.

(b) Secondly, we may rely on previous experience of a random method of selection to justify its use on new occasions. This is evidently an extrapolation, and though most people would regard it as reasonable, the fact has to be realized. The subjective theory of probability can embrace this extrapolation within its scope, for the probabilities given by the method are assessable in terms of prior knowledge ; but the frequency theory has to take the extrapolation as an additional assumption.

9.10 One of the most reliable methods of drawing random samples consists of constructing a model of the population and sampling from the model. We may, for instance, note down the characteristics of each member on a card and sample by choosing cards from the pack corresponding to the whole population. This is the method adopted in lotteries and the process is known as lottery or ticket sampling. It is moderately effective but suffers in practice from two disadvantages : the labour of constructing the card population, and the danger of bias in the drawing of cards. To be reasonably satisfied about the randomness of the shuffling entails a good deal of

(*) At least, this is our interpretation of the position ; but the writers on the subjective theory have not discussed practical sampling problems at any length, and we may be putting a gloss on their views which they would not accept.

trouble and labour, and the same object can be attained much more simply by the use of random sampling numbers, which we now consider.

Random sampling numbers

9.11 The easiest way of constructing a miniature population is to attach a positive integer to each member, most simply by numbering the members from 1 onwards. The set of integers so obtained is the miniature population and the problem of drawing a random sample reduces to finding a series of random numbers. The advantages of this method are obvious: no physical model population has to be constructed; the numbering can be carried out in any convenient manner; and the series of random numbers can be applied to any enumerable population so that any series of random numbers has a very wide range of application.

One point should be made clear here. If the numbering of the population is carried out in such a way as to be independent of certain characteristics of the population, any set of numbers will serve to draw a sample random with respect to those characteristics. The randomness in such a case lies, so to speak, in the allocation of integers to the population, not in deciding which ordinals to select for the sample. But in practice a procedure of this kind is of no value, since it only throws back to the difficulty of numbering the population " at random." The usual course is to number the population in any convenient way, related to the characteristics or not, and then seek for a set of numbers which are a random set from the possible integers of the population.

9.12 One of the more obvious ways of drawing random samples from an enumerated population is to use haphazard numbers taken from some totally unrelated source. Suppose, for instance, we wished to take a sample from the visible stars in the sky. We will ignore the small complications due to the existence of double stars and unresolved objects. Since the position of a star on the celestial sphere is defined by latitude and longitude, what is then required is a series of random pairs of latitudes and longitudes. At first sight it seems plausible to take an ordinary atlas and choose the figures set out in the index for place-names arranged alphabetically; for there is little reason to expect any relationship between the distribution of stars in the sky and the distribution of places on the Earth's surface. A little reflection, however, will show that the method is unsound. There are large stretches of territory and sea on the Earth which have no place-names on them—the poles, deserts and oceans; consequently no numbers will occur for these regions and there will be corresponding areas on the celestial sphere which have no chance of being included.

9.13 As a next attempt we might take a book containing a number of digits, e.g. a telephone directory, or a set of statistical tables or mathematical tables, open it at hazard and choose the digits which first strike the eye, or which occur at the top of the page, and so on. This is an improvement, but it is still open to some objection.

(a) *Telephone directories.* Table 1.4 on page 6 shows the distribution of 10,000 digits taken from the London telephone directory. Pages were chosen by opening the directory haphazardly; numbers of less than four digits and numbers in heavy type were ignored; and of the four-figure numbers remaining the two right-hand ones

were taken for all numbers on the page. If the numbers were random we should expect about 1000 of each digit in the total of 10,000. Actually there are very considerable deviations from this expectation, and we shall see in a later chapter that they cannot be explained as sampling fluctuations. There are significant deficiencies in 5's and 9's, due to several causes such as the tendency to avoid these digits because they sound alike, the reservation of numbers ending in 99 for testing purposes by telephone engineers and so on. It is evident that tables of random numbers could not be constructed from directories such as this.

(b) *Mathematical tables.* Evidently care has to be exercised in using mathematical tables in constructing random series. Suppose, for instance, we take a set of logarithm tables. There are clearly relationships between successive logarithms, expressible by the fact that differences are approximately constant if the interval is small. Moreover there is a very curious theorem about digits in certain classes of table which throws theoretical doubt on the method. Consider the logarithms to base 10 of the natural numbers from 1 onwards. Suppose we choose the kth digit in each and so obtain a series of digits 0–9. Then the proportional frequency of any digit in this series does *not* tend to a limit as the length of the series increases, whatever k may be.(*) Just what does happen does not appear to be known, but it would seem that certain systematic effects begin to show themselves and these will obviously endanger the randomness of the series.

(c) *Statistical tables.* If we have a volume of statistics such as populations of towns and rural districts, there are some grounds for supposing that if the numbers are large —say, four figures or more—the final digits will be random. Here again, however, the use of such tables requires care—they may have been compiled by an observer with number preferences, and some rounding up may have taken place.

9.14 However, the necessity for the ordinary student to construct random series of his own has been obviated by the publication of various tables of Random Sampling Numbers. There are several such available :

(a) Tippett's numbers comprise 41,600 digits taken from census reports combined into fours to make 10,400 four-figure numbers (*Tracts for Computers*, No. 15).

(b) Kendall and Babington Smith's numbers comprise 100,000 digits grouped in twos and fours and in 100 separate thousands (*Tracts for Computers*, No. 24). These numbers were obtained from a machine specially constructed for the purpose on the lines very briefly described in Example 9.3.

(c) Fisher and Yates' numbers comprise 15,000 digits arranged in twos (*Statistical Tables for Biological, Agricultural and Medical Research*). These numbers were obtained from the 15th–19th digits in A. J. Thompson's tables of logarithms and were subsequently adjusted, it having been found that there were too many sixes.

(d) The Rand Corporation has published *A Million Random Digits* (1955) arranged in groups of five. They were constructed by a kind of electronic roulette wheel ; but it is interesting to observe that even after searches for systematization had been made in the circuits, the numbers were still not quite random and had to be adjusted.

(*) Cf. J. Franel, *Vierteljahrschrift der Naturforschenden Gesellschaft in Zürich* (1917), **62**, 286.

The above-mentioned tables consist of randomized digits. Wold (*Tracts for Computers*, No. 25) has compiled a table of 25,000 normal deviates, by converting the Kendall–Babington Smith tables with the aid of tables of the normal integral. Fieller and others (*Tracts for Computers*, No. 26) have given 27,000 pairs of normal deviates correlated to given extents. The Rand Corporation's book also contains 100,000 normal deviates.(*)

9.15 Before considering the basis of these tables it may be helpful to give some examples of their use. Here are the first 200 of the Kendall–Babington Smith tables:—

Table 9.5—Random sampling numbers (*Tracts for Computers*, No. 24)

23 15	75 48	59 01	83 72	59 93	76 24	97 08	86 95	23 03	67 44
05 54	55 50	43 10	53 74	35 08	90 61	18 37	44 10	96 22	13 43
14 87	16 03	50 32	40 43	62 23	50 05	10 03	22 11	54 38	08 34
38 97	67 49	51 94	05 17	58 53	78 80	59 01	94 32	42 87	16 95
97 31	26 17	18 99	75 53	08 70	94 25	12 58	41 54	88 21	05 13

Example 9.5

To draw a sample of 10 men from the population of 8585 men of Table 1.7.

The first process is to number the population; and here, as in most similar cases, one numbering has already been provided by the frequency-distribution. We take numbers 1 and 2 to be those in the group 57– inches, numbers 3 to 6 those in the group 58–, and so on, those in the group 77– inches being numbers 8584 and 8585.

Now we take 10 four-figure numbers from the tables, e.g. reading across in Table 9.5 we have 2315, 7548, 5901, 8372, 5993, 7624, [9708], [8695], 2303, 6744, 0554, 5550.

The two numbers in square brackets are greater than 8585 and we ignore them. We now select the individuals corresponding to the remaining 10 numbers. They will be found to be in the intervals 65–, 70–, 68–, 72–, 68–, 70–, 65–, 69–, 63–, 68– inches respectively.

The mean of these values considered as located at the centres of intervals is 68·24, as against a value in the population of 67·46.

Example 9.6

To draw a sample of 12 from the population in the following bivariate table, showing the relation between inoculation and attack in cholera.

	Not attacked	Attacked	TOTAL
Inoculated	276 (0001–3312)	3 (3313–3348)	279
Not inoculated . . .	473 (3349–9024)	66 (9025–9816)	539
TOTALS	749	69	818

(*) Quenouille (1959, *Biometrika*, **46**, 178) gives 1000 random deviates from each of eight distributions, these being the normal, lognormal, exponential, double exponential, rectangular, and three Edgeworth expansions.

There are now 818 members. We could, of course, take three-figure numbers from the tables, obtaining, e.g. from Table 9.5,

231, 575, 485, etc.

But this is rather troublesome as the numbers are not grouped in threes. It is more convenient to take four-figure numbers as before and to associate each member of the population with 12 numbers in the tables, e.g. the first would correspond to 0000–0011, the second to 0012–0023, and so on. We then get the numbers shown in brackets in the above table. We ignore numbers above 9816 as before.

The two numbers omitted in the previous Example can now be used, and, from all 12 numbers listed there, we find the following results:

	Not attacked	Attacked	TOTAL
Inoculated	3	0	3
Not inoculated . . .	8	1	9
TOTALS	11	1	12

Here, for example, the member corresponding to the number 2315 falls in the not-attacked : inoculated class, and so on.

It has so happened in this example that no member in the very small class inoculated : attacked class has been selected. Suppose we had had a series containing

3314, 3323, 3333, 3341.

All these fall into the group and there are four of them, as against only three members in the population. Had we been confronted with this position we should have had to decide whether the sampling was to be with or without replacement. If it was without replacement, we should have to suppose that the first three numbers in the group 3313–3348 exhausted that part of the population and ignore all numbers of the group occurring subsequently.

While on this point, we might remark that the use of random sampling numbers involves sampling with replacement as a general rule. If we wish to sample without replacement we should not use the same number twice. This involves keeping a record of what has already been used, a tedious procedure except in cases where the sample or the population is small.

Example 9.7

To construct a series of random permutations of the digits 1 to 5.

Here we are not concerned with the digits 0, 6, 7, 8 and 9 and so ignore them in the table of random numbers. We read through the table and note the digits as they occur, e.g. in Table 9.5 we have 2315, 7548, etc. The 7 is to be ignored and also the second 5, for one 5 has already occurred. We then reach the permutation 23154. Then we start again, the next series being 8, 5901, 8372, 5993, 7624, etc., giving the permutation 51324; and so on.(*)

(*) Tables of 400 random permutations of the integers 1 to 20 are given in Kendall's *Rank Correlation Methods*, 3rd edn. (Griffin, 1962). Permutations of fewer than 20 can be derived

Example 9.8

To take a random sample of 10 from the normal population $dF = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2} dx$.

This is a particularly interesting case, for we have to select a sample from an infinite population. Such a process, as has been seen, can only be considered as a limiting one.

Suppose that we divide the distribution into variate-ranges of width 0·1. We obtain from a table of the normal integral (e.g. Appendix Table 2):

Variate-value x	$F(x)$	x	$F(x)$
$-\infty$	0·00000	0·0	0·5000
.	.	.	.
.	.	.	.
.	.	.	.
−3·7	0·00011	0·1	0·5398
−3·6	0·00016	0·2	0·5793
−3·5	0·00023	0·3	0·6179
.		.	
.		.	
.		.	
−0·4	0·3085	3·5	0·99977
−0·3	0·3821	3·6	0·99984
−0·2	0·4207	3·7	0·99989
		.	.
		.	.
		.	.
−0·1	0·4602	∞	1·00000

We may now use a four-figure random number to select a value of x through the population formed by $F(x)$: e.g. the number 5461 corresponds to a variate-value +0·1– and the number 3500 to −0·4–.

Had we taken the table to n places of decimals we should have required n– figure numbers. Furthermore, we can make the approximation more exact by taking a finer variate interval. Such matters as this are to be decided in the light of the degree of approximation required.

9.16 Random sampling numbers must obey certain conditions before they can be used. Any set of numbers whatever is random in the sense that it might arise from random sampling; but such a set might not be suitable as a table of random sampling numbers. From the examples already given it is clear that we desire such a table to have very great flexibility. It should give random results in as many cases as possible, whether used in part or in whole.

Now it is impossible to construct a table of random sampling numbers which will satisfy this requirement entirely. Suppose, to take an extreme case, we constructed a table of $10^{10^{10}}$ digits. The chance of any digit being a zero is $\frac{1}{10}$ and thus the chance

by omitting unwanted integers, and there can be derived by obvious procedures 800 permutations of the integers 1 to 10, 1600 of the integers 1 to 5 and so on. Cochran and Cox, in their *Experimental Designs* (Wiley, 1950), give 1000 permutations of each of 9 and 16 integers.

that any given block of a million digits are all zeros is 10^{-10^6}. Such a set should therefore arise fairly often in the set of $10^{10^{10}-6}$ blocks of a million. If it did not, the whole set would not be satisfactory for certain sampling experiments. Clearly, however, the set of a million zeros is not suitable for drawing samples in an experiment requiring less than a million digits.

Thus, it is to be expected that in a table of random sampling numbers there will occur patches which are not suitable for use by themselves. The unusual must be given a chance of occurring in its due proportion, however small. Kendall and Babington Smith attempted to deal with this problem by indicating the portions of their table (5 thousands out of 100) which it would be better to avoid in sampling experiments requiring fewer than 1000 digits.

9.17 If a table of random numbers is used to draw members from a population of ten, we expect the members to appear in approximately equal proportions. In other words we expect such a table to contain the ten digits 0–9 in approximately equal proportions. Similarly we expect the hundred pairs 00-99 to appear in approximately equal proportions, and so on. Various tests of this kind, based on a comparison between actual frequencies and those required to satisfy the laws of probability, can be devised. No table can satisfy them all, but if it satisfies tests which (a) ensure the randomness of the numbers for the commoner types of sampling inquiry for which it is likely to be used and (b) are capable of revealing any particular sort of bias to which the numbers are susceptible in virtue of their mode of formation, it is likely to be of general application.

9.18 The development of high-speed computers has made it possible to incorporate a " randomizing unit " into some machines which will select random samples *ad hoc*. The use of such mechanisms requires investigation ; it seems difficult to generate sets of numbers which pass the usual tests for randomness.

A further method which has been employed is to raise numbers to very high powers and to select sets of digits from the middle of the very large numbers so generated. Such " pseudo-random " numbers also may fail to pass tests for randomness. On the other hand, there are circumstances where they may be useful in sampling experiments, for one of two distinct reasons : they may be " nearly " random ; or, if enough are taken, they may enumerate all possible samples, though in an irregular order. It is unsafe to generalize on this subject without considerable experience.

Sampling from a continuous population

9.19 Random sampling numbers offer the best method known at the present time of drawing random samples from an enumerable population, and as was seen in Example 9.8, may also be used to draw samples from a continuous population specified mathematically. But cases sometimes occur in which they cannot be employed. For instance, if we wish to take a sample of milk or flour, we cannot in practice number each particle and extract it from the population for examination. In such cases we are usually compelled to fall back on more intuitively founded procedure. To take a random sample from a milk churn, for instance, we might stir the contents thoroughly and scoop up a

sample haphazardly. Sometimes, when the population is of manageable size, we can proceed systematically by dividing it into a number of parcels and selecting parcels by the ordinary technique of random numbers. Most sciences have their own peculiar sampling problems and no attempt can be made here to discuss them all. At this point we leave the technique of random sampling and assume hereafter, unless the contrary is stated, that the material we are discussing has been obtained by a random process. We shall return to the comparison of the efficiencies of different sampling methods in Volume 3.

Sampling for attributes

9.20 As an introduction to the general sampling problems we shall consider the sampling of attributes, which raises many of the difficulties of principle but is not obscured by too much mathematics.

Suppose we have a random sample from a population whose members all exhibit either an attribute A or its negative not-A. Our sample is n in number, and a proportion p, or a number pn, exhibit the attribute; and consequently a proportion q, or number qn $(p+q = 1)$ do not. We will assume that the population is large, or that sampling is with replacement, so that the probability of obtaining an A at any drawing is not affected by other drawings and is therefore a constant, say ϖ.

The problems we have to consider are of three types:—

(a) Suppose we have some reason for supposing that the proportion of A's in the population is given by a known ϖ. Does the observed proportion p bear out this hypothesis or is it so divergent from ϖ as to lead us to doubt the hypothesis? For example, in an experiment with plants exhibiting two strains of a quality such as height in pea plants, we may wish to test whether the breeding follows the simple Mendelian law of dominant and recessive. If we begin with two pure strains tall and short, cross-breed a first generation and then produce a second generation by interbreeding, the proportional frequencies of "short" and "tall" in this generation will be $\frac{3}{4}$ and $\frac{1}{4}$ if "short" is dominant and $\frac{1}{4}$ and $\frac{3}{4}$ if "tall" is dominant, provided that the simple Mendelian law holds. Suppose we carry out such an experiment and find that for 400 plants the frequencies are 70 and 330. Can the divergence from the theoretical values 100 and 300 have arisen by chance, or is it large enough to throw doubt on the hypothesis that the simple Mendelian law is operating?

(b) In the foregoing type of problem we have some reason for testing a value of ϖ given *a priori*; but we may know nothing of ϖ, and in such a case our principal problem is to estimate it from the sample.

(c) Then, having estimated it, we wish to know the degree of reliability of the estimate. How far is the estimate likely to deviate from the real value of ϖ?

9.21 Consider the first type of problem, in which ϖ is given *a priori*. If we select repeated samples of size n from the population, the distribution of the number of A-individuals in the sample will be arrayed by the binomial expansion of $(\chi+\varpi)^n$, where $\chi = 1-\varpi$. (This is, in fact, the binomial distribution dealt with in **5.2**, but we are now using Greek symbols instead of p and q to distinguish the fact that these are population parameters.) The probability of obtaining np or fewer A-individuals

is the sum of the first $(np+1)$ terms in this distribution. Call this P. We test the hypothesis by determining beforehand a (generally small) probability of error, say α, which we are prepared to tolerate in rejecting the hypothesis, and compare P with α. If P is the larger, we do not reject the hypothesis, while if α is the larger we do.

Superficially, this procedure is a reasonable one; we reject the hypothesis if the observed value of the statistic is located in the most " unlikely " portion of its sampling distribution. However, a satisfactory analysis of the logic of test inference must await a more sophisticated discussion in the second volume.

It will have been noted that P, determined as above, relates to only one of the " tails " of the sampling distribution. It is, in fact, more common in practice for tests to employ a P calculated from *both* tails of the distribution. That is to say, quite frequently we are interested in departure from the hypothesis under test which would have the effect either of decreasing or of increasing the expected number of A-individuals in the sample, while on other occasions we are interested only in departures in one direction. Once again, full discussion of the rationale of these procedures must be deferred to the second volume. Here, it is sufficient to note that in our Mendelian law example we should be following the intuitively sound procedure in regarding only low numbers of " tall " plants as providing evidence against the hypothesis that " tall " is dominant.

Example 9.9

In certain coin-tossing experiments a coin was tossed 20 times and came down " heads " 15 times. Does this conflict with the hypothesis that the coin was unbiassed ?

Here we have to test the hypothesis that $\varpi = \frac{1}{2}$. Bias in the coin can operate in either direction (towards " heads " or " tails "), so we calculate P including both extremes of the sampling distribution. The probability that in 20 tosses we should get 15 or more heads is the sum of the first six terms in $(\frac{1}{2}+\frac{1}{2})^{20}$, and from Table 5.2 this is 0·0207. The probability of 15 or more tails is, by symmetry, the same. P is therefore 0·0414. If we had only been prepared to tolerate a risk of wrongly rejecting the hypothesis which is lower than this, the hypothesis would not be rejected. In the contrary case, the hypothesis would have been rejected. Commonly, a risk of ·05 is tolerated, and in this case the hypothesis would be rejected.

9.22 In the example just given we purposely took a fairly low value of n in order that the terms of the binomial could be calculated directly. In practice n is often fairly large—100 or more—and the evaluation and summation of individual terms would be most tedious. We can, if complete accuracy is desired, use the method of summation given in **5.7**, and evaluate from the incomplete B-function or from the tables listed there. But for all ordinary purposes it is quite enough to use the normal approximation to the binomial. We saw in Example 4.6 that as $n \to \infty$ the binomial tends to the normal distribution, with mean $n\varpi$ and variance $n\varpi\chi$ in our present notation. Probabilities can therefore be evaluated from the normal integral. In fact, for many purposes, it is not even necessary to carry out the actual evaluations. From the tables of the integral (Appendix Table 2) we note that the probability of an absolute deviation as great as, or greater than, the standard deviation is 0·3173; for twice the s.d. is

0·0455 ; for thrice the s.d. is 0·0027 ; for four times the s.d. is 0·00006. Thus, if we set our region of rejection outside $n\varpi \pm 2\sqrt{(n\varpi\chi)}$, we shall have a probability of 0·0455 of wrongly rejecting the hypothesis and similarly for other multiples of $\sqrt{(n\varpi\chi)}$.

It makes no difference whether we compare the actual frequencies or the proportions, since mean and standard deviation are equally affected when the variate is divided by a constant.

Example 9.10

In some dice-throwing experiments Weldon threw dice 49,152 times, and of these 25,145 yielded a 4, 5, or 6. Is this consonant with the hypothesis that the dice were unbiassed ?

If the dice are unbiassed the probability of a 4, 5 or 6 is $\frac{1}{2}$. Thus $n\varpi$ is 24,576 and the observed np is 569 in excess of this value.

$$\sqrt{(n\varpi\chi)} = \sqrt{(49{,}152 \times \tfrac{1}{2} \times \tfrac{1}{2})} = 110{\cdot}9.$$

The observed deviation is more than 5 times this quantity and we accordingly suspect very strongly that the dice were biassed.

Standard error

9.23 The quantity $\sqrt{(n\varpi\chi)}$ is a particular case, appropriate to the binomial, of an important statistical concept known as the standard error. It is the standard deviation of the sampling distribution of the statistic, here np. It is particularly important in the class of cases, which is relatively large, wherein that sampling distribution can be taken to be normal either exactly or to an adequate degree of approximation.

9.24 Let us now turn to the case (b) in which no value of ϖ is given *a priori*. If the sample gives a proportion of A's equal to p, what shall we take as our estimate of ϖ ? The most obvious course is to take p itself ; and this is the course dictated by the more sophisticated ideas described in Chapter 8.

Consider first of all the method of maximum likelihood. The probability of obtaining np A's and nq not-A's is

$$\binom{n}{np} \varpi^{np} \chi^{nq}. \tag{9.1}$$

This is proportional to the likelihood, and neglecting constants we have to maximize

$$\mathrm{L} = k\varpi^{np}(1-\varpi)^{nq}$$

for variations of ϖ. Since L is non-negative, $\dfrac{\partial \mathrm{L}}{\partial \varpi}$ and $\dfrac{\partial}{\partial \varpi}(\log \mathrm{L})\left(= \dfrac{1}{\mathrm{L}}\dfrac{\partial \mathrm{L}}{\partial \varpi}\right)$ vanish together and it is therefore sufficient to maximize log L. We have

$$\frac{\partial}{\partial \varpi}(\log \mathrm{L}) = \frac{np}{\varpi} - \frac{nq}{1-\varpi} = 0$$

giving, as our estimator $\hat{\varpi}$,

$$\hat{\varpi} = p. \tag{9.2}$$

The method of Bayes will give the same result if we suppose the possible values

of ϖ equally distributed between 0 and 1 for $d\varpi \to 0$. For then

$$\mathrm{P}(\varpi \mid p) \propto \mathrm{P}(\varpi \mid H)\mathrm{P}(p \mid \varpi H) \tag{9.3}$$

$$\propto \varpi^{np}\chi^{nq}\,d\varpi \tag{9.4}$$

which, as before, is maximized when $\varpi = p$.

There is another way of looking at this problem of estimation. Suppose we took a large number of samples of the same size n from the population with a proportion of ϖ A's and χ not-A's. Our estimate of ϖ would be p in each case, p varying from sample to sample; and the mean value of all such estimates would be

$$\begin{aligned}
\mathrm{E}(p) &= \sum_{p=0}^{1} p\binom{n}{np}\varpi^{np}\chi^{nq} \qquad (9.5)\\
&= \frac{1}{n}\Sigma\left\{np\binom{n}{np}\varpi^{np}\chi^{nq}\right\}\\
&= \frac{1}{n}\varpi.n\Sigma\left\{\binom{n-1}{np-1}\varpi^{np-1}\chi^{nq}\right\}\\
&= \varpi\{\varpi+\chi\}^{n-1}\\
&= \varpi,
\end{aligned}$$

so that the mean value of our estimate over all possible samples is ϖ. Such an estimate is called *unbiassed*—if we follow the rule of estimation the average of our estimates in a large number of cases will be exactly the correct value ϖ. It may thus be argued that the unbiassed estimate should be taken as a reliable estimate of ϖ.

9.25 In this case, therefore, all the approaches lead to the same conclusion (a happy state of affairs which, as we shall see in the sequel, does not always exist). Consider now the next stage of the problem: (c) what is the reliability of the estimate? In other words, how far is the estimate likely to differ from the true value?

We know that if the sample value p differs from ϖ by $t\sqrt{(\varpi\chi/n)}$, the probability of the difference becomes smaller as t increases. Thus, with an assigned degree of probability we can say that it is improbable that p will differ from ϖ by more than an assigned amount. But to specify this amount exactly we require to know ϖ; and this is precisely the quantity we are trying to find.

The problem as put in this form can only be solved as an approximation. If n is large the standard error of p is of the order $n^{-\frac{1}{2}}$, so that we may put

$$\varpi = p + \frac{k}{n^{\frac{1}{2}}}.$$

Thus

$$\begin{aligned}
\sqrt{\frac{\varpi\chi}{n}} &= \sqrt{\left\{\frac{1}{n}\left(p+\frac{k}{n^{\frac{1}{2}}}\right)\left(q-\frac{k}{n^{\frac{1}{2}}}\right)\right\}}\\
&= \sqrt{\left\{\frac{pq}{n}\left(1+\frac{k(q-p)}{pqn^{\frac{1}{2}}}\right)\right\}},
\end{aligned}$$

neglecting terms of order n^{-1},

$$= \sqrt{\frac{pq}{n}}\left\{1+\frac{1}{2}\frac{k(q-p)}{pqn^{\frac{1}{2}}}\right\}. \tag{9.6}$$

Hence for large n the standard error of ϖ is approximately equal to $\sqrt{(pq/n)}$; and we thus reach the fundamental result that in large samples for attributes the standard error may be calculated by using the estimates of the parameters under estimate instead of the (unknown) values of those parameters themselves.

In Volume 2 we shall consider a different approach which frees us from the necessity of making this type of approximation.

Example 9.11

In a sample of 600, 240 are found to possess the attribute A. Thus $p = 0{\cdot}40$, $np = 240$, $\sqrt{(npq)} = 12$. We can therefore regard it as somewhat improbable that $n\varpi$ differs from 240 by more than twice this amount, 24, and highly improbable that it differs by more than 36. We thus can say with some assurance that $n\varpi$ lies in the range 240 ± 24 and with great assurance that it lies in the range 240 ± 36.

9.26 We now turn to a general consideration of the problems of sampling which have been exemplified above. In the first place, let us note the role of the sampling distribution in this branch of the subject. We construct from the observations some statistic t. The sampling distribution of this statistic will in general (but not always) depend on some parameters of the parent population. The probability of the observed t then permits the making of statements, by inverse probability, likelihood or otherwise, about these parameters, and so we are enabled to draw inferences about the parent population. The sampling distribution is thus fundamental to the whole subject and several subsequent chapters will be devoted entirely to the methods of finding sampling distributions when the population is specified.

If we wish to test some hypothesis about the parent which is expressible by the determination of certain parameters *a priori*, the problem is fairly simple. Given the values of the parameters, we can determine from the sampling distribution the probability of the observed value of the statistic, and use this to assess the acceptability of the hypothesis. Complications can arise even here, however, for in general, several statistics can be computed from the same sample, and they need not necessarily all lead to the same conclusion about the hypothesis; for instance, a sample might have a mean which throws doubt on the hypothesis and a variance which does not. We shall discuss this difficulty more fully in the second volume.

9.27 When the parameters of the population are not given *a priori*, we have the double problem of estimating the parameters from the sample and assigning probable limits to the estimates so obtained. We have already touched on some of the principles of estimation and shall develop the topic more systematically in due course. When we have obtained an estimate—itself a statistic—we seek its sampling distribution and therefrom can assign probable limits to the population value. A special class of cases arises when we can find a statistic whose sampling distribution depends on only one parameter of the population (as in the case of attributes).

9.28 These latter types of problem permit of certain important approximations, namely in the case when the sample is large. We saw in Chapter 7 that under very

general conditions the sum of n independent variables, distributed in whatever form, tends to normality as n tends to infinity. Now many of the ordinary statistics in current use can be expressed as the sum of variates, e.g. all the moments; and many others may also be shown to tend to normality for large samples. Thus we may approximate—

(a) by taking a statistic, calculated from the sample as if it were a population, to be the estimate of the corresponding parameter in that population, e.g. the variance of the sample may be taken as an estimate of the variance of the population;

(b) by calculating the mean and variance of the sampling distribution using, instead of the unknown parameter values, the statistic values calculated according to (a);

(c) by assuming that the distribution is normal and hence determining probabilities from the normal integral with the aid of the sampling mean and sampling variance (the latter being the square of the standard error).

9.29 Just how large n must be for such approximations to be valid it is not always easy to say. For some distributions, particularly that of the mean, quite a satisfactory approximation is given by low values of n, say $n>30$. For others n has to be much higher before the approximation begins to give satisfactory results, e.g. for the product-moment correlation coefficient even values as high as 500 are not good enough in samples from a normal population. In fact, the form of the parent distribution, as well as the statistic considered, influences the rapidity of approach of the sampling distribution to the normal form.

Example 9.12

Returning to the binomial distribution, let us consider how far our procedure need be modified when sampling without replacement.

If a random sample of n is drawn without replacement from a population of N, a proportion ϖ of which bear the attribute A, the sampling distribution of the proportion p of attribute-bearers in the sample is arrayed by the hypergeometric distribution (**5.18**) and a typical term is, in our present notation,

$$\binom{n}{np}\frac{(N\varpi)^{[np]}(N\chi)^{[nq]}}{N^{[n]}}.$$

The mean value of p in this distribution, from (5.53), is known to be ϖ; and thus the sample proportion remains an unbiassed estimator of the population proportion. It is not a Bayes estimator or a maximum likelihood estimator (see Exercise 9.10)—these do not take simple forms in this case—but it differs from them only by terms of order n^{-1}.

The variance of p, from (5.55), is given by

$$\operatorname{var}p = \frac{N-n}{N-1}\cdot\frac{\varpi\chi}{n}. \tag{9.7}$$

This differs from the ordinary binomial case (sampling with replacement) by the factor $(N-n)/(N-1)$ which is less than unity for $n>1$. The variance in sampling without replacement is therefore smaller than with replacement, as is intuitively obvious from the fact that the extreme sampling possibilities are greater in the latter case.

Example 9.13 Sequential sampling

Cases sometimes occur, especially in sampling for attributes, where the sample number is not fixed in advance. Thus, instead of drawing a sample of size n and noting the number r $(= np)$ of " successes," or appearances of the attribute, we decide to go on adding to the sample number until some (predetermined) number of successes r appear. In this case r is fixed and n is a random variable.

The distribution of n is easily obtained. The probability that there are $r-1$ successes in the first $n-1$ sample-members is $\binom{n-1}{r-1}\varpi^{r-1}\chi^{n-r}$. The probability that the next drawing yields a success is ϖ. Thus (since the sampling cannot terminate until at least r members have been drawn) the distribution of n is given by

$$\text{Prob}(n = s) = \binom{s-1}{r-1}\varpi^r\chi^{s-r}, \qquad s = r, r+1, \ldots, \infty. \tag{9.8}$$

The successive terms are

$$\varpi^r\{1, r\chi, \tfrac{1}{2}r(r+1)\chi^2, \ldots\},$$

i.e. those of the negative binomial

$$\frac{\varpi^r}{(1-\chi)^r}.$$

This is another form of the result we reached in **5.14**. The mean and variance of the distribution are, by (5.38), $r\chi/\varpi$ and $r\chi/\varpi^2$ respectively.

For the maximum likelihood estimator of ϖ we find, as for the ordinary binomial, the value p. This, however, is no longer unbiassed. In fact, we shall show that $(r-1)/(n-1)$ is an unbiassed estimator of ϖ.

For

$$\begin{aligned} E\left(\frac{r-1}{n-1}\right) &= \varpi^r \sum_{n=r}^{\infty} \frac{r-1}{n-1}\chi^{n-r}\binom{n-1}{r-1} \\ &= \varpi^r \Sigma \binom{n-2}{r-2}\chi^{n-r} \\ &= \varpi^r(1-\chi)^{-(r-1)} \\ &= \varpi. \end{aligned}$$

The biassed character of r/n follows, for it cannot have the same expectation as $(r-1)/(n-1)$, being always greater.

It is rather curious, nevertheless, that n/r is an unbiassed estimator of $1/\varpi$, as $E(n) = r/\varpi$.

At first sight this seems rather paradoxical. n/r is an unbiassed estimator of $1/\varpi$, but r/n is a biassed estimator of ϖ. The reason is, of course, our somewhat arbitrary definition of bias as a departure from a mean value. It should not occasion surprise that if $E(x) = \alpha$, $E(1/x)$ is not equal to $1/\alpha$. For a positive variate this is, in fact, inevitable (cf. Exercise 9.13).

The method of sampling under which we proceed with the drawing of the sample until a certain number of " successes " have turned up is sometimes called " Inverse Sampling," a term to which there are some objections. It is a particular case of a method

which we shall study in Volume 2, and is most conveniently known as Sequential Sampling (for attributes), the characteristic feature of which is that we decide as we go along whether to draw further members for the sample, in the light of what has been drawn up to date.

EXERCISES

9.1 Of 10,000 babies born in a particular country 5100 are male. Taking this to be a random sample of the births in that country, show that it throws considerable doubt on the hypothesis that the sexes are born in equal proportions.

Consider how far this conclusion would be modified if the sample consisted of 1000 births, 510 of which were male.

9.2 In a sample of n_1 from a population π_1, the proportion observed with an attribute A is p_1. In another independent sample of n_2 from a population π_2 the proportion is p_2. To test whether the populations have identical proportions of A in them, it is proposed to consider the difference p_1-p_2. Show that on the hypothesis of identical proportions the difference has zero mean and a variance which may be estimated as

$$(p_1n_1+p_2n_2)(q_1n_1+q_2n_2)/\{n_1n_2(n_1+n_2)\},$$

where $q_1 = 1-p_1$, $q_2 = 1-p_2$.

9.3 If a proportion ϖ has to be estimated from a simple random sample with proportion p, and if f is the prior probability of ϖ, then the posterior probability of ϖ is, according to Bayes' theorem, proportional to

$$f\varpi^{np}(1-\varpi)^{nq}.$$

Show that this is a maximum if

$$\frac{1}{f}\frac{\partial f}{\partial \varpi}+n\frac{p-\varpi}{\varpi(1-\varpi)} = 0.$$

Hence, in general, as n increases, the solution tends to $\varpi = p$, whatever the prior probability of ϖ.

9.4 By using the Bienaymé-Tchebycheff inequality, show that in sampling from a large population for an attribute, the probability that the observed proportion p in a sample of n differs from the population proportion ϖ by more than k is not greater than $1/(4nk^2)$.

(This is an exact result, no assumptions about the limiting normality of the binomial or the use of estimates in calculating standard errors being involved. The limits are, however, much too wide in general.)

9.5 In a large population the proportion ϖ of members bearing an attribute A is small. Show that in samples of size n the proportion of A's is a maximum likelihood estimator of ϖ; that it is unbiassed; and that the variance of the number, say n_1, of A's observed may be estimated as equal to n_1.

9.6 In k different populations the proportions of an attribute are $\varpi_1, \varpi_2, \ldots, \varpi_k$. One population is chosen at random (with probability $1/k$) and from it a sample of n members is drawn with replacement at random, yielding a proportion p bearing the attribute. Show that, under repetitions of this process (including the renewed selection of a parent population), the mean value of p is $\varpi\left(\equiv \sum_{i=1}^{k}\varpi_i/k\right)$ and that its variance is $\varpi(1-\varpi)/n+(n-1)\Sigma(\varpi_i-\varpi)^2/(nk)$.

Obtain the corresponding formulae if the probability of selection of the ith population is α_i, where $\sum_{i=1}^{k}\alpha_i = 1$.

9.7 In the previous exercise, a sample of kn is taken by drawing n members from each population. Show that the proportion p in the kn sample members together has mean ϖ and variance

$$\varpi(1-\varpi)/(nk) - \Sigma(\varpi_i - \varpi)^2/(nk^2).$$

9.8 From a population in which the proportion of individuals possessing a certain attribute is ϖ, a sample of size n is drawn with replacement. From this a sub-sample of n_1 is drawn with replacement. If p, p_1 are the proportions in the sample and sub-sample respectively show that $p-p_1$ has a zero mean and variance

$$\varpi(1-\varpi)\frac{n-1}{nn_1}.$$

9.9 In the previous Exercise, if the sample of size n is selected from a population of size N without replacement and the sub-sample of n_1 is selected from the n without replacement, show that $p-p_1$ has zero mean and variance

$$\varpi(1-\varpi)\frac{N}{N-1}\cdot\frac{n-n_1}{nn_1}.$$

9.10 Show that, to order N^{-1}, the maximum likelihood estimator of the proportion ϖ in the hypergeometric distribution is

$$p+\frac{1}{2N}(p-q),$$

and that, to the same order, the variance of this estimator is

$$\frac{N-n}{N-3}\cdot\frac{\varpi\chi(1-\varpi)}{n}.$$

9.11 Define a variable x as unity if the individual concerned possesses an attribute A and zero in the contrary case. If p is the proportion in a sample of n bearing the attribute show that, for sampling either with or without replacement, $E(x) = E(x^2) = E(x^k) = \varpi$. Hence obtain equation (9.7).

9.12 In Example 9.13 show that

$$E(r/n) = r\varpi^r\chi^{-r}\int_0^\chi t^{r-1}(1-t)^{-r}\,dt$$

$$= \varpi\sum_{j=0}^{\infty}\frac{r!j!}{(r+j)!}\chi^j.$$

9.13 For a variate x which is defined in the range $0 < x \leqslant \infty$, show that if $E(x)$ and $E(1/x)$ exist,

$$E(x)E(1/x) \geqslant 1,$$

the inequality becoming an equality only if the distribution of x is wholly concentrated at a single value. Hence show that if t is an unbiassed estimator of θ, $1/t$ cannot be an unbiassed estimator of $1/\theta$.

9.14 For the ordinary binomial distribution show that $E(n/r)$ does not exist.
Show also that

$$E\left(\frac{n+1}{r+1}\right) = (1-\chi^{n+1})/\varpi$$

and hence that, asymptotically in n, $(n+1)/(r+1)$ is an unbiassed estimator of $1/\varpi$.

CHAPTER 10

STANDARD ERRORS

10.1 Towards the close of the last chapter we discussed the estimation of statistical parameters from large samples and the type of judgement of their reliability which depends on the use of the standard error. It was remarked that, for large samples, an estimate of a parameter may be obtained by calculating from the sample values the value of the parameter in the sub-population composed by the sample; and it was established that for samples of size n the standard error gives a valid measure of precision, provided (a) that the sampling distribution of the statistic under discussion approaches normality and (b) that n is large in the sense there defined. It was also pointed out that a sufficiently accurate estimate of standard errors involving parent parameters could be obtained by using as the parameter values the corresponding statistics from the sample itself.

Since the majority of statistics in current use do tend to normality the theory of large samples is, in the main, devoted to the determination of standard errors. In this chapter we describe the principal methods available for the purpose, and incidentally derive formulae for the standard errors of the various statistics considered in previous chapters. To avoid the usual square roots associated with the standard error we shall write our results as sampling variances and covariances. Thus, for a statistic t we write the variance of its sampling distribution as $\operatorname{var} t$. The covariance of the joint distribution of two statistics t and u, that is, the first product-moment of their joint sampling distribution, is written $\operatorname{cov}(t, u)$.

10.2 By definition, the rth moment of a statistic t, that is the rth moment of its sampling distribution, is the mean value of t^r taken over all possible samples, and may be written $E(t^r)$ (cf. **2.27**). If the joint distribution of the variates $x_1, \ldots, x_n$, from which t is calculated, is $dF(x_1, \ldots, x_n)$, then the rth moment of t is the integral of $t^r\,dF$ (considered as a function of the x's) over the domain of the x's. In particular, if the sample is simple random and the parent distribution is dF, we have

$$E(t^r) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} t^r\, dF(x_1) \ldots dF(x_n). \tag{10.1}$$

We are particularly interested in this chapter in the first and second moments of t, that is, the mean and variance of its sampling distribution. It may be recalled that the mean value of a sum is the sum of the mean values and that, if the variables are independent, the mean value of a product is the product of the mean values (**2.28**). These two results will be repeatedly required.

Standard errors of moments

10.3 In the following sections, we shall adopt the usual convention in regard to the distinction of population values and statistics by writing Greek letters to represent

the former and Roman letters to represent the latter. We have, then, for the rth moment-statistic m'_r, corresponding to the rth moment μ'_r of the population,

$$m'_r = \frac{1}{n}\sum_{j=1}^{n} x_j^r \tag{10.2}$$

and for the mean-moment

$$m_r = \frac{1}{n}\sum_{j=1}^{n} (x_j - m'_1)^r. \tag{10.3}$$

10.4 Consider now the mean value of m'_r. We have

$$\begin{aligned} E(m'_r) &= \frac{1}{n}\Sigma E(x^r) = E(x^r) \\ &= \mu'_r. \end{aligned} \tag{10.4}$$

The sampling variance of m_r is, by definition, $E(m'_r - \mu'_r)^2$ and thus assuming, as we do throughout, that the appropriate moments exist,

$$\begin{aligned} \operatorname{var} m'_r &= E\left\{\frac{1}{n}\Sigma x^r - \mu'_r\right\}^2 \\ &= \frac{1}{n^2}E[\{\Sigma x^r\}^2 - 2n\mu'_r \Sigma x^r + n^2 \mu'^2_r] \\ &= \frac{1}{n^2}E[\{\Sigma x^r\}^2] - \mu'^2_r \\ &= \frac{1}{n^2}E\{\Sigma x_j^{2r} + \Sigma (x_j^r x_k^r)\} - \mu'^2_r \end{aligned}$$

the second summation extending over the $n(n-1)$ cases in which $j \neq k$ (permutations of j and k thus being allowed). Since the x's are independent the mean value of the product is the product of the mean values, and thus

$$\begin{aligned} \operatorname{var} m'_r &= \frac{1}{n^2}\{n\mu'_{2r} + n(n-1)\mu'^2_r\} - \mu'^2_r, \\ &= \frac{1}{n}(\mu'_{2r} - \mu'^2_r). \end{aligned} \tag{10.5}$$

This is an exact result.

In a similar way, if we have two moments, m'_q, m'_r, their sampling covariance is given by

$$\begin{aligned} \operatorname{cov}(m'_q, m'_r) &= E\{(m'_q - \mu'_q)(m'_r - \mu'_r)\} \\ &= E\left\{\left(\frac{1}{n}\Sigma x^q - \mu'_q\right)\left(\frac{1}{n}\Sigma x^r - \mu'_r\right)\right\} \\ &= \frac{1}{n^2}E(\Sigma x^{q+r}) + \frac{1}{n^2}E\sum_{j\neq k}(x_j^q x_k^r) - \frac{1}{n}\mu'_q E(\Sigma x^r) - \frac{1}{n}\mu'_r E(\Sigma x^q) + E(\mu'_q \mu'_r) \\ &= \frac{1}{n}(\mu'_{q+r} - \mu'_q \mu'_r), \end{aligned} \tag{10.6}$$

which reduces to (10.5) if $q = r$, as it must, for the covariance of a variable with itself is its variance.

10.5 The formulae for moments about the mean are not so simple, for the mean itself is subject to sampling fluctuations. We see from (10.5) with $r = 1$ that

$$\operatorname{var} m_1' = (\mu_2' - \mu_1'^2)/n = \mu_2/n. \tag{10.7}$$

Thus fluctuations in m_1' are of order $n^{-\frac{1}{2}}$. We may then take an origin at the parent mean and neglect powers of m_1' higher than the first. We then have, neglecting terms of order $n^{-\frac{1}{2}}$,

$$\begin{aligned} E(m_r) &= E\{\Sigma(x - m_1')^r/n\} \\ &= E\{\Sigma x^r - r m_1' \Sigma x^{r-1}\}/n \\ &= E\{(1 - r/n)\Sigma x^r - (r/n)\Sigma x_j x_k^{r-1}\}/n, \qquad j \neq k. \end{aligned}$$

Now the second term in the expectation on the right will involve the moments $\mu_1' \mu_{r-1}'$ and will vanish since we have chosen our origin at the mean of the parent distribution ($\mu_1' = 0$). We shall then have, neglecting terms of order $n^{-\frac{1}{2}}$,

$$E(m_r) = \mu_r, \tag{10.8}$$

a result which, unlike (10.4), is not exact but is an approximation to order $n^{-\frac{1}{2}}$. To the same order we have

$$\begin{aligned} \operatorname{var} m_r &= E(m_r^2) - \mu_r^2 \\ &= E\{\Sigma x^r - r\Sigma x_j x_k^{r-1}/n\}^2/n^2 - \mu_r^2 \\ &= \frac{1}{n^2} E\left\{\Sigma x^{2r} + \Sigma x_j^r x_k^r + \frac{r^2}{n^2}\Sigma x_j^2 x_k^{2r-2} + \frac{r^2}{n^2}\Sigma x_j^2 x_k^{r-1} x_l^{r-1} - \frac{2r}{n}\Sigma x_j^{r+1} x_k^{r-1}\right\} - \mu_r^2, \end{aligned}$$

where $j \neq k \neq l$. The expectations of other terms occurring in the squaring vanish, since they contain μ_1'. The expectation of $\Sigma x_j^2 x_k^{2r-2}$ is of order $n(n-1)$ and the term involving it is therefore of order n^{-2} and may be neglected. The others give us, to the first order in n,

$$\operatorname{var} m_r = \frac{1}{n}(\mu_{2r} - \mu_r^2 + r^2 \mu_2 \mu_{r-1}^2 - 2r\mu_{r-1}\mu_{r+1}). \tag{10.9}$$

Similarly it appears that

$$\operatorname{cov}(m_r, m_q) = \frac{1}{n}(\mu_{r+q} - \mu_r \mu_q + rq\mu_2 \mu_{r-1}\mu_{q-1} - r\mu_{r-1}\mu_{q+1} - q\mu_{r+1}\mu_{q-1}). \tag{10.10}$$

Example 10.1

For the height distribution of Table 1.7 we found (Examples 2.1 and 2.7) that $m_1' = 67{\cdot}46$, $\sqrt{m_2} = 2{\cdot}57$. Suppose we regard this distribution as a simple random sample from the adult male inhabitants of the United Kingdom living at the time when the data were collected. What can we say about the mean of the population?

The standard error of the mean depends on μ_2. This is an unknown quantity, but we may, in accordance with the general principles of large sample theory, use m_2 instead. We then find

$$\text{Standard error of } m_1' = \frac{2{\cdot}57}{\sqrt{8585}} = 0{\cdot}028 \text{ approximately.}$$

Thus we can say that the population mean probably lies in the range of twice this

amount on either side of the sample mean, i.e. in the range $67{\cdot}46 \pm 0{\cdot}056$, and very probably in thrice the range, i.e. $67{\cdot}46 \pm 0{\cdot}084$. Our estimate of the mean would almost certainly be less than a tenth of an inch in error.

Example 10.2

From equation (10.9) with $r = 4$ we find

$$\operatorname{var} m_4 = \frac{1}{n}(\mu_8 - \mu_4^2 - 8\mu_5\mu_3 + 16\mu_2\mu_3^2).$$

In Chapter 12 we shall show how to obtain this result by other methods of a more exact character and confirm that it is, in fact, correct to order n^{-1}.

Example 10.3

To show that in samples from a symmetrical population the first product-moment between the mean and any mean-moment of even order vanishes to order n^{-1}.

We have, by definition, as in **10.5**,

$$\begin{aligned}\operatorname{cov}(m_1', m_r) &= \frac{1}{n^2} E\left[\left\{\Sigma x \left\{\Sigma x^r - \frac{r}{n}\Sigma x_j x_k^{r-1}\right\}\right\}\right] \\ &= \frac{1}{n^2} E\left\{\Sigma x^{r+1} - \frac{r}{n}\Sigma x_j^2 x_k^{r-1}\right\},\end{aligned}$$

the other terms vanishing, since they involve the unit power of x, if we take an origin at the mean of the parent population,

$$= \frac{1}{n}(\mu_{r+1} - r\mu_2\mu_{r-1}).$$

Now if r is even, μ_{r+1} and μ_{r-1}, being moments of odd order, will vanish for a symmetrical population and hence

$$\operatorname{cov}(m_1', m_r) = 0.$$

In the language of the theory of correlation, the mean and the even moment about the mean are uncorrelated to order n^{-1}.

Standard errors of functions of random variables

10.6 If $\xi_1, \xi_2, \ldots, \xi_k$ are random variables with corresponding variates $x_1, x_2, \ldots, x_k$, we may define a function $g(\xi_1, \xi_2, \ldots, \xi_k)$ as that function which takes the value $g(x_1, x_2, \ldots, x_k)$ when $\xi_1 = x_1, \xi_2 = x_2, \ldots, \xi_k = x_k$.

Let us suppose that x_i has a mean θ_i and a variance of order n^{-1} and that g has finite mean and variance. If g is differentiable at $x_i = \theta_i$ we have, writing $\partial g/\partial\theta_i$ for $\partial g/\partial x_i$ at $x_i = \theta_i$,

$$\Delta g = \sum_{i=1}^{k} \frac{\partial g}{\partial \theta_i}\Delta x_i + O(\Delta x)^2, \tag{10.11}$$

where we measure the x_i from their expected values.

For large samples the expectation of $(\Delta x)^2$ is $o(1)$, and we have, to the first order, since $E(\Delta x_i) = 0$,

$$E(\Delta g) = 0.$$

Hence the mean value of g is $g(\theta_1, \theta_2, \ldots, \theta_k)$. We also have, to the first order, if the derivatives are not equal to zero,

$$\begin{aligned} \operatorname{var} g &= E\left(\Sigma \frac{\partial g}{\partial \theta_i} \Delta x_i\right)^2 \\ &= \sum_{i,j=1}^{k} \left\{ \frac{\partial g}{\partial \theta_i} \frac{\partial g}{\partial \theta_j} \operatorname{cov}(x_i, x_j) \right\} \end{aligned} \tag{10.12}$$

where now i, j may be equal (in which case $\operatorname{cov}(x_i, x_j) = \operatorname{var} x_i$).

This is a formula of fundamental importance. It gives us the variance of a function of random variables in terms of their variances and covariances.

In a similar manner it may be shown that for two functions of random variables, say g and h,

$$\operatorname{cov}(g, h) = \sum_{i,j=1}^{k} \left\{ \frac{\partial g}{\partial \theta_i} \frac{\partial h}{\partial \theta_j} \operatorname{cov}(x_i, x_j) \right\}. \tag{10.13}$$

10.7 Three particular specializations of (10.12) are of interest :—

(a) g is a function of a single random variable. We then find

$$\operatorname{var} g(x) = \left(\frac{dg}{d\theta}\right)^2 \operatorname{var} x. \tag{10.14}$$

In particular, if $g(x)$ is simply a linear function of the form $cx+e$ we find that $\operatorname{var} g = c^2 \operatorname{var} x$, as is otherwise obvious, the result in this case being exact.

(b) g is a linear function of the random variables. If

$$g(x_1, \ldots, x_k) = \sum_{i=1}^{k} a_i x_i \tag{10.15}$$

then (10.12) gives

$$\operatorname{var} g = \Sigma a_i^2 \operatorname{var} x_i + \sum_{i \neq j} a_i a_j \operatorname{cov}(x_i, x_j), \tag{10.16}$$

again an exact result. In particular, if the x's are independent, the variance of g is a weighted sum of their individual variances.

(c) g is a ratio of two random variables, say x_1/x_2. To avoid difficulties to be discussed in the next chapter, we assume here that $x_2 > 0$ if it is discrete and $\geqslant 0$ if it is continuous. Equation (10.12) gives

$$\begin{aligned} \operatorname{var}(x_1/x_2) &= \frac{\operatorname{var} x_1}{\theta_2^2} + \frac{\theta_1^2 \operatorname{var} x_2}{\theta_2^4} - \frac{2\theta_1 \operatorname{cov}(x_1, x_2)}{\theta_2^3} \\ &= \left\{\frac{E(x_1)}{E(x_2)}\right\}^2 \left\{ \frac{\operatorname{var} x_1}{E^2(x_1)} + \frac{\operatorname{var} x_2}{E^2(x_2)} - \frac{2 \operatorname{cov}(x_1, x_2)}{E(x_1) E(x_2)} \right\}. \end{aligned} \tag{10.17}$$

The second factor in (10.17) is the sum of the squares of the coefficients of variation of the two variables, minus twice the square of what may analogously be called their coefficient of covariation.

Example 10.4

To find the sampling variance of the fourth cumulant, we shall, for reasons which will become clear in Chapter 12, continue to write κ, not k, for the *sample* cumulant. Thus we require the variance of

$$\kappa_4 = m_4 - 3m_2^2.$$

From (10.12) with $x_1 = m_4$, $x_2 = m_2$ and $g = x_1 - 3x_2^2$, we obtain

$$\operatorname{var}\kappa_4 = \operatorname{var} m_4 + 36\mu_2^2 \operatorname{var} m_2 - 12\mu_2 \operatorname{cov}(m_4, m_2).$$

Substituting from (10.9) and (10.10) we find

$$\operatorname{var}\kappa_4 = \frac{1}{n}\{\mu_8 - 12\mu_6\mu_2 - 8\mu_5\mu_3 - \mu_4^2 + 48\mu_4\mu_2^2 + 64\mu_3^2\mu_2 - 36\mu_2^4\}.$$

For a normal population, $\mu_4 = 3\mu_2^2 = 3\sigma^4$, $\mu_6 = 15\sigma^6$, $\mu_8 = 105\sigma^8$ and we find in this case

$$\operatorname{var}\kappa_4 = 24\sigma^8/n.$$

Example 10.5

To find the sampling variance of the coefficient of variation

$$V = 100\sqrt{m_2}/m_1'.$$

Using (10.17), we have, writing $\mathbf{V}$ for the population coefficient (taken to be non-zero, as is the population mean μ_1'),

$$\operatorname{var}(V/100) = \left(\frac{\mathbf{V}}{100}\right)^2 \left\{\frac{\operatorname{var}(\sqrt{m_2})}{E^2(\sqrt{m_2})} + \frac{\operatorname{var} m_1'}{E^2(m_1')} - \frac{2\operatorname{cov}(\sqrt{m_2}, m_1')}{E(\sqrt{m_2})E(m_1')}\right\}. \tag{10.18}$$

We saw in (10.7) that

$$\operatorname{var} m_1 = \mu_2/n.$$

Also, from (10.14) and (10.9),

$$\operatorname{var}(\sqrt{m_2}) = \frac{\operatorname{var} m_2}{4\mu_2} = \frac{\mu_4 - \mu_2^2}{4n\mu_2}.$$

Finally, in Example 10.3 we saw that to order n^{-1}

$$\operatorname{cov}(m_2, m_1') = \mu_3/n,$$

and from (10.13) we have

$$\operatorname{cov}(\sqrt{m_2}, m_1') = \frac{1}{2\sqrt{\mu_2}} \operatorname{cov}(m_2, m_1') = \mu_3/(2n\sqrt{\mu_2}).$$

Substituting these various values in (10.18) we find

$$\operatorname{var} V = \frac{\mathbf{V}^2}{n}\left\{\frac{\mu_4 - \mu_2^2}{4\mu_2^2} + \frac{\mu_2}{\mu_1'^2} - \frac{\mu_3}{\mu_2\mu_1'}\right\}. \tag{10.19}$$

In the normal case ($\mu_3 = 0$, $\mu_4 = 3\mu_2^2$) this becomes

$$\operatorname{var} V = \frac{\mathbf{V}^2}{n}\left\{\frac{1}{2} + \frac{\mu_2}{\mu_1'^2}\right\} = \frac{\mathbf{V}^2}{2n}\left\{1 + 2\left(\frac{\mathbf{V}}{100}\right)^2\right\}$$

$$= \frac{\mathbf{V}^2}{2n} \text{ approximately.} \tag{10.20}$$

10.8 A few points on the use of standard errors may be noted.

(a) The standard error, strictly speaking, has been justified only in the case where the statistic tends to normality. In other cases it may be used, in conjunction with the Bienaymé-Tchebycheff inequality, to make statements in probability, but the more exact statements which are customarily made rely on the normal integral.

(b) Where there is serious doubt, it may be necessary to conduct a separate inquiry into the limiting normality of statistics. Many of those in current use, especially those dependent on sums of variates, like the moments, tend to normality under a Central Limit effect; but certain others, especially those dependent on extreme values, do not.

(c) On a similar point, it should also be remembered that some statistics tend to normality more rapidly than others, and a given n may be large for some purposes but not for others. So far as it is possible to generalize with safety, we can usually (but not always) assume values of n greater than 500 to be "large"; values greater than 100 are often great enough to be "large" for our purposes; values below 100 are suspect in many instances; and values below 30 are very rarely "large." Some measures calculated from the moments tend to normality very slowly. $\sqrt{b_1}$ or b_1 (the sample values of $\sqrt{\beta_1}$ or β_1) are cases in point, and more refined methods which we discuss in Chapter 12 are preferable to the use of the standard error.

(d) The sampling variances must relate to the statistic under consideration. For instance, if the standard deviation of a normal curve is estimated by taking $\sqrt{(\frac{1}{2}\pi)}$ times the mean deviation of the sample, instead of the more usual root-mean-square, the formula $\operatorname{var} s = (\mu_4 - \mu_2^2)/(4n\mu_2)$ derivable from (10.9) is not applicable.

(e) From (10.5) and (10.9) it will be seen that the sampling variance of a moment depends on the population moment of twice the order, i.e. becomes very large for higher moments, even when n is large. This is the reason why such moments have very limited practical application.

(f) The order of the approximation makes it necessary to exercise care in the neighbourhood of vanishing values of standard errors. For instance, if the coefficient of variation $V = 0$ in a sample, the formula of Example 10.5 would be estimated from the sample as $\operatorname{var} V = 0$. But it does not, of course, follow that there is no variation at all in the population, though none exists in the sample and the presence of variation in the parent will be unlikely if the sample is at all large.

(g) It is interesting to compare the sampling fluctuations, as expressed in the sampling variance, with Sheppard's corrections to the moments. Writing temporarily s_1^2 for the uncorrected variance in the sample, s_2^2 for the corrected variance, we have

$$\frac{s_2^2}{s_1^2} = 1 - \frac{1}{12}\frac{h^2}{s_1^2},$$

where h is the interval width. For many practical cases, if d is the number of intervals, dh is about equal to $6s_1$, and thus

$$\frac{s_2^2}{s_1^2} = 1 - \frac{3}{d^2}$$

$$\frac{s_2}{s_1} = 1 - \frac{3}{2d^2} \text{ approximately.}$$

Writing s for the square root of the sample variance m_2 we have

$$\operatorname{var} s = (\mu_4 - \mu_2^2)/(4n\mu_2).$$

For a normal population this is equal to $\sigma^2/(2n)$.

Thus if n is, say, 1000, the standard error of s is about $0{\cdot}0224\sigma = 2{\cdot}24$ per cent. of σ. Sheppard's correction in a case where $d = 20$ is only 0·375 per cent. of s_1, i.e. only about a sixth of the standard error. It is as well to make the corrections, even when n is smaller than 1000, in order to avoid systematic error; but the correction should not be misinterpreted as implying a higher degree of reliability in the corrected value than actually exists.

Similar considerations apply *a fortiori* to the higher moments.

Standard errors of bivariate moments

10.9 Extensions of the above formulae to the bivariate case are made without difficulty, only slightly more complicated algebra being involved. The reader will be able to verify the following formulae for himself:

$$\operatorname{var}(m'_{r,s}) = \frac{1}{n}(\mu'_{2r,2s} - \mu'^2_{r,s}) \tag{10.21}$$

$$\operatorname{cov}(m'_{r,s}, m'_{u,v}) = \frac{1}{n}(\mu'_{r+u,s+v} - \mu'_{r,s}\mu'_{u,v}) \tag{10.22}$$

$$\begin{aligned}\operatorname{var}(m_{r,s}) = \frac{1}{n}(&\mu_{2r,2s} - \mu^2_{r,s} + r^2\mu_{2,0}\mu^2_{r-1,s} + s^2\mu_{0,2}\mu^2_{r,s-1} \\ &+ 2rs\mu_{1,1}\mu_{r-1,s}\mu_{r,s-1} - 2r\mu_{r+1,s}\mu_{r-1,s} - 2s\mu_{r,s+1}\mu_{r,s-1})\end{aligned} \tag{10.23}$$

$$\begin{aligned}\operatorname{cov}(m_{r,s}, m_{u,v}) = \frac{1}{n}(&\mu_{r+u,s+v} - \mu_{r,s}\mu_{u,v} + ru\mu_{2,0}\mu_{r-1,s}\mu_{u-1,v} \\ &+ sv\mu_{0,2}\mu_{r,s-1}\mu_{u,v-1} + rv\mu_{1,1}\mu_{r-1,s}\mu_{u,v-1} \\ &+ su\mu_{1,1}\mu_{r,s-1}\mu_{u-1,v} - u\mu_{r+1,s}\mu_{u-1,v} \\ &- v\mu_{r,s+1}\mu_{u,v-1} - r\mu_{r-1,s}\mu_{u+1,v} - s\mu_{r,s-1}\mu_{u,v+1}).\end{aligned} \tag{10.24}$$

Where no ambiguity is involved we may write $\mu_{r,s}$ as μ_{rs}.

Example 10.6

The coefficient of correlation in a sample is defined by $r = m_{11}/(m_{20}m_{02})^{\frac{1}{2}}$ and in the population by $\rho = \mu_{11}/(\mu_{20}\mu_{02})^{\frac{1}{2}}$. From (10.17), we have

$$\operatorname{var} r = \rho^2\left\{\frac{\operatorname{var} m_{11}}{\mu_{11}^2} + \frac{\operatorname{var}(m_{20}m_{02})^{\frac{1}{2}}}{\mu_{20}\mu_{02}} - 2\,\frac{\operatorname{cov}\{m_{11}, (m_{20}m_{02})^{\frac{1}{2}}\}}{\mu_{11}(\mu_{20}\mu_{02})^{\frac{1}{2}}}\right\}$$

and evaluating $\operatorname{var}(m_{20}m_{02})^{\frac{1}{2}}$ and $\operatorname{cov}\{m_{11}(m_{20}m_{02})^{\frac{1}{2}}\}$ from (10.12) and (10.13), we find

$$\begin{aligned}\operatorname{var} r = \rho^2\Bigg\{&\frac{\operatorname{var} m_{11}}{\mu_{11}^2} + \frac{1}{4}\left(\frac{\operatorname{var} m_{20}}{\mu_{20}^2} + \frac{\operatorname{var} m_{02}}{\mu_{02}^2} + \frac{2\operatorname{cov}(m_{20}m_{02})}{\mu_{20}\mu_{02}}\right) \\ &- \left(\frac{\operatorname{cov}(m_{11}m_{02})}{\mu_{11}\mu_{02}} + \frac{\operatorname{cov}(m_{11}m_{20})}{\mu_{11}\mu_{20}}\right)\Bigg\}\end{aligned}$$

from which, on substituting for variances and covariances from (10.23) and (10.24), we have

$$\operatorname{var} r = \frac{\rho^2}{n}\left\{\frac{\mu_{22}}{\mu_{11}^2}+\frac{1}{4}\left(\frac{\mu_{40}}{\mu_{20}^2}+\frac{\mu_{04}}{\mu_{02}^2}+\frac{2\mu_{22}}{\mu_{20}\mu_{02}}\right)-\left(\frac{\mu_{31}}{\mu_{11}\mu_{20}}+\frac{\mu_{13}}{\mu_{11}\mu_{02}}\right)\right\}.$$

For the bivariate normal distribution the substitution of values of Example 3.17 gives

$$\operatorname{var} r = \frac{1}{n}(1-\rho^2)^2.$$

The use of the standard error to test a hypothetical non-zero value of ρ is not, however, to be recommended, since the sampling distribution of r tends to normality very slowly. We return to this point in **16.23–33** below.

Standard errors of quantiles

10.10 Among the various quantities measuring location and dispersion which we considered in Chapter 2 there was one group, namely the quantiles, which are not algebraic functions of the observations and whose sampling variances cannot accordingly be determined by the above methods. We proceed to consider them now.

Suppose the parent distribution is represented by $dF(x) = f(x)\,dx$. The probability that, of a sample of n, $(l-1)$ fall below a value x_1, one falls in the range $x_1 \pm \frac{1}{2}dx_1$ and the remaining $(n-l)$ fall above x_1 is proportional to

$$\{F(x_1)\}^{l-1} f(x_1)\,dx_1\,\{1-F(x_1)\}^{n-l} = F_1^{l-1}(1-F_1)^{n-l}\,dF_1, \tag{10.25}$$

where $F_1 = F(x_1)$. This expression is accordingly the frequency function of the value not exceeded by a proportion l/n of the sample, i.e. the lth quantile.

Put $l = nq$

so that $n-l = n(1-q)$

$= np$, say.

The distribution (10.25) has a modal value given by differentiating the frequency function with respect to x_1, i.e. (taking logarithms first) by

$$(l-1)\frac{f_1}{F_1}-(n-l)\frac{f_1}{(1-F_1)}+\frac{f_1'}{f_1} = 0, \tag{10.26}$$

this equation being satisfied by the modal value $\check{x}$. Now for large n, the factor f_1'/f_1 will in general be small compared with the other terms in (10.26), l and $n-l$ being large. We may therefore neglect it, and (10.26) becomes, to order n^{-1},

$$\frac{q}{F}-\frac{p}{1-F} = 0$$

or

$$F(\check{x}) = q.$$

This is in accordance with our general assumptions. To order n^{-1} the quantile of the sample equals the quantile of the parent.

Now let us investigate the distribution (10.25) in the neighbourhood of the modal value. Put

$$F_1 = q+\xi.$$

(10.25) becomes (neglecting constants)

$$(q+\xi)^{nq}(p-\xi)^{np}.$$

Taking logarithms and expanding we have, except for constants,

$$\begin{aligned} & nq\log\left(1+\frac{\xi}{q}\right)+np\log\left(1-\frac{\xi}{p}\right) \\ &= nq\left(+\frac{\xi}{q}-\frac{1}{2}\frac{\xi^2}{q^2}\cdots\right)+np\left(-\frac{\xi}{p}-\frac{1}{2}\frac{\xi^2}{p^2}\cdots\right) \\ &= -\frac{n\xi^2}{2pq}+O(\xi^3). \end{aligned}$$

Now for large samples ξ will be small compared with q, and we neglect the terms of higher order. Thus the distribution of ξ is

$$dF \propto \exp\left(\frac{-n\xi^2}{2pq}\right)d\xi$$

or, evaluating the necessary constant by integration,

$$dF = \frac{1}{\sqrt{(2\pi)}\sqrt{(pq/n)}}\exp\left(\frac{-n\xi^2}{2pq}\right)d\xi, \tag{10.27}$$

showing that ξ is in the limit distributed normally with variance

$$\operatorname{var}\xi = \frac{pq}{n}. \tag{10.28}$$

This is the variance of ξ, which is a proportion. To find the variance of x_1 we note that $d\xi = dF_1 = f_1\,dx_1$ and hence that

$$\operatorname{var}x_1 = \frac{pq}{nf_1^2}. \tag{10.29}$$

In practice this formula is often applied to grouped frequency-distributions, and in such applications it is to be remembered that f_1, the ordinate of the parent, is to be taken as the frequency *per unit interval* at x_1, this being the best estimate of the ordinate.

Example 10.7

If x_1 is the median, $p = q = \frac{1}{2}$ and we have var(median) $= 1/(4nf_1^2)$, where f_1 is the median ordinate. For instance, if the parent population is normal with variance σ^2, the median ordinate is (from Appendix Table 1) $0{\cdot}39894/\sigma$. Hence the standard error of the median is

$$\begin{aligned} & \frac{\sigma}{\sqrt{n}}\cdot\frac{1}{2\times 0{\cdot}39894} \\ &= 1{\cdot}2533\frac{\sigma}{\sqrt{n}}. \end{aligned}$$

The standard error of the mean in samples of n from a normal population is $\frac{\sigma}{\sqrt{n}}$, which is thus considerably smaller than the standard error of the median.

10.11 To find the covariance of two quantiles we generalize equation (10.25). If we have a random sample of n individuals the probability that $(l-1)$ lie below x_1, one lies at $x_1 \pm \frac{1}{2}dx_1$, $(n-l-m)$ lie between x_1 and x_2, one at $x_2 \pm \frac{1}{2}dx_2$, and the remaining $(m-1)$ above x_2 is

$$dF \propto F_1^{l-1}(F_2-F_1)^{n-l-m}(1-F_2)^{m-1}\,dF_1\,dF_2, \tag{10.30}$$

where $F_1 = F(x_1)$, $F_2 = F(x_2)$.

We put

$$\begin{aligned} l &= q_1 n \\ m &= p_2 n \end{aligned}$$

and find, for the equations giving the modal values corresponding to (10.26),

$$\frac{q_1}{F_1} - \frac{(q_2-q_1)}{F_2-F_1} = 0$$

$$\frac{q_2-q_1}{F_2-F_1} - \frac{p_2}{1-F_2} = 0$$

giving, for the limiting modal values,

$$\left.\begin{aligned} F(\check{x}_1) &= q_1 \\ F(\check{x}_2) &= q_2 \end{aligned}\right\} \tag{10.31}$$

The conditions as to the relative smallness of $f'(\check{x})/f(\check{x})$ are satisfied in any ordinary case. Now put

$$\begin{aligned} F_1 &= q_1+\xi_1 \\ F_2 &= q_2+\xi_2. \end{aligned}$$

The joint distribution of ξ_1 and ξ_2 then becomes

$$dF \propto (q_1+\xi_1)^{q_1 n}(q_2-q_1+\xi_2-\xi_1)^{(q_2-q_1)n}(p_2-\xi_2)^{p_2 n}\,d\xi_1\,d\xi_2.$$

On proceeding as in the previous section, taking logarithms, expanding and neglecting terms in ξ^3 and higher, we find ultimately

$$dF \propto \exp\left\{-\frac{n}{2(q_2-q_1)}\left(\frac{q_2}{q_1}\xi_1^2 - 2\xi_1\xi_2 + \frac{p_1}{p_2}\xi_2^2\right)\right\} d\xi_1\,d\xi_2. \tag{10.32}$$

Thus the joint distribution of ξ_1 and ξ_2 tends to the bivariate normal form, and on comparing (10.32) with the standard form of Example 3.17, we see that

$$\begin{aligned} \frac{1}{(1-\rho^2)\operatorname{var}\xi_1} &= \frac{nq_2}{(q_2-q_1)q_1} \\ \frac{1}{(1-\rho^2)\operatorname{var}\xi_2} &= \frac{np_1}{(q_2-q_1)p_2} \\ \frac{\rho^2}{(1-\rho^2)\operatorname{cov}(\xi_1,\xi_2)} &= \frac{\rho}{(1-\rho^2)\sqrt{\{\operatorname{var}\xi_1 \operatorname{var}\xi_2\}}} = \frac{n}{(q_2-q_1)}, \end{aligned}$$

whence it is easy to find

$$\left.\begin{aligned} \operatorname{var}\xi_1 &= \frac{p_1 q_1}{n} \\ \operatorname{var}\xi_2 &= \frac{p_2 q_2}{n} \\ \operatorname{cov}(\xi_1,\xi_2) &= \frac{p_2 q_1}{n} \end{aligned}\right\} \tag{10.33}$$

The asymmetry of the result for the covariance is due to the fact that p_2 relates necessarily to the *upper* quantile. For the corresponding expression in x_1 and x_2 we have

$$\operatorname{cov}(x_1, x_2) = \frac{p_2 q_1}{n f_2 f_1}. \tag{10.34}$$

With equations (10.33) and (10.34) we can find expressions for the variances of the interquartile range and similar statistics.

Example 10.8

The variance of the difference δ of two quantiles at x_1 and x_2 is given, from (10.16), by

$$\begin{aligned} d\delta &= dx_1 - dx_2, \\ \operatorname{var}\delta &= \operatorname{var}x_1 + \operatorname{var}x_2 - 2\operatorname{cov}(x_1, x_2) \\ &= \frac{1}{n}\left\{\frac{p_1 q_1}{f_1^2} + \frac{p_2 q_2}{f_2^2} - \frac{2p_2 q_1}{f_1 f_2}\right\}. \end{aligned}$$

When the quantiles are the two quartiles, $p_2 = q_1 = \frac{1}{4}$, $p_1 = q_2 = \frac{3}{4}$, and for the variance of the *semi*-interquartile range we have

$$\operatorname{var}(\text{s.i.q.}) = \frac{1}{64n}\left(\frac{3}{f_1^2} + \frac{3}{f_2^2} - \frac{2}{f_1 f_2}\right),$$

where f_1, f_2 are the frequencies per unit interval at the two quartiles, f_2 relating to the upper quartile.

For instance, if the parent distribution is normal, $f_1 = f_2$ and we find

$$\operatorname{var}(\text{s.i.q.}) = \frac{1}{16nf_1^2}.$$

From the tables the deviate corresponding to the quartile is 0·6745 and the ordinate at this point, $f_1 = 0{\cdot}3178/\sigma$, so that the standard error of the semi-interquartile range is

$$\begin{aligned} &\frac{\sigma}{\sqrt{n}(4 \times 0{\cdot}3178)} \\ &= 0{\cdot}7867\frac{\sigma}{\sqrt{n}}. \end{aligned}$$

10.12 In amplification of the point mentioned in **10.8** (d), it is worth while stressing again the fact that a standard error is related to the way in which a parameter is estimated. For instance, the standard deviation of a normal distribution can be estimated from a sample in several ways: from the second moment; by taking $\sqrt{(\frac{1}{2}\pi)}$ times the mean deviation; by taking 1/0·6745 times the semi-interquartile range; and so on. Each method will have its appropriate standard error, that for the first (**10.8**) being $\sigma/\sqrt{(2n)}$, and that for the third $1{\cdot}6495\sigma/\sqrt{(2n)}$, as is easily deduced from Example 10.8. At a later stage considerations such as this will lead us to the inquiry, what is the estimate, if any, with the *minimum* sampling variance? For present purposes it is enough to note the importance of not using a quoted formula without reference to the method of estimation of the parameter concerned.

Standard error of the mean deviation

10.13 Suppose that we are sampling from a population with mean μ and variance σ^2. Let d_α be the sample mean deviation measured about the *fixed* point α, i.e.

$$d_\alpha = \frac{1}{n}\sum_{i=1}^{n} |x_i - \alpha|.$$

Then

$$E(d_\alpha) = E|x_i - \alpha| = \delta_\alpha, \tag{10.35}$$

where δ_α is the corresponding population mean deviation. Also, since the n sample observations are independent,

$$\begin{aligned} \operatorname{var} d_\alpha &= \frac{1}{n}\operatorname{var}(|x-\alpha|) \\ &= \frac{1}{n}\{E(|x-\alpha|^2) - E^2(|x-\alpha|)\} \\ &= \frac{1}{n}\{\sigma^2 + (\alpha-\mu)^2 - \delta_\alpha^2\}. \end{aligned} \tag{10.36}$$

If $\alpha = \mu$, the parent mean, this becomes

$$\operatorname{var} d_\mu = \frac{1}{n}(\sigma^2 - \delta_\mu^2). \tag{10.37}$$

Consider now the sample mean deviation d_a about a value a, itself subject to sampling variation, which has the properties

$$\begin{aligned} E(a) &= \alpha \\ \operatorname{var} a &= O(n^{-1}). \end{aligned}$$

By an argument of the kind used in **10.5**, it follows that *approximately*

$$\begin{aligned} E(d_a) &= \delta_\alpha \\ \operatorname{var} d_a &= \operatorname{var} d_\alpha. \end{aligned}$$

In particular, if $a = \bar{x}$ we have for the mean deviation from the sample mean

$$\operatorname{var} d = \frac{1}{n}(\sigma^2 - \delta_\mu^2). \tag{10.38}$$

For a normal parent we have seen at (5.75) that

$$\delta_\mu = \sigma\sqrt{(2/\pi)}$$

and hence, in this case,

$$\operatorname{var} d = \sigma^2(1 - 2/\pi)/n. \tag{10.39}$$

In the normal case the sampling variance is known exactly, Helmert (1876) and Fisher (1920) independently establishing the result

$$\operatorname{var} d = \frac{2\sigma^2(n-1)}{\pi n^2}[\tfrac{1}{2}\pi + \sqrt{\{n(n-1)\}} - n - \arcsin\{1/(n-1)\}]. \tag{10.40}$$

For large n, this expression tends to (10.39). Godwin (1945) gives an exact expression for the d.f. of d, which is tabulated for $n = 2(1)10$ by Hartley (1945).

Standard error of the mean difference

10.14 Let us write g for the sample value of Gini's mean difference without repetition, for which the parental value is Δ (**2.21**). We may then write

$$g = \frac{1}{n(n-1)} \sum_{i,j=1}^{n} |x_i - x_j|. \tag{10.41}$$

We have at once

$$E(g) = E|x_i - x_j| = \Delta. \tag{10.42}$$

Also

$$E(g^2) = \frac{1}{n^2(n-1)^2} E\{\Sigma |x_i - x_j|\}^2$$

$$= \frac{1}{n^2(n-1)^2} E\{\Sigma (x_i - x_j)^2 + \Sigma' |x_i - x_j||x_i - x_k| + \Sigma'' |x_i - x_j||x_k - x_l|\},$$

$$i \neq j \neq k \neq l,$$

where there are $2n(n-1)$ terms in the summation Σ, $4n(n-1)(n-2)$ in Σ' and $n(n-1)(n-2)(n-3)$ in Σ''. We then find

$$E(g^2) = \frac{1}{n(n-1)} \{2E(x_i - x_j)^2 + 4(n-2)E(|x_i - x_j||x_i - x_k|)$$

$$+ (n-2)(n-3)E(|x_i - x_j||x_k - x_l|)\}.$$

The first expected value is $2\sigma^2$ and the third is Δ^2. Writing $\mathcal{J}$ for the second we have

$$E(g^2) = \frac{1}{n(n-1)} \{4\sigma^2 + 4(n-2)\mathcal{J} + (n-2)(n-3)\Delta^2\}$$

and hence

$$\text{var}g = \frac{1}{n(n-1)} \{4\sigma^2 + 4(n-2)\mathcal{J} - 2(2n-3)\Delta^2\}. \tag{10.43}$$

The quantity $\mathcal{J}$ may be put in various forms. If $f(x)$ is the population frequency function we may write it as

$$\mathcal{J} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |x-y||x-z| f(x)f(y)f(z)\,dx\,dy\,dz. \tag{10.44}$$

If Z is the incomplete moment function defined by

$$Z(x) = \int_{-\infty}^{x} tf(t)\,dt$$

(10.44) may be reduced—we leave the reduction to the reader as an exercise—to

$$\mathcal{J} = 4\int_{-\infty}^{\infty} \{(xF - Z)^2 + (\mu - x)(xF - Z)\} f(x)\,dx. \tag{10.45}$$

This derivation is due to Lomnicki (1952), a more complicated proof having been first given by Nair (1936), who worked out the variance in three cases :—

Normal population : any mean, variance σ^2.

$$E(g) = 2\sigma/\sqrt{\pi}$$

$$\text{var}g = \frac{4\sigma^2}{n(n-1)} \{\tfrac{1}{3}(n+1) + 2(n-2)\sqrt{3}/\pi - 2(2n-3)/\pi\} \sim \sigma^2(0{\cdot}8068)^2/n. \tag{10.46}$$

Exponential population : $dF = \exp(-x/\sigma)\,dx/\sigma. \quad (0 \leqslant x \leqslant \infty)$

$$E(g) = \sigma$$

$$\text{var}\, g = \sigma^2 \frac{2(2n-1)}{3n(n-1)} \sim 4\sigma^2/(3n). \tag{10.47}$$

Rectangular population : $dF = dx/k. \quad (0 \leqslant x \leqslant k)$

$$E(g) = \tfrac{1}{2}k$$

$$\text{var}\, g = \frac{k^2(n+3)}{45n(n-1)} \sim k^2/(45n). \tag{10.48}$$

Some commonly required standard errors

10.15 It may be useful to list here for reference (see table opposite) some of the standard errors commonly required. Certain of the results have been obtained already in this chapter ; others are direct consequences of the methods we have developed ; some relate to statistics which we shall meet at a later stage.

For convenience in writing, the variance is usually given, not the standard error ; but sometimes the numerical coefficient of a variance is expressed as a square so that the square root is obtainable at sight.

EXERCISES

10.1 In the height distribution of Table 1.7 it has been found that

$$m_2 = \ \ 6{\cdot}616$$
$$m_3 = -0{\cdot}207$$
$$m_4 = 137{\cdot}689.$$

Regarding the distribution as a random sample from a population which is approximately normal, show that m_3 does not differ significantly from zero (which, of course, must be so if the assumption of normality is to be maintained) and that m_4 has a standard error of about 4 per cent. of its value.

10.2 A multinomial distribution has k classes with probabilities $\varpi_1, \varpi_2, \ldots, \varpi_k$. In a sample of size n the observed frequencies in these classes are $f_1, f_2, \ldots, f_k$. Show that for any function T of these frequencies

$$\frac{1}{n}\text{var}\, T = \sum_{i=1}^{k} \varpi_i \left(\frac{\partial T}{\partial f_i}\right)^2 - \left(\frac{\partial T}{\partial n}\right)^2$$

asymptotically, the derivatives being taken at the values $f_i = n\varpi_i$.

(Fisher, *Statistical Methods for Research Workers*)

10.3 With a random sample of 100 it is found that the odds are 99 to 1 against a certain statistic differing from the parent value by more than two units. Find how big the sample would have to be to increase these odds to 199 to 1 ; and find the odds that with the sample of 100 the statistic would differ from the parent value by one unit.

10.4 If k random samples of different sizes are drawn from a population in which the proportion of members bearing an attribute A is $\bar{\omega}$, show that the standard error of the mean proportion of A in the samples is $\sqrt{\dfrac{\bar{\omega}(1-\bar{\omega})}{kH}}$ where H is the harmonic mean of the numbers in the samples.

Statistic	Variance multiplied by n	Notes
Mean, m_1'	$\mu_2\ (=\sigma^2)$	True for any population with finite second moment μ_2.
Sample variance, m_2	$\mu_4-\mu_2^2$	For normal parent, $=2\sigma^4$.
Sample s.d., s	$(\mu_4-\mu_2^2)/(4\mu_2)$	For normal parent, $=\sigma^2/2$.
Third moment, m_3	$\mu_6-\mu_3^2-6\mu_4\,\mu_2+9\mu_2^3$	For normal parent, $=6\sigma^6$.
Fourth moment, m_4	$\mu_8-\mu_4^2-8\mu_5\,\mu_3+16\mu_2\mu_3^2$	For normal parent, $=96\sigma^8$.
b_1	$\beta_1\,(4\beta_4-24\beta_2+36+9\beta_1\,\beta_2-12\beta_3+35\beta_1)$*	For normal parent, variance is $6/n$. Note failure of usual formula for symmetrical distributions with $\beta_1=0$. Cf. **10.8** (f) and **12.18.**
b_2	$\beta_6-4\beta_2\,\beta_4+4\beta_2^3-\beta_2^2+16\beta_2\,\beta_1-8\beta_3+16\beta_1$*	For normal parent, variance is $24/n$. See also **12.18** and Exercise 12.12.
Coefficient of variation, V	See Example 10.5	For normal parent, $=\mathbf{V}^2/2$ approx.
Pearson mode	See Yasukawa (1926) for formulae and tables	Distribution skew for moderate n.
Mean deviation	See **10.13**	For normal parent, $=\sigma^2(1-2/\pi)$.
Gini's mean difference	See **10.14**	For normal parent, $=(0{\cdot}8068)^2\,\sigma^2$.
Median	$1/(4y_0^2)$ when y_0 is ordinate at median	For normal parent and small samples see Hojo (1931) and K. Pearson (1931). For large normal samples equals $(1{\cdot}2533)^2\sigma^2$.
Quartile	$3/(16y^2)$ where y is the ordinate at the quartile	For normal parent, $=(1{\cdot}3626)^2\,\sigma^2$. See also Hojo (1931).
Deciles	See **10.10**	For normal parent, (deciles 4, 6) $=(1{\cdot}2680)^2\,\sigma^2$; (deciles 3, 7) $=(1{\cdot}3180)^2\,\sigma^2$; (deciles 2, 8) $=(1{\cdot}4288)^2\,\sigma^2$; (deciles 1, 9) $=(1{\cdot}7094)^2\,\sigma^2$.
Semi-interquartile range	$\frac{1}{4}\{3/16y_1^2+3/16y_2^2-1/8y_1\,y_2\}\sigma^2$, where y_1, y_2 are the quartile ordinates	For normal parent, $=(0{\cdot}7867)^2\,\sigma^2$.
Correlation coefficient r (product-moment)	See Example 10.6	For bivariate normal parent, $=(1-\rho^2)^2$. Not to be recommended except for very large samples. See Chapter 16.

* See equation (3.88) for the definition of the β-coefficients.

10.5 If the variate-range is divided into sub-ranges and the frequency of a large sample falling into the pth range is f_p, show that

$$\text{var}\, f_p = f_p\left(1-\frac{f_p}{n}\right)$$

$$\text{cov}\,(f_p, f_q) = -\frac{1}{n}f_p f_q$$

and hence find expressions for the sampling variance of the rth moment about an arbitrary point.

10.6 Show that for large samples from a symmetrical distribution any sample moment of odd order is uncorrelated (has zero covariance) with any other sample moment of even order.

10.7 In the notation of this chapter show that, to order n^{-1},

$$\operatorname{cov}(m_r, m'_q) = (\mu_{q+r} - \mu_q \mu_r - r\mu_{q+1}\mu_{r-1})/n.$$

10.8 Show that, for a symmetrical distribution, the covariance in large samples of two sample-moments of even order cannot be negative.

10.9 The variance of m_2 in large samples is $(\mu_4 - \mu_2^2)/n$. If this is estimated as $(m_4 - m_2^2)/n$, show that the variance of the estimated $\operatorname{var} m_2$ in samples from a normal population is $56\sigma^8/n^2$. Compare with the value of $\operatorname{var} m_2$ in normal variation, $2\sigma^4/n$.

10.10 Verify the formula for $\operatorname{var} b_1$ given in **10.15**.

10.11 Verify equation (10.46).

10.12 Show that the sampling variances of the first four cumulants, as calculated from the moments, are given to order n^{-1} by

$$\operatorname{var}\ \kappa_1 = \frac{1}{n}\kappa_2$$

$$\operatorname{var}\ \kappa_2 = \frac{1}{n}(\kappa_4 + 2\kappa_2^2)$$

$$\operatorname{var}\ \kappa_3 = \frac{1}{n}(\kappa_6 + 9\kappa_4\kappa_2 + 9\kappa_3^2 + 6\kappa_2^3)$$

$$\operatorname{var}\ \kappa_4 = \frac{1}{n}(\kappa_8 + 16\kappa_6\kappa_2 + 48\kappa_5\kappa_3 + 34\kappa_4^2 + 72\kappa_4\kappa_2^2 + 144\kappa_3^2\kappa_2 + 24\kappa_2^4).$$

(Note. The sample values of the κ's on the left have been written with Greek letters; to replace them in the customary way by k's would involve confusion with the k-statistics introduced in Chapter 12.)

10.13 Show that the mean value of the variance is given exactly by

$$E(m_2) = \frac{n-1}{n}\mu_2$$

and that its variance is given exactly by

$$\operatorname{var}\ m_2 = \left(\frac{n-1}{n}\right)^2 \frac{\mu_4 - \mu_2^2}{n} + \frac{2(n-1)}{n^3}\mu_2^2.$$

Hence verify that the formulae of this chapter as applied to the variance of a sample are accurate to order n^{-1}.

10.14 Verify that the standard error of the first decile in samples from a normal population is $1{\cdot}7094\sigma/\sqrt{n}$.

10.15 Show that in odd samples of n from a population with frequency-function uniform over a unit range the sampling variance of the distribution of the median is given exactly by $1/\{4(n+2)\}$.

10.16 In the notation of this chapter, if two variates are independent, if the distribution of each is symmetric, and if $r+s$ is an even number, show that

$$n \operatorname{cov}(m_{r,s}, m_{u,v}) = \mu_{r+u,0}\mu_{0,s+v} - \mu_{r0}\mu_{u0}\mu_{0s}\mu_{0v}.$$

10.17 A statistic t has a corresponding population value τ and to order n^{-1}, $\operatorname{var} t = f(\tau)$. Show that if

$$u = \int^t \{f(t)\}^{-\frac{1}{2}}\, dt,$$

then for large samples $\operatorname{var} u = 1$. Examine the case of the correlation coefficient, for which $\operatorname{var} r = (1-\rho^2)^2/n$ in large samples.

10.18 In large normal samples, the fourth moment μ_4 is estimated (a) as equal to m_4; (b) as equal to $3m_2^2$. Show that the variance of estimate (a) is 4/3 times that of (b).

10.19 Show that in samples of size n from the rectangular population $dF = k\,dx$, $0 \leqslant x \leqslant 1/k$, the variance of the mean deviation is approximately $1/(48k^2n)$.

10.20 A sample of values $x_1, x_2, \ldots, x_n$ are drawn from one population and another sample of $y_1, y_2, \ldots, y_m$ from a second population with the same variance μ_2. The mean of the $m+n$ values is z. Show that the standard error of the difference of z and $\bar{x}$ is given by

$$\sqrt{\frac{m\mu_2}{n(m+n)}}.$$

10.21 Two lists of M and N items have D items in common. Independent random samples of sizes m and n respectively ($m \geqslant n \geqslant 2$) are taken without replacement and found to have d items in common. Show that an unbiassed estimate of D is given by

$$\hat{D} = \frac{MN}{mn}d$$

and that

$$\operatorname{var} \hat{D} = \frac{MND}{mn}\left\{1 + \frac{m-1}{M-1}\cdot\frac{n-1}{N-1}(D-1)\right\} - D^2.$$

Show that as d approaches n, $\hat{D}$ exceeds N and thus that the estimate takes " impossible " values.
(Cf. L. A. Goodman (1952), *Ann. Math. Statist.*, **23**, 632 ; also W. E. Deming and G. J. Glasser (1959), *J. Amer. Statist. Ass.*, **54**, 403.)

10.22 In Exercise 10.21, show that an unbiassed estimate of $\operatorname{var} \hat{D}$ is given by

$$\hat{V} = \hat{D}\left(\frac{M-1}{m-1}\cdot\frac{N-1}{n-1} - 1\right) + \hat{D}^2\left(1 - \frac{mn}{MN}\cdot\frac{M-1}{m-1}\cdot\frac{N-1}{n-1}\right)$$

and hence that $\hat{V}$ may be negative, and thus an " impossible " estimate of $\operatorname{var} \hat{D}$, when m, n, M and N are all small.

10.23 x and y are jointly distributed with $E(x) = \mu'_{10} \neq 0$, $E(y) = \mu'_{01} \neq 0$.
Show that

$$\operatorname{var}(xy) = \mu'^2_{10}\mu_{02} + \mu_{20}\mu'^2_{01} + 2\mu'_{10}\mu_{11}\mu'_{01} + 2\mu'_{10}\mu_{12} + 2\mu_{21}\mu'_{01} + \mu_{22} - \mu^2_{11},$$

exactly, and that when x is independent of y, this reduces to

$$\operatorname{var}(xy) = \mu'^2_{10}\mu_{02} + \mu_{20}\mu'^2_{01} + \mu_{20}\mu_{02}. \qquad \text{(A)}$$

Show that this agrees to order n^{-1} with the large-sample result obtained from (10.12) (which is simply (A) with the last term on the right omitted) if μ_{20} and μ_{02} are of order n^{-1} as there required.

If m_{20}, m_{02} are unbiassed estimates of μ_{20}, μ_{02}, show that an unbiassed estimate of (A) is

$$x^2 m_{02} + y^2 m_{20} - m_{20}m_{02}.$$

(L. A. Goodman (1960), *J. Amer. Statist. Ass.*, **55**, 708)

CHAPTER 11

EXACT SAMPLING DISTRIBUTIONS

11.1 The role of the sampling distribution in statistical inference has been indicated in Chapter 7. In the present chapter we propose to give an account of the main methods of finding such distributions when the population from which the sample was derived is specified. It will, as usual, be assumed that the sampling is simple and random. Thus, if the parent distribution is $dF(x)$ the simultaneous distribution of n values $x_1, \ldots, x_n$ is $dF(x_1)\,dF(x_2)\ldots dF(x_n)$; and if z is a statistic

$$z = z(x_1, \ldots, x_n) \tag{11.1}$$

the distribution function of z is given by

$$F(z_0) = \int \ldots \int dF(x_1) \ldots dF(x_n), \tag{11.2}$$

the integration being taken over the domain of the x's such that $z(x_1, \ldots, x_n) \leqslant z_0$.

Formally, (11.2) is the solution of our problem, which thus reduces to the purely mathematical one of evaluating certain multiple integrals or sums. The methods with which we are here concerned are fundamentally devices of various kinds to facilitate the integrative process. They may be classified into four groups:—

(a) straightforward evaluation of the integral (11.2) by ordinary analytical processes such as a convenient change of variable;
(b) the use of geometrical terminology to effect the same object and to avoid cumbrous analytical formulae;
(c) the use of characteristic functions; and
(d) other analytical methods, including mathematical induction.

11.2 As an illustration of the straightforward analytical approach, let us find the distribution of the sums of squares of n independent variables, each of which is distributed normally with unit variance and zero mean. The joint distribution of the n variables is then the product of n quantities of type $\frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2}$, that is to say

$$dF = \frac{1}{(2\pi)^{\frac{1}{2}n}} \exp\{-\tfrac{1}{2}(x_1^2 + x_2^2 + \ldots + x_n^2)\}\, dx_1 \ldots dx_n. \tag{11.3}$$

We require the sampling distribution of

$$z = x_1^2 + x_2^2 + \ldots + x_n^2. \tag{11.4}$$

We have thus to evaluate the multiple integral

$$F = \int \ldots \int \frac{1}{(2\pi)^{\frac{1}{2}n}} \exp(-\tfrac{1}{2}\Sigma x^2)\, dx_1 \ldots dx_n$$

over the domain of x's conditioned by (11.4).

Make the transformation to variables $z, \theta_1, \theta_2, \ldots, \theta_{n-1}$

$$\left.\begin{aligned} x_1 &= z^{\frac{1}{2}} \cos\theta_1 \cos\theta_2 \ldots \cos\theta_{n-1} \\ x_2 &= z^{\frac{1}{2}} \cos\theta_1 \cos\theta_2 \ldots \cos\theta_{n-2} \sin\theta_{n-1} \\ &\ldots \quad \ldots \quad \ldots \quad \ldots \\ x_j &= z^{\frac{1}{2}} \cos\theta_1 \cos\theta_2 \ldots \cos\theta_{n-j} \sin\theta_{n-j+1} \\ &\ldots \quad \ldots \quad \ldots \quad \ldots \\ x_n &= z^{\frac{1}{2}} \sin\theta_1 \end{aligned}\right\} \tag{11.5}$$

The Jacobian of this transformation is given by

$$\frac{\partial(x_1, \ldots, x_n)}{\partial(z, \theta_1, \ldots, \theta_{n-1})},$$

which is equal to $\frac{1}{2} z^{\frac{1}{2}n-1}$ times the determinant

$$\begin{vmatrix} \cos\theta_1 \cos\theta_2 \ldots \cos\theta_{n-1} & \cos\theta_1 \cos\theta_2 \ldots \cos\theta_{n-2} \sin\theta_{n-1} \ldots & \sin\theta_1 \\ -\sin\theta_1 \cos\theta_2 \ldots \cos\theta_{n-1} & -\sin\theta_1 \cos\theta_2 \ldots \cos\theta_{n-2} \sin\theta_{n-1} \ldots & \cos\theta_1 \\ -\cos\theta_1 \sin\theta_2 \ldots \cos\theta_{n-1} & -\cos\theta_1 \sin\theta_2 \ldots \cos\theta_{n-2} \sin\theta_{n-1} \ldots & 0 \\ \ldots \quad \ldots & \ldots \quad \ldots \quad \ldots & \ldots \\ -\cos\theta_1 \cos\theta_2 \ldots \sin\theta_{n-1} & +\cos\theta_1 \cos\theta_2 \ldots \cos\theta_{n-1} \quad \ldots & 0 \end{vmatrix}$$

Taking out common factors in columns we find that this determinant is equal to

$$\cos^{n-1}\theta_1 \cos^{n-2}\theta_2 \ldots \cos\theta_{n-1} \sin\theta_1 \sin\theta_2 \ldots \sin\theta_{n-1} \text{ times}$$

$$\begin{vmatrix} 1 & 1 & 1 & \ldots & 1 \\ -\tan\theta_1 & -\tan\theta_1 & -\tan\theta_1 & \ldots & \cot\theta_1 \\ -\tan\theta_2 & -\tan\theta_2 & -\tan\theta_2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ -\tan\theta_{n-2} & -\tan\theta_{n-2} & \cot\theta_{n-2} & \ldots & 0 \\ -\tan\theta_{n-1} & \cot\theta_{n-1} & 0 & \ldots & 0 \end{vmatrix}$$

and, on subtracting each column from the preceding one, the determinant is found to reduce to $\cos^{n-2}\theta_1 \cos^{n-3}\theta_2 \ldots \cos\theta_{n-2}$.

Thus our integral becomes

$$\int \ldots \int \frac{1}{(2\pi)^{\frac{1}{2}n}} e^{-\frac{1}{2}z} \tfrac{1}{2} z^{\frac{1}{2}n-1} \cos^{n-2}\theta_1 \ldots \cos\theta_{n-2}\, dz\, d\theta_1 \ldots d\theta_{n-1}. \tag{11.6}$$

The advantage of the transformation is that the limits of the variables are now much simpler. z itself can vary from 0 to z, θ_{n-1} from 0 to 2π and the other θ's from $-\frac{1}{2}\pi$ to $\frac{1}{2}\pi$. Thus the integral (11.6) divides into a product of integrals, those in θ being constant, and we find for our distribution function of z

$$F(z) = k \int_0^z e^{-\frac{1}{2}z} z^{\frac{1}{2}n-1}\, dz. \tag{11.7}$$

The constant k may be evaluated by integration between 0 and ∞ and hence the distribution sought is

$$dF = \frac{1}{2^{\frac{1}{2}n}\Gamma(\frac{1}{2}n)} e^{-\frac{1}{2}z} z^{\frac{1}{2}n-1}\, dz, \qquad 0 \leqslant z \leqslant \infty, \tag{11.8}$$

a Pearson Type III distribution.

11.3 The essential feature of the change of variables is the simplification of the

domain of integration as defined by the limits of the new variables. In general, we take the statistic whose sampling distribution is being sought as one of the new variables and choose $n-1$ others in any way which may be convenient to the particular problem. Then, if J is the Jacobian of the transformation, namely

$$J = \frac{\partial(x_1, \ldots, x_n)}{\partial(z, \theta_1, \ldots, \theta_{n-1})},$$

the integral (11.2) becomes

$$F(z) = \int \ldots \int f(x_1) \ldots f(x_n) \frac{\partial(x_1, \ldots, x_n)}{\partial(z, \theta_1, \ldots, \theta_{n-1})} dz\, d\theta_1 \ldots d\theta_{n-1}, \tag{11.9}$$

$f(x_j)$ being the frequency function of the parent and x_j being expressed in terms of z and the θ's. The integration now takes place with respect to the θ's, which can usually be chosen so as to vary between limits which are independent of z; and thus the integral (11.2) is replaced by more easily calculable definite integrals.

As always in such cases J is subject to an ambiguity of sign which must be determined so as to make the transformed integral positive. The validity of the variate-transformation depends on the familiar conditions governing the change of variable in a multiple integral. For example, it is a sufficient condition that the new variables and their first derivatives shall be continuous in the x's and that J does not change sign in the domain of integration. Some further examples will make the general type of investigation clear.

Example 11.1

To find the distribution of the mean of a sample of n values $x_1, \ldots, x_n$ from the distribution

$$dF = \frac{dx}{\pi(1+x^2)}, \qquad -\infty \leqslant x \leqslant \infty.$$

The joint distribution is

$$\frac{1}{\pi^n} \prod_{j=1} \frac{dx_j}{(1+x_j^2)}, \tag{11.10}$$

and the statistic z is given by

$$nz = \sum_{j=1}^{n} x_j. \tag{11.11}$$

We have to integrate (11.10) over a domain of x's subject to $\Sigma x \leqslant nz$. Let us take new variables $u_1 = x_1$, $u_2 = x_2$, . . . , $u_{n-1} = x_{n-1}$, and $u_n = \frac{1}{n}\sum_1^n x_j = z$. Here J is evidently equal to the constant n. Our new variables $u_1, \ldots, u_{n-1}$ may extend from $-\infty$ to $+\infty$ and the new variable u_n from $-\infty$ to z. We then have

$$F(z) = \int_{-\infty}^{z} du_n \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \frac{n}{\pi^n} \prod_{j=1}^{n-1} \frac{du_j}{(1+u_j^2)\{1+(nu_n - u_1 - \ldots - u_{n-1})^2\}} \tag{11.12}$$

and the *frequency* function of z is given by the $(n-1)$-fold multiple integral in u_1,

. . , u_{n-1} in (11.12). This integral may be evaluated by step-by-step integration We have

$$\frac{1}{(1+x^2)\{r^2+(a-x)^2\}} = \frac{1}{\{a^2+(r+1)^2\}\{a^2+(r-1)^2\}} \left[\frac{2ax}{x^2+1}+\frac{a^2+r^2-1}{x^2+1}+\frac{2a^2-2ax}{r^2+(a-x)^2}+\frac{a^2-r^2+1}{r^2+(a-x)^2}\right]$$

whence, integrating with respect to x from $-\infty$ to $+\infty$, we find on the right

$$\frac{1}{\{a^2+(r+1)^2\}\{a^2+(r-1)^2\}}\left[a\log(x^2+1)-a\log\{r^2+(a-x)^2\} + (a^2+r^2-1)\arctan x+\frac{a^2-r^2+1}{r}\arctan\frac{x-a}{r}\right]_{-\infty}^{\infty}$$

reducing to

$$\pi\left(\frac{r+1}{r}\right)\left\{\frac{1}{a^2+(r+1)^2}\right\}. \tag{11.13}$$

Thus in (11.12), taking $x = u_{n-1}$, $r = 1$, $a = nu_n - u_1 - \ldots - u_{n-2}$, we find that the $(n-1)$-fold integral reduces to

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\frac{n}{\pi^{n-1}}\prod_{j=1}^{n-2}\frac{du_j}{(1+u_j^2)}\{2^2+(nu_n-u_1-\ldots-u_{n-2})^2\}.$$

Integrating with respect to u_{n-2}, u_{n-3}, . . . successively, we reduce this eventually to

$$\frac{n^2}{\pi\{n^2+(nu_n)^2\}} = \frac{1}{\pi(1+u_n^2)}. \tag{11.14}$$

Thus the distribution of z is given by

$$dF = \frac{dz}{\pi(1+z^2)}, \qquad -\infty \leqslant z \leqslant \infty, \tag{11.15}$$

and is thus the same as that of a single observation.

This is an interesting example of the failure of the Central Limit theorem, the mean of samples of n observations failing to tend to normality for large n. The moments of the distribution do not exist.

Example 11.2 ✓

To find the distribution of a linear function of n independent variables x_1, . . , x_n where x_j is distributed normally with zero mean and variance σ_j^2.

Let the linear function be

$$z = a_1x_1 + \ldots + a_nx_n. \tag{11.16}$$

Then by a transformation $u_j = x_j/\sigma_j$ we have

$$z = \Sigma a_j\sigma_j u_j \tag{11.17}$$

and u_j is now distributed with zero mean and unit variance. Our problem is thus equivalent to finding the distribution of a linear function of variables each of which is normally distributed with zero mean and unit variance.

Consider a transformation of type

$$v_i = \sum_{j=1}^{n} c_{ij}u_j \tag{11.18}$$

and let us determine the c's so that they are orthogonal, i.e. so that

$$\left.\begin{aligned} \sum_{i=1}^{n} c_{ij}c_{ik} &= 0, \quad j \neq k \\ &= 1, \quad j = k \end{aligned}\right\}. \tag{11.19}$$

This can always be done, for (11.19) imposes $\frac{1}{2}n(n-1)+n$ conditions on n^2 constants. We then find

$$\sum_{i=1}^{n} v_i^2 = \Sigma(\Sigma c_{ij}u_j)^2 = \sum_{j=1}^{n} u_j^2.$$

Also, for the Jacobian of the transformation

$$J = \left|\frac{\partial v}{\partial u}\right| = |c_{ij}|.$$

If we multiply $|c_{ij}|$ by its transpose we find, in virtue of (11.19), a determinant whose elements are all zero except for units in the diagonals, namely a determinant which is equal to unity. Thus $J^2 = 1$ and $J = 1$. The joint distribution of the u's is, by hypothesis,

$$dF = \frac{1}{(2\pi)^{\frac{1}{2}n}} \exp\left(-\tfrac{1}{2}\sum_{j=1}^{n} u_j^2\right) \prod_j du \tag{11.20}$$

and hence that of the v's is of the same form :—

$$dF = \frac{1}{(2\pi)^{\frac{1}{2}n}} \exp\left(-\tfrac{1}{2}\sum_{j=1}^{n} v_j^2\right) \Pi\, dv. \tag{11.21}$$

The distribution of the other variables v is then, from (11.21), seen to be independent of v_1 and we have, for the (marginal) distribution of v_1,

$$dF = \frac{1}{\sqrt{(2\pi)}} \exp(-\tfrac{1}{2}v_1^2)\, dv_1. \tag{11.22}$$

Thus v_1 is distributed normally with unit variance and zero mean, independently of the other v_i. By symmetry, the same is true of any v_j. Now v_1 is an arbitrary linear function of the x_j. If we identify it with z of (11.16), we have the result that z is normally distributed with zero mean and variance obtained from (11.17) as

$$\text{var}\, z = \sum_j a_j^2 \sigma_j^2. \tag{11.23}$$

Example 11.3 Helmert's transformation

The kind of transformation we used in the previous example is especially useful for samples from a normal population ; for the Jacobian is a constant and the density function remains of the same form. There is one special transformation which has a particular interest. Let us put

$$\left.\begin{aligned} u_1 &= (x_1-x_2)/\sqrt{2} \\ u_2 &= (x_1+x_2-2x_3)/\sqrt{6} \\ u_3 &= (x_1+x_2+x_3-3x_4)/\sqrt{12} \\ \cdots &\quad \cdots \quad \cdots \\ u_{n-1} &= \{x_1+x_2+\ldots+x_{n-1}-(n-1)x_n\}/\sqrt{\{n(n-1)\}} \\ u_n &= (x_1+x_2+\ldots+x_n)/\sqrt{n}. \end{aligned}\right\} \tag{11.24}$$

It is readily verified that the Jacobian of the transformation is unity, that the sum of squares of coefficients in any of these equations is unity and that the sum of cross-products of coefficients in any pair of equations is zero. Thus, if the x's are all independent and normally distributed about zero mean with unit variance, so also are the u's.

Now

$$\begin{aligned}\Sigma(x-\bar{x})^2 &= \Sigma x^2 - n\bar{x}^2 \\ &= \sum_{i=1}^{n} u_i^2 - u_n^2 \\ &= \sum_{i=1}^{n-1} u_i^2.\end{aligned}$$

Thus we see that the sum of squares of n standardized variates *measured from their mean* is distributed like the sum of squares of $n-1$ normal variates with zero mean. It follows from (11.8) that the distribution of

$$w = \Sigma(x-\bar{x})^2$$

is given by

$$dF = \frac{1}{2^{\frac{1}{2}(n-1)}\Gamma\{\frac{1}{2}(n-1)\}} e^{-\frac{1}{2}w} w^{\frac{1}{2}(n-3)}\, dw. \tag{11.25}$$

Moreover, this quantity w (a simple multiple of the sample variance) is independent of $\bar{x}$. Thus: in normal samples the mean is distributed independently of the variance.

This interesting and important result is characteristic of the normal distribution; that is to say, if the mean and variance of random samples from a population are independent, that population must be normal. (Cf. Exercise 11.19.)

Order-statistics

11.4 It may not always be possible to find variate-transformations in which z, the statistic of interest, is one new variate and there are $(n-1)$ others whose domains of integration are not dependent on z. A case of particular interest is provided by the so-called "order-statistics," the theory of which we shall consider in detail in Chapter 14 and in the second volume. If we have a sample of n observations, y_1, y_2, . . . , y_n, and rearrange them in ascending order of magnitude, renumbering them so that the suffix 1 is attached to the smallest, y_r is called the rth order-statistic. The median and the quantiles are order-statistics in this sense (the median, for example, being the $\frac{1}{2}(n+1)$th observation or midway between the $\frac{1}{2}n$th and the $\frac{1}{2}(n+2)$th); so also would be the lowest value of the sample ($r = 1$) or the highest value ($r = n$).

If we have a sample of n values x_1, x_2, . . . , x_n from a distribution function $F(x)$, their distribution is given by

$$dG = dF(x_1)\,dF(x_2)\ldots dF(x_n). \tag{11.26}$$

A transformation to

$$y = \int_{-\infty}^{x} dF(t) = F(x)$$

gives for the distribution of the y's

$$dH = dy_1 dy_2 \ldots dy_n, \qquad (0 \leqslant y_i \leqslant 1, \text{ all } i). \tag{11.27}$$

The y's and the x's are in the same order since y is a non-decreasing function of x.

If we now order the y's so that y_1 is the smallest, y_2 the next smallest and so on, the domain of variation is given by

$$0 \leqslant y_1 \leqslant y_2 \leqslant \ldots \leqslant y_n \leqslant 1. \tag{11.28}$$

To integrate (11.27) over this domain we require

$$\int_0^1 \ldots \left[\int_0^{y_4} \left\{ \int_0^{y_3} \left(\int_0^{y_2} dy_1 \right) dy_2 \right\} dy_3 \right] \ldots dy_n. \tag{11.29}$$

This reduces, on step-by-step integration, to $1/n!$. For the *ordered* variates, then, the joint distribution of the y's is

$$dK = n! \prod_{i=1}^{n} dy_i. \tag{11.30}$$

Suppose now that we are interested in the (marginal) distribution of y_r. To simplify the derivation slightly we make the further transformation

$$\left.\begin{aligned} z_i &= y_i, & i &= 1, 2, \ldots, r \\ &= 1 - y_i, & i &= r+1, r+2, \ldots, n \end{aligned}\right\} \tag{11.31}$$

The Jacobian is unity and the distribution of the z's is accordingly

$$dL = n! \prod_{i=1}^{n} dz_i \tag{11.32}$$

over the domain

$$0 \leqslant z_1 \leqslant z_2 \leqslant \ldots \leqslant z_r \leqslant 1 - z_{r+1} \leqslant \ldots \leqslant 1 - z_n \leqslant 1$$

which may be re-written

$$0 \leqslant z_1 \leqslant z_2 \leqslant \ldots \leqslant z_r\,; \quad 0 \leqslant z_n \leqslant z_{n-1} \leqslant \ldots \leqslant z_{r+1} \leqslant 1 - z_r.$$

To find the distribution of z_r we integrate out the other variables in two separate blocks, as follows:

$$\begin{aligned} dM &= dz_r n! \int_0^{z_r} \ldots \int_0^{z_3} \int_0^{z_2} dz_1 dz_2 \ldots dz_{r-1} \times \int_0^{1-z_r} \int_0^{z_{r+1}} \ldots \int_0^{z_n} {}^{1} dz_n dz_{n-1} \ldots dz_{r+1} \\ &= dz_r n! \frac{z_r^{r-1}}{(r-1)!} \frac{(1-z_r)^{n-r}}{(n-r)!} \\ &= \frac{n!}{(r-1)!(n-r)!} z_r^{r-1} (1-z_r)^{n-r} dz_r, \qquad 0 \leqslant z_r \leqslant 1. \end{aligned} \tag{11.33}$$

Thus the distribution of z_r is a Beta-distribution of the first kind. To obtain that of the rth order-statistic x_r we simply put $F(x_r) = z_r$, obtaining

$$dN = \frac{n!}{(r-1)!(n-r)!} \{F(x_r)\}^{r-1} \{1 - F(x_r)\}^{n-r} f(x_r)\, dx_r. \tag{11.34}$$

This expression could have been obtained more directly (cf. (10.25)), but we have given a full derivation to illustrate the processes involved.

Example 11.4

If n is an odd number, say $2m+1$, the $(m+1)$th sample member is the median and for the distribution of z_{m+1}, say z, we have

$$dF = \frac{(2m+1)!}{(m!)^2} z^m (1-z)^{m+1} dz.$$

The variance of this is easily found to be $1/\{4(2m+3)\} = 1/\{4(n+2)\}$.

Since the median x_{m+1} is related to z_{m+1} by

$$z_{m+1} = \int_{-\infty}^{x_{m+1}} dF(x),$$

we have

$$\frac{dz_{m+1}}{dx_{m+1}} = f(x_{m+1})$$

and, by (10.14), it follows that in large samples

$$\begin{aligned} \operatorname{var} x_{m+1} &= \operatorname{var} z_{m+1} \{f(x_{m+1})\}^{-2} \\ &= 1/\{4nf^2(x_{m+1})\}, \end{aligned}$$

agreeing with (10.29).

The geometrical method

11.5 A considerable amount of cumbrous analysis may usually be avoided by the use of geometrical representation of the domain of integration. We may imagine the values $x_1, \ldots, x_n$ of any given sample as the co-ordinates of a point in an n-dimensional Euclidean hyperspace. The function $dF(x_1) \ldots dF(x_n)$ may then be regarded as the *density* at the point and the total frequency between z_1 and z_2 will be the integral of this density (the *weight*) in a region lying between the two loci $z(x_1, \ldots, x_n) = z_1$ and $z(x_1, \ldots, x_n) = z_2$, which in general will be hypersurfaces in the n-fold space, i.e. will themselves be spaces of $(n-1)$ dimensions. The distribution function of z will be the total weight between the hypersurface corresponding to $z = -\infty$ and that corresponding to z; and the frequency function will be the element of weight between the hypersurfaces $z - \frac{1}{2}dz$ and $z + \frac{1}{2}dz$.

Example 11.5

Consider again the problem of Example 11.2. In the n-fold u-space the density is given by

$$\frac{1}{(2\pi)^{\frac{1}{2}n}} \exp(-\tfrac{1}{2}\Sigma u^2).$$

The statistic z $(= \Sigma a_j x_j)$ determines a hyperplane

$$z = \Sigma a_j \sigma_j u_j \tag{11.35}$$

and we have to find the total weight between this hyperplane and the corresponding hyperplane at $-\infty$, i.e. the weight on one side—the "lower" side—of the hyperplane (11.35).

Now Σu^2 is the square of the distance of the point $u_1, \ldots, u_n$ from the origin

and is therefore unchanged by any rotation of the co-ordinate axes. Choose such a rotation which brings the axis of one variable perpendicular to the hyperplane (11.35), meeting it in Q. Let P be the sample point $u_1, \ldots, u_n$ and O the origin. Then

$$\Sigma u^2 = OP^2 = OQ^2 + QP^2,$$

so that the density at P is

$$\frac{1}{(2\pi)^{\frac{1}{2}n}} \exp(-\tfrac{1}{2}OQ^2) \exp(-\tfrac{1}{2}QP^2).$$

For variation over the hyperplane, OQ^2 is constant and the integral of $\exp(-\frac{1}{2}QP^2)$ is thus a constant independent of OQ. Hence the frequency function of z is given by

$$f(z) = k \exp(-\tfrac{1}{2}OQ^2),$$

k being some constant.

But OQ is the distance from O to the hyperplane and is given by

$$OQ^2 = \frac{z^2}{\Sigma a_j^2 \sigma_j^2}.$$

Hence

$$f(z) = k \exp\left\{-\frac{1}{2}\frac{z^2}{\Sigma a_j^2 \sigma_j^2}\right\},$$

i.e. z is distributed normally with variance $\Sigma a_j^2 \sigma_j^2$ about zero mean.

The reader will find it instructive to compare this example with Example 11.2. They are, in effect, the same thing expressed in different language.

Example 11.6

Consider again the illustration of **11.2.** The elegance of the geometrical approach is well brought out by the analogous derivation of the result there obtained.

In fact, our density function, as before, is given by $k \exp(-\frac{1}{2}OP^2)$. We require the distribution of the statistic $z = OP^2$, and the density is obviously constant over the surface z = constant, that is to say the surface of an n-dimensional hypersphere. The frequency function of z is then the integral of this constant density between the hyperspheres z and $z+dz$, i.e. is proportional to $\exp(-\frac{1}{2}OP^2)$ times the element of the volume of the hypersphere, which itself is proportional to the nth power of the radius OP. Thus we have

$$dF = k \exp(-\tfrac{1}{2}OP^2) \frac{d}{dz} OP^n \, dz$$

$$\propto e^{-\frac{1}{2}z} z^{\frac{1}{2}(n-2)} \, dz,$$

giving, on evaluation of the constant,

$$dF = \frac{1}{2^{\frac{1}{2}n}\Gamma(\frac{1}{2}n)} e^{-\frac{1}{2}z} z^{\frac{1}{2}(n-2)}$$

as at (11.8).

Now suppose that the quantities $x_1, \ldots, x_n$, while still being normally distributed with unit variance, are subject to p homogeneous linear restrictions of type

$$a_1 x_1 + a_2 x_2 + \ldots + a_n x_n = 0.$$

In the n-space the variables x will then be constrained to lie on p hyperplanes. The first will cut the hypersphere of constant density in a hypersphere of one lower dimension, also, of course, of constant density; the second will cut this in a hypersphere of one lower dimension still, and so on. The result of the linear restrictions will be to constrain the variables to a hypersphere of p fewer dimensions, and thus the distribution of z in these circumstances will be as before, but with $n-p$ instead of n, i.e.

$$dF = \frac{1}{2^{\frac{1}{2}(n-p)}\Gamma\{\frac{1}{2}(n-p)\}} e^{-\frac{1}{2}z} z^{\frac{1}{2}(n-p-2)}\, dz. \tag{11.36}$$

Example 11.7 The sampling distribution of the mean and variance in normal samples

Writing $\bar{x}$ for the mean of a sample, we have, for the variance s^2,

$$\begin{aligned} s^2 &= \frac{1}{n}\Sigma(x-\bar{x})^2 \\ &= \frac{1}{n}\Sigma x^2 - \bar{x}^2. \end{aligned}$$

In samples from a normal population with zero mean and unit variance the density at the point $x_1, \ldots, x_n$ is proportional to

$$\exp(-\tfrac{1}{2}\Sigma x^2) = \exp\{-\tfrac{1}{2}(ns^2+n\bar{x}^2)\}. \tag{11.37}$$

Let us find the sampling distributions of s and $\bar{x}$. From (11.37) it is seen that the density function can be expressed simply in terms of those quantities, and we then have to find some transformation of the volume element $dx_1 \ldots dx_n$. We have considered this from the analytical viewpoint in Example 11.3. Let us now consider it geometrically.

In the n-space consider the unit vector whose direction cosines are $\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}}$, say OQ where O is the origin. If P is the sample point, let PM be the perpendicular from P on to OQ. Then the length of OM is

$$\frac{x_1}{\sqrt{n}} + \frac{x_2}{\sqrt{n}} + \ldots + \frac{x_n}{\sqrt{n}} = \bar{x}\sqrt{n}.$$

The length of OP is $\sqrt{\Sigma x^2}$. Thus the length of PM is $(\Sigma x^2 - n\bar{x}^2)^{\frac{1}{2}} = s\sqrt{n}$.

The element of volume at P may be regarded as the product of an elemental increment in OM, equal to $d\bar{x}\sqrt{n}$, and the elemental volume in the perpendicular hyperplane through M. In the hyperplane the contours of equal density, as in the last example, are hyperspheres of radius $s\sqrt{n}$ centred at M, and consequently the element of volume is equal to $k\, d\bar{x}\, s^{n-2}\, ds$ multiplied by other elements which need not concern us since they are independent of $\bar{x}$ and s. We have then for the element of frequency

$$dF \propto \exp\{-\tfrac{1}{2}(ns^2+n\bar{x}^2)\}\, s^{n-2}\, d\bar{x}\, ds, \tag{11.38}$$

and this splits into two factors

$$dF_1 \propto \exp(-\tfrac{1}{2}n\bar{x}^2)\, d\bar{x}, \tag{11.39}$$

$$dF_2 \propto \exp(-\tfrac{1}{2}ns^2)\, s^{n-2}\, ds. \tag{11.40}$$

Thus, as we have seen, in samples from a normal population the distributions of mean

and variance are independent. We have from (11.40), for the distribution of s^2,

$$dF \propto \exp(-\tfrac{1}{2}ns^2)\, s^{n-3}\, ds^2$$

and, on evaluation of the constant,

$$dF = \frac{n^{\frac{1}{2}(n-1)}}{2^{\frac{1}{2}(n-1)}\Gamma\{\frac{1}{2}(n-1)\}} \exp(-\tfrac{1}{2}ns^2)\,(s^2)^{\frac{1}{2}(n-3)}\, d(s^2), \qquad 0 \leqslant s^2 \leqslant \infty, \qquad (11.41)$$

which is (11.25) again.

It is interesting to compare this with the distribution of the previous example. In the latter case we found the distribution of the sum of squares of the variables *measured from a fixed point.* In this case we have found the distribution of $1/n$ times the sum of the squares *measured from the sample mean.* A comparison of the form (11.41) with that of (11.36) shows that the distribution of variances is, except for constants, the same as that of sums of squares when subject to one linear constraint, as we observed in Example 11.3.

Example 11.8 " Student's " distribution

In the previous example we have

$$\frac{\bar{x}\sqrt{n}}{s\sqrt{n}} = \frac{OM}{PM} = \cot\phi,$$

where ϕ is the angle POM.

If, then, we define a statistic $z = \bar{x}/s$, z will be constant over the cone obtained by rotating OP about the unit vector, keeping the angle ϕ constant. The distribution of z will then be given by determining the weight between the cones defined by ϕ and $\phi + d\phi$.

Consider the intersection of these cones with the hypersphere of radius OP. They will cut off an annulus on the sphere whose " content " (the n-dimensional analogue of volume) will be proportional to $OP\, d\phi . PM^{n-2}$

$$= OP^{n-1} \sin^{n-2}\phi\, d\phi.$$

The density function is constant and proportional to $\exp(-\frac{1}{2}OP^2)$ on the hypersphere, and thus the total frequency between the cones will be proportional to

$$\int_0^\infty e^{-\frac{1}{2}OP^2} OP^{n-1} \sin^{n-2}\phi\, d\phi\, d(OP)$$

$$\propto \sin^{n-2}\phi\, d\phi, \qquad 0 \leqslant \phi \leqslant \pi.$$

The distribution of $z\,(= \cot\phi)$ is then given by

$$dF \propto \frac{k\, dz}{(1+z^2)^{\frac{1}{2}n}},$$

or, on evaluation of the constant,

$$dF = \frac{1}{B\{\frac{1}{2}(n-1), \frac{1}{2}\}} \frac{dz}{(1+z^2)^{\frac{1}{2}n}}. \qquad (11.42)$$

Since z is the ratio of two functions of unit dimension in the variables, this distribution holds for samples from a normal population irrespective of the scale, that is to say, irrespective of the variance of the parent population.

The distribution is usually put in a slightly different form.

Put $$t = \frac{\bar{x}\sqrt{n}}{\left\{\frac{1}{n-1}\Sigma(x-\bar{x})^2\right\}^{\frac{1}{2}}} = \sqrt{(n-1)}z.$$

(11.42) then becomes

$$dF = \frac{1}{\sqrt{(n-1)}B\{\frac{1}{2}(n-1), \frac{1}{2}\}} \frac{dt}{\left(1+\frac{t^2}{n-1}\right)^{\frac{1}{2}n}} \tag{11.43}$$

$$= \frac{\Gamma\{\frac{1}{2}(\nu+1)\}}{\sqrt{(\nu\pi)}\,\Gamma(\frac{1}{2}\nu)} \frac{dt}{(1+t^2/\nu)^{\frac{1}{2}(\nu+1)}} \tag{11.44}$$

where $\nu = n-1$.

This celebrated expression is known as " Student's " distribution after the pen-name of its discoverer (1908). It will be discussed in detail in Chapter 16.

Example 11.9 *Distribution of the mean of samples from a rectangular population* (Hall, P., 1927, *Biometrika*, **19**, 240)

Consider now a sample of n values from the rectangular distribution

$$dF = dx, \quad 0 \leqslant x \leqslant 1.$$

In the n-space the density function will be a constant everywhere inside a hypercube

$$0 \leqslant x_j \leqslant 1, \; j = 1, \ldots, n \tag{11.45}$$

and zero elsewhere. The unit vector will be along the diagonal of this cube. If P is the sample point $(x_1, \ldots, x_n)$ and PM the perpendicular on to this diagonal, then, as shown in Example 11.7, $OM = \bar{x}\sqrt{n}$. Thus, for the distribution of $\bar{x}$ we require the element of weight (which in this case is proportional to the element of volume) between the hyperplanes $\bar{x}$ and $\bar{x}+d\bar{x}$; and this is equivalent to finding the content of the hyperplane (its " area ") cut off by the various faces of the hypercube. The complication of the problem arises from the fact that as $\bar{x}$ increases this region changes its shape according to the number of edges of the hypercube cut by the hyperplane.

Consider the " quadrants "

$$\left.\begin{matrix} x_j \geqslant r_j \\ r_j = 0 \text{ or } 1 \end{matrix}\right\} j = 1, 2, \ldots, n, \tag{11.46}$$

whose corners are the corners of the hypercube. Any one of the corners may have 0 or 1 or 2 . . . or n of its co-ordinates equal to unity and the rest zero. We divide the quadrants into $(n+1)$ sets according as the corner has 0, 1, . . , n of its co-ordinates equal to unity, that is, according as

$$r = \sum_{j=1}^{n} r_j$$

is equal to 0, 1, . . , n. A quadrant of the tth set may be called Q_t. There will be $\binom{n}{t}$ different Q_t's. Let S be any point of Q_0, i.e. any point whose co-ordinates are all $\geqslant 0$, and let just s of its co-ordinates be $\geqslant 1$. Then S will belong to just $\binom{s}{0}$, $= 1$, Q_0; $\binom{s}{1}$ Q_1's; $\binom{s}{2}$ Q_2's, and so on. Now if $s > 0$,

$$\sum_{t=0}^{s} (-1)^t \binom{s}{t} = (1-1)^s = 0. \tag{11.47}$$

Hence, if whenever a point belongs to a Q_t we give it a density $(-1)^t$ and then sum over all Q, the resultant density will be 1 or 0 according as the point belongs to the hypercube or not.

Let the segment of the hyperplane

$$z = \Sigma x \tag{11.48}$$

lying in Q_0 have content $V_n(z)$. Then the segment lying in any member of (11.46) will have content $V_n(z-r)$ which is zero if $r \geqslant z$. Further, the segment of (11.48) lying in any member of (11.46) will have the content

$$\sum_{r=0}^{k} (-1)^r \binom{n}{r} V_n(z-r) \tag{11.49}$$

where $k = [z]$, = the greatest integer less than z.

To find $V_n(z)$, let $\bar{V}_{n-1}(z)$ be the projection of $V_n(z)$ perpendicular to one of the axes, so that

$$V_n(z) = \sqrt{n}\bar{V}_{n-1}(z).$$

Now $\bar{V}_n(z)$ is the content of the n-dimensional region bounded by (11.48) and the coordinate hyperplanes—a region whose base is therefore of content $V_n(z)$. The perpendicular from O to this base is $\frac{z}{\sqrt{n}}$. Hence

$$\bar{V}_n(z) = \frac{1}{n}\left(\frac{z}{\sqrt{n}}\right) V_n(z)$$

and

$$\bar{V}_{n-1}(z) = \frac{1}{n-1}\left(\frac{z}{\sqrt{(n-1)}}\right) V_{n-1}(z)$$

or

$$V_n(z) = \frac{1}{n-1}\sqrt{\left(\frac{n}{n-1}\right)}\, z V_{n-1}(z). \tag{11.50}$$

Since $V_2(z) = z\sqrt{2}$ repeated applications of this formula give

$$V_n(z) = \frac{\sqrt{n}}{(n-1)!} z^{n-1}.$$

Substituting in (11.49) we find for the content of the region common to the hypercube and the hyperplane

$$f(z) = \frac{\sqrt{n}}{(n-1)!} \sum_{r=0}^{k} (-1)^r \binom{n}{r} (z-r)^{n-1} \tag{11.51}$$

for values of z between k and $k+1$.

Since

$$\int_0^n f(z)\frac{dz}{\sqrt{n}} = 1$$

the distribution of the mean $m = \frac{z}{n}$ is given by

$$(m) = \frac{n^n}{(n-1)!} \sum_{r=0}^{k} (-1)^r \binom{n}{r} \left(m - \frac{r}{n}\right)^{n-1}, \quad \frac{k}{n} \leqslant m \leqslant \frac{k+1}{n}. \tag{11.52}$$

This is the required distribution. It is unusual in consisting of n arcs of degree $(n-1)$ in m, having $(n-1)$-point contact at their joins, that is at the points $\frac{k}{n}$ $(k = 1, 2, \ldots, n)$.

The distribution is symmetrical since the hyperplane z = constant is perpendicular to the unit vector, which itself is an axis of symmetry of the hypercube.

For particular values $n = 2, 3, 4$, (11.52) gives the following results for the frequency function :—

$$
\begin{array}{lll}
n=2: & 4m, & 0 \leqslant m \leqslant \tfrac{1}{2}, \\
 & 4(1-m), & \tfrac{1}{2} \leqslant m \leqslant 1. \\
n=3: & \dfrac{27m^2}{2}, & 0 \leqslant m \leqslant \tfrac{1}{3}, \\
 & \dfrac{27}{2}\{m^2-3(m-\tfrac{1}{3})^2\}, & \tfrac{1}{3} \leqslant m \leqslant \tfrac{2}{3}, \\
 & \dfrac{27}{2}(1-m)^2, & \tfrac{2}{3} \leqslant m \leqslant 1. \\
n=4: & \dfrac{128}{3}m^3, & 0 \leqslant m \leqslant \tfrac{1}{4}, \\
 & \dfrac{128}{3}\{m^3-4(m-\tfrac{1}{4})^3\}, & \tfrac{1}{4} \leqslant m \leqslant \tfrac{1}{2}, \\
 & \dfrac{128}{3}\{(1-m)^3-4(\tfrac{3}{4}-m)^3\}, & \tfrac{1}{2} \leqslant m \leqslant \tfrac{3}{4}, \\
 & \dfrac{128}{3}(1-m)^3, & \tfrac{3}{4} \leqslant m \leqslant 1.
\end{array}
$$

For $n = 2$ the distribution is " triangular." For larger n it approaches normality in virtue of the Central Limit theorem.

The method of characteristic functions

11.6 It has already been noted (**7.18**) that the characteristic function of the sum of n independent variables is the product of their characteristic functions. This simple property enables us to find the sampling distribution of a wide class of statistics which are expressible as sums, and particularly of the mean.

If we have a sample of n values from a population whose characteristic function is $\phi(t)$, the characteristic function of their sum is ϕ^n. Thus the distribution function of their sum z is given by $F(z)$ where (cf. (4.4) and (4.33))

$$F(z)-F(0) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{1-e^{-itz}}{it}\phi^n\,dt \qquad (11.53)$$

and the frequency function is

$$f(z) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itz}\phi^n\,dt. \qquad (11.54)$$

The following examples will illustrate the power of these results.

Example 11.10 Distribution of the mean for the binomial distribution

The characteristic function of the binomial distribution arrayed by $(q+p)^r$ is

$$(q+pe^{it})^r.$$

The c.f. of the sampling distribution of the sum of n values is then

$$(q+pe^{it})^{rn}$$

and that of the distribution of the mean ($1/n$ of that sum) is

$$(q+pe^{it/n})^{rn}.$$

But this is the c.f. of the binomial distribution arrayed by

$$(q+p)^{rn}, \tag{11.55}$$

the interval being $1/n$ instead of unity ; and hence this distribution is that of the mean.

Example 11.11 Distribution of the mean for the Poisson distribution

The characteristic function of the Poisson distribution whose general term is $e^{-\lambda}\lambda^r/r!$ is

$$\exp\{\lambda(e^{it}-1)\}.$$

The c.f. of the mean is then

$$\exp\{n\lambda(e^{it/n}-1)\}$$

and hence the distribution of the mean is the Poisson distribution whose general term is

$$e^{-n\lambda}\frac{(n\lambda)^r}{r!}, \tag{11.56}$$

the interval being $1/n$ instead of unity.

Example 11.12 Distribution of the mean for the normal population

The characteristic function of the normal distribution

$$dF = \frac{1}{\sigma\sqrt{(2\pi)}}\exp\left\{-\tfrac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}dx$$

is seen from Example 3.4 to be

$$\exp(-\tfrac{1}{2}t^2\sigma^2+it\mu).$$

The c.f. of the distribution of the mean of n values is then

$$\exp\left\{n\left(-\tfrac{1}{2}\frac{t^2\sigma^2}{n^2}+\frac{it\mu}{n}\right)\right\} = \exp\left\{-\tfrac{1}{2}\frac{t^2\sigma^2}{n}+it\mu\right\}. \tag{11.57}$$

This is the c.f. of a normal distribution with mean μ and variance σ^2/n, which is therefore the distribution required.

Example 11.13 Distribution of the mean for the Type III population

The characteristic function of the distribution

$$dF = \frac{1}{\Gamma(\gamma)}e^{-x/a}\left(\frac{x}{a}\right)^{\gamma-1}\frac{dx}{a}, \qquad a>0,$$

is (Example 3.6)

$$\phi(t) = \frac{1}{(1-ita)^{\gamma}}.$$

The c.f. of the distribution of the mean of n values is then

$$\frac{1}{\left(1-\frac{ita}{n}\right)^{n\gamma}}.$$

This is the c.f. of the distribution

$$dF = \frac{1}{\Gamma(\gamma n)} e^{-nx/a} \left(\frac{nx}{a}\right)^{n\gamma-1} \frac{n\,dx}{a}. \tag{11.58}$$

Example 11.14 Distribution of the mean for the rectangular population

The characteristic function of the distribution $dF = dx$ $(0 \leqslant x \leqslant 1)$ is

$$\int_0^1 e^{itx}\,dx = \frac{e^{it}-1}{it}.$$

The c.f. of the mean of n values is then $\left(\frac{e^{it/n}-1}{it/n}\right)^n$, and the frequency function is thus

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\left(\frac{e^{it/n}-1}{it/n}\right)^n dt. \tag{11.59}$$

This integral is everywhere analytic and the range of integration may be changed to the contour Γ consisting of the real axis from $-\infty$ to $-c$, the small semicircle of radius c and centre at the origin, and the real axis from c to ∞. Thus

$$\begin{aligned} f(x) &= \frac{1}{2\pi}\int_\Gamma e^{-itx}\left(\frac{e^{it/n}-1}{it/n}\right)^n dt \\ &= \frac{(-1)^n}{2\pi}\int_\Gamma e^{-itx} \sum_{j=0}^{n} (-1)^j \binom{n}{j} \frac{e^{itj/n}}{(it/n)^n}\,dt. \end{aligned} \tag{11.60}$$

Now

$$\begin{aligned} \int_\Gamma \frac{e^{igz}}{z^n}\,dz &= 0 \text{ if } g>0 \\ &= -2\pi i^n \frac{g^{n-1}}{(n-1)!} \text{ if } g \leqslant 0. \end{aligned}$$

This may be seen by integrating along a contour consisting of Γ and the infinite semicircle *above* the real axis if $g>0$ and *below* it if $g \leqslant 0$.

Substituting in (11.60) we find

$$\begin{aligned} f(x) &= \frac{(-1)^n}{2\pi}\int_\Gamma \Sigma (-1)^j \binom{n}{j} \frac{e^{itj/n-itx}}{(it/n)^n}\,dt \\ &= \frac{(-1)^{n-1}}{(n-1)!} \sum_{j<nx} (-1)^j \binom{n}{j} \frac{\left(\frac{j}{n}-x\right)^{n-1}}{\left(\frac{1}{n}\right)^n} \\ &= \frac{n^n}{(n-1)!} \sum_{j=0}^{[nx]} (-1)^j \binom{n}{j} \left(x-\frac{j}{n}\right)^{n-1} \end{aligned}$$

This, with a few changes of notation, is the same as (11.52).

11.7 General expressions may also be derived for the distributions of geometric means and the moments about fixed points.

In fact, if $y = \log x$, the characteristic function of y is

$$\alpha(t) = \int_{-\infty}^{\infty} e^{it\log x}\,dF = \int_{-\infty}^{\infty} x^{it}\,dF.$$

The distribution of the sum of n independent values of y, say nz, is then given by

$$F(nz)-F(0) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{1-e^{-itnz}}{it}\alpha^n\,dt, \tag{11.61}$$

and the distribution of the mean is that of z. But $z = \log u$, where u is the geometric mean, and hence the distribution of u may be found.

The frequency function, when it exists, is

$$f(nz) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itnz}\alpha^n\,dt.$$

Similarly the characteristic function of a power of the variate, say x^r, is given by

$$\beta(t) = \int_{-\infty}^{\infty} \exp(itx^r)\,dF$$

and thus the distribution of the rth moment, say z, by

$$F(nz)-F(0) = \int_{-\infty}^{\infty} \frac{e^{-itnz}-1}{it}\beta^n\,dt. \tag{11.62}$$

Example 11.15 Distribution of the geometric mean in samples from a rectangular population

If the population is

$$dF = dx/a, \qquad 0 \leqslant x \leqslant a,$$

the characteristic function of $\log x$ is

$$\int_0^a x^{it}\frac{dx}{a} = \frac{a^{it}}{1+it}.$$

The *frequency* function of $u = \Sigma \log x$ is then given by

$$\begin{aligned} f(u) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{-itu}a^{nit}}{(1+it)^n}\,dt \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{it(n\log a-u)}}{(1+it)^n}\,dt, \qquad n\log a-u \geqslant 0. \end{aligned}$$

This integral may be evaluated by contour integration and we find

$$f(u) = \frac{(n\log a-u)^{n-1}e^{-(n\log a-u)}}{\Gamma(n)},$$

whence, putting $z = e^{u/n}$, we find for the distribution of the geometric mean z

$$f(z) = \frac{n^n z^{n-1}}{a^n\Gamma(n)}\left(\log\frac{a}{z}\right)^{n-1}. \tag{11.63}$$

Example 11.16 Distribution of the second-order moment about the population mean in samples from a normal population

If the distribution is

$$dF = \frac{1}{\sigma\sqrt{(2\pi)}}\exp(-\tfrac{1}{2}x^2/\sigma^2)\,dx$$

the characteristic function of x^2 is

$$\frac{1}{\sigma\sqrt{(2\pi)}}\int_{-\infty}^{\infty}\exp(itx^2)\exp(-\tfrac{1}{2}x^2/\sigma^2)\,dx$$
$$=\frac{1}{(1-2\sigma^2 it)^{\frac{1}{2}}}.$$

The c.f. of the mean of n values, say m_2, is then

$$\frac{1}{(1-2\sigma^2 it/n)^{\frac{1}{2}n}} \tag{11.64}$$

and the frequency function of this is

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{\exp(-itm_2)}{(1-2\sigma^2 it/n)^{\frac{1}{2}n}}\,dt.$$

This may be integrated by a contour integration, or the result written down directly from the consideration that (11.64) is the characteristic function of the distribution

$$dF=\frac{n^{\frac{1}{2}n}}{(2\sigma^2)^{\frac{1}{2}n}\Gamma(\frac{1}{2}n)}\exp(-nm_2/2\sigma^2)\,m_2^{\frac{1}{2}n-1}\,dm_2, \tag{11.65}$$

a result which may be compared with that of (11.8), to which it reduces on writing $z=nm_2/\sigma^2$.

The method of induction

11.8 The distribution of the sum of two independent variates may be obtained directly as follows. Their joint distribution is

$$dF(x_1,x_2)=dF_1(x_1)\,dF_2(x_2)=f_1(x_1)f_2(x_2)\,dx_1\,dx_2.$$

The transformation $z=x_1+x_2$, $y=x_2$, has Jacobian equal to 1, and gives

$$dG(z,y)=f_1(z-y)f_2(y)\,dz\,dy. \tag{11.66}$$

Integrating out y, we obtain, for the marginal distribution of z,

$$dH(z)=\left\{\int_{-\infty}^{\infty}f_1(z-y)f_2(y)\,dy\right\}dz$$

or, for the frequency function of z,

$$h(z)=\int_{-\infty}^{\infty}f_1(z-y)f_2(y)\,dy. \tag{11.67}$$

The distribution function of z is obtained by integrating $h(t)$ from $-\infty$ to z, giving,

$$H(z)=\int_{-\infty}^{\infty}F_1(z-y)f_2(y)\,dy. \tag{11.68}$$

Note that if $x_1\geqslant 0$, y is to be integrated from $-\infty$ to z; if x_2 also is $\geqslant 0$, y is integrated from 0 to z.

(11.67) can be used to obtain successively the distribution of the sum of any number of variables whose individual distributions are known. If all the variables have the same distribution the general form may be suggested when the results for two or three variates have been worked out. Its correctness can then be verified by induction. The following examples illustrate the method.

Example 11.17

Consider again the distribution of Example 11.1,

$$dF = \frac{dx}{\pi(1+x^2)}, \qquad -\infty \leqslant x \leqslant \infty.$$

By (11.67) the distribution of the sum of two independent variables, each of which has this distribution, has the frequency function

$$\int_{-\infty}^{\infty} \frac{1}{\pi^2}\left(\frac{1}{1+(z-x)^2}\right)\left(\frac{1}{1+x^2}\right)dx = \frac{2}{\pi(z^2+2^2)}.$$

This suggests the general form

$$\frac{n}{\pi(z^2+n^2)}.$$

If this is correct, then the form for $(n+1)$ variables is

$$\begin{aligned}\int_{-\infty}^{\infty} \frac{n}{\pi^2}\left(\frac{1}{n^2+(z-x)^2}\right)\left(\frac{1}{1+x^2}\right)dx &= \frac{n}{\pi}\left(\frac{n+1}{n}\right)\Big/\{z^2+(n+1)^2\} \\ &= \frac{(n+1)}{\pi\{z^2+(n+1)^2\}}.\end{aligned}$$

The result holds for $n = 1, 2$, and is therefore true in general.

Example 11.18

In Example 11.6 we found that the distribution of the sum of squares of n independent normal variates is given by

$$dF = \frac{1}{2^{\frac{1}{2}n}\Gamma(\frac{1}{2}n)} e^{-\frac{1}{2}z} z^{\frac{1}{2}n-1}\, dz. \tag{11.69}$$

Suppose we had surmised this form from an examination of a few cases for low n. Let x be another variate distributed normally about zero mean with unit variance, and $x^2 = v$. Then v has the distribution

$$dF = \frac{1}{2^{\frac{1}{2}}\Gamma(\frac{1}{2})} e^{-\frac{1}{2}v} v^{-\frac{1}{2}}\, dv, \qquad 0 \leqslant v \leqslant \infty.$$

Then, from (11.67) the frequency function of the distribution of $u = z+v$ is given, since z and v are both non-negative, by

$$\begin{aligned}&\int_0^u \frac{1}{2^{\frac{1}{2}}\Gamma(\frac{1}{2})} e^{-\frac{1}{2}(u-y)}(u-y)^{-\frac{1}{2}} \frac{1}{2^{\frac{1}{2}n}\Gamma(\frac{1}{2}n)} e^{-\frac{1}{2}y} y^{\frac{1}{2}(n-2)}\, dy \\ &= \frac{e^{-\frac{1}{2}u}}{2^{\frac{1}{2}(n+1)}\Gamma(\frac{1}{2})\Gamma(\frac{1}{2}n)} \int_0^u (u-y)^{-\frac{1}{2}} y^{\frac{1}{2}(n-2)}\, dy \\ &= \frac{B(\frac{1}{2}, \frac{1}{2}n)}{2^{\frac{1}{2}(n+1)}\Gamma(\frac{1}{2})\Gamma(\frac{1}{2}n)} e^{-\frac{1}{2}u} u^{\frac{1}{2}(n-1)} \\ &= \frac{1}{2^{\frac{1}{2}(n+1)}\Gamma\{\frac{1}{2}(n+1)\}} e^{-\frac{1}{2}u} u^{\frac{1}{2}(n-1)}\end{aligned}$$

which is the same as (11.69) with $n+1$ for n. Hence the distribution holds generally.

The distribution of a ratio of random variables

11.9 Cases sometimes arise in which we require the sampling distribution of the ratio of two independent variates, x_1 and x_2. Their joint distribution is, as before,

$$dF(x_1, x_2) = dF_1(x_1)\,dF_2(x_2) = f_1(x_1)f_2(x_2)\,dx_1\,dx_2. \tag{11.70}$$

We now transform to new variables

$$u = x_1/x_2,$$
$$v = x_2,$$

and the Jacobian of the transformation is found to be

$$J = v. \tag{11.71}$$

Care is necessary here to ensure that the change of variable obeys the validity conditions outlined in **11.3**. Essentially, we must ensure that $J \neq 0$ within the domain of integration, although zero values on the boundary of the domain do not affect the validity of the transformation. From (11.71) we see that this implies

$$x_2 = v \neq 0 \tag{11.72}$$

except on the boundary. Thus we must in the first place restrict our discussion to ratios of random variables such that the *denominator* can range *either* from 0 to ∞, *or* from $-\infty$ to 0, but not both. In this case we have, from (11.70) and (11.71),

$$dG(u, v) = f_1(uv)f_2(v)\,v\,du\,dv, \tag{11.73}$$

and thus the marginal distribution of u is

$$dH(u) = \left\{\int_{-\infty}^{\infty} f_1(uv)f_2(v)\,v\,dv\right\}du, \tag{11.74}$$

and its distribution function

$$H(u) = \int_{-\infty}^{\infty} F_1(uv)f_2(v)\,dv. \tag{11.75}$$

Example 11.19

Consider again the distribution of the ratio $\bar{x}/s$ discussed in Example 11.8. Here $\bar{x}$ is the mean of samples of n from a normal population and is thus distributed as

$$dF_1 \propto \exp(n\bar{x}^2/2\sigma^2)\,d\bar{x}.$$

s is distributed as

$$dF_2 \propto \exp(-ns^2/2\sigma^2)\,s^{n-2}\,ds,$$

as we have seen at (11.39) and (11.40), where we had $\sigma^2 = 1$.

Then the distribution of z $(= \bar{x}/s)$ is, from (11.74), a constant times

$$\int_0^{\infty} \exp(-nz^2s^2/2\sigma^2)\exp(-ns^2/2\sigma^2)\,s^{n-2}.s.\,ds$$

$$\propto \frac{1}{(1+z^2)^{\frac{1}{2}n}},$$

which gives us (11.42) on evaluation of the constant.

11.10 The distribution of the ratio u of **11.9** can also be obtained in terms of characteristic functions if x_2 is non-negative with finite mean. From (11.75), we have for non-negative v

$$H(u) = \int_0^\infty F_1(uv)\,dF_2(v),$$

which on using (4.4) becomes, writing $\phi_1(t)$, $\phi_2(t)$ for the c.f.'s of x_1, x_2 respectively,

$$\begin{aligned} H(u) &= \int_0^\infty \left\{ F_1(0) + \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{1-e^{-ituv}}{it}\,\phi_1(t)\,dt \right\} dF_2(v) \\ &= F_1(0) + \frac{1}{2\pi i}\int_{-\infty}^{\infty} \frac{\phi_1(t)}{t}\left\{ \int_0^\infty (1-e^{-ituv})\,dF_2(v) \right\} dt \\ &= F_1(0) + \frac{1}{2\pi i}\int_{-\infty}^{\infty} \frac{\phi_1(t)}{t}\{1-\phi_2(-tu)\}\,dt. \end{aligned} \tag{11.76}$$

Differentiating with respect to u, we find that the f.f. of u, if it exists, is given by

$$h(u) = \frac{1}{2\pi i}\int_{-\infty}^{\infty} \phi_1(t)\phi_2'(-tu)\,dt, \tag{11.77}$$

provided that the integral converges, a result originally due to Cramér (1937). Daniels (1954) pointed out that the finiteness of the mean of x_2 is a sufficient condition for the integral on the right of (11.77) to converge absolutely for all u.

Geary (1944, *J. Roy. Statist. Soc.*, **107**, 56) generalized (11.77) to the case of non-independent x_1, x_2 with joint c.f. $\phi(t_1, t_2)$, where x_2 is non-negative with finite means as before. His result is

$$h(u) = \frac{1}{2\pi i}\int_{-\infty}^{\infty} \left[\frac{\partial\phi(t_1,t_2)}{\partial t_2} \right]_{t_2=-t_1u} dt_1. \tag{11.78}$$

A proof of (11.78) is sketched in Exercise 11.24 below, which indicates the importance of the condition that $E(x_2)$ be finite.

Example 11.20 Fisher's variance-ratio distribution

Suppose that we have two independent samples of n_1, n_2 members from normal populations with variances σ_1^2 and σ_2^2 respectively. We define the sample variances as of type $\Sigma(x-\bar{x})^2/n$ and call them s_1^2 and s_2^2. Their distributions are then, by (11.40),

$$dF_1 \propto \exp\left(-\frac{n_1 s_1^2}{2\sigma_1^2}\right) s_1^{n_1-3}\,ds_1^2,$$

$$dF_2 \propto \exp\left(-\frac{n_2 s_2^2}{2\sigma_2^2}\right) s_2^{n_2-3}\,ds_2^2.$$

The distribution of the ratio $t^2 = s_1^2/s_2^2$ is then, from (11.74), given by

$$\begin{aligned} f(t^2) &\propto \int_0^\infty \exp\left(-\frac{n_1 t^2 s_2^2}{2\sigma_1^2}\right)(s_2 t)^{n_1-3} \exp\left(-\frac{n_2 s_2^2}{2\sigma_2^2}\right) s_2^{n_2-3}.s_2^2.ds_2^2 \\ &\propto \int_0^\infty \exp\left\{-\frac{s_2^2}{2}\left(\frac{n_1 t^2}{\sigma_1^2}+\frac{n_2}{\sigma_2^2}\right)\right\} s_2^{n_1+n_2-4}\,t^{n_1-3}\,ds_2^2 \end{aligned}$$

$$\propto \frac{t^{n_1-3}}{\left(\frac{n_1 t^2}{\sigma_1^2}+\frac{n_2}{\sigma_2^2}\right)^{\frac{1}{2}(n_1+n_2-2)}}, \qquad 0 \leqslant t^2 \leqslant \infty. \tag{11.79}$$

This is a Beta-distribution of the second kind (cf. (6.17)). It is usually put in different form. Writing

$$z = \tfrac{1}{2}\log\frac{n_1(n_2-1)s_1^2/\sigma_1^2}{n_2(n_1-1)s_2^2/\sigma_2^2} = \tfrac{1}{2}\log\frac{n_1(n_2-1)}{n_2(n_1-1)}t^2\Big/\left(\frac{\sigma_1^2}{\sigma_2^2}\right), \tag{11.80}$$

we find for the frequency function of z

$$f(z) \propto \frac{e^{(n_1-1)z}}{\{(n_1-1)e^{2z}+(n_2-1)\}^{\frac{1}{2}(n_1+n_2-2)}}, \qquad -\infty \leqslant z \leqslant \infty, \tag{11.81}$$

or, writing $\nu_1 = n_1-1$ and $\nu_2 = n_2-1$ and evaluating the constant term,

$$f(z) = \frac{2\nu_1^{\frac{1}{2}\nu_1}\nu_2^{\frac{1}{2}\nu_2}}{B(\frac{1}{2}\nu_1, \frac{1}{2}\nu_2)}\frac{e^{\nu_1 z}}{(\nu_1 e^{2z}+\nu_2)^{\frac{1}{2}(\nu_1+\nu_2)}}. \tag{11.82}$$

This form was chosen by R. A. Fisher in preference to (11.79) because interpolation in tables of z is easier than in tables of t. The distribution is the basis of many tests of hypotheses, and at a later stage (Chapter 16) we shall study its properties at some length.

Let us obtain the distribution by Cramér's theorem on the c.f.'s of ratios. For the respective c.f.'s of s_1^2 and s_2^2 we have (cf. Examples 11.3 and 11.13)

$$\phi_1(t) = \left(1-\frac{2\sigma_1^2 it_1}{n_1}\right)^{-\frac{1}{2}(n_1-1)}$$

$$\phi_2(t) = \left(1-\frac{2\sigma_2^2 it_2}{n_2}\right)^{-\frac{1}{2}(n_2-1)}$$

Formula (11.77) then gives us for the ratio $v = \dfrac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2}$

$$f(v) \propto \frac{1}{2\pi i}\int_{-\infty}^{\infty}\left(1-\frac{2it}{n_1}\right)^{-\frac{1}{2}(n_1-1)}\left(1+\frac{2itv}{n_2}\right)^{-\frac{1}{2}(n_2+1)}dt.$$

A contour integration over the imaginary axis and the infinite semicircle to the left of it then gives us

$$f(v) \propto \left[\frac{d^\alpha}{dt^\alpha}\left(1-\frac{2tv}{n_2}\right)^{-\frac{1}{2}(n_2+1)}\right]_{t=-\frac{1}{2}n_1}$$

where $\alpha = \frac{1}{2}(n_1-3)$, giving

$$f(v) \propto \frac{v^{\frac{1}{2}(n_1-3)}}{\left(1+\frac{n_1 v}{n_2}\right)^{\frac{1}{2}(n_1+n_2-2)}}.$$

This, on the substitution $v = \dfrac{t^2}{\sigma_1^2/\sigma_2^2}$, reduces to the form (11.79). It will be noticed that this proof applies only for odd n_1, since α must be integral.

11.11 In **11.10** we referred to the problem of the distribution of a ratio when the denominator has a range overlapping zero. One way of handling the difficulty is to split the range into two parts at zero. We assume the denominator random variable

to be continuous at zero, since otherwise there is a finite probability that the ratio is infinite. We obtain from (11.70), separately for the two parts of the range,

$$dG(u, v)\begin{cases} = f_1(uv)f_2(v)\,|\,v\,|\,du\,dv, & v<0, \\ = f_1(uv)f_2(v)\,v\,du\,dv, & v\geqslant 0, \end{cases} \tag{11.83}$$

so that the analogue of (11.74) is

$$dH(u) = \left\{\int_0^\infty f_1(uv)f_2(v)\,v\,dv - \int_{-\infty}^0 f_1(uv)f_2(v)\,v\,dv\right\}du, \tag{11.84}$$

and the distribution function follows as before.

In practice, we are generally interested in a ratio whose denominator is non-negative, so that the methods of **11.10** apply, as in Examples 11.19 and 11.20. But even when the denominator can take values on either side of zero, the problem can sometimes be simplified by considering the square of the ratio, thus getting back to **11.10.**

Example 11.21

Consider the ratio of two independent normal variables with zero mean and unit variance. From (11.84), its frequency function is

$$dH(u) = \frac{du}{2\pi}\left\{\int_0^\infty e^{-\frac{1}{2}(uv)^2}e^{-\frac{1}{2}v^2}\,v\,dv - \int_{-\infty}^0 e^{-\frac{1}{2}(uv)^2}e^{-\frac{1}{2}v^2}\,v\,dv\right\}.$$

Simplifying, and making use of the symmetry of the normal distribution, this becomes

$$dH(u) = \frac{du}{\pi}\int_0^\infty \exp\{-\tfrac{1}{2}v^2(u^2+1)\}\,v\,dv.$$

This integral is immediately seen to be (cf. the mean deviation of a normal distribution at (5.26)) equal to $1/(1+u^2)$. Thus

$$dH(u) = \frac{du}{\pi(1+u^2)}, \qquad -\infty\leqslant u\leqslant\infty, \tag{11.85}$$

which is a Cauchy distribution.

Here, as suggested in **11.11**, consideration of the square of the ratio is fruitful, leading to the distribution of the ratio of two squares of standardized normal variables, since each of these is distributed in the Type III form of (11.8) with $n = 1$. Thus the distribution of the squared ratio is equivalent to Fisher's distribution of t in Example 11.20. It is easily confirmed that equation (11.79) reduces to our result above if we put $n_1 = n_2 = 2$,(*) $\sigma_1 = \sigma_2 = 1$.

11.12 Up to this point we have been mainly concerned with the distribution of a single statistic compiled from the members of a simple random sample. The methods may, however, readily be generalized to obtain the simultaneous distribution of several statistics. For example, if there are several statistics $z_1, z_2, \ldots, z_p$, and the joint distribution of the sample values $x_1, \ldots, x_n$ is represented by $dF(x_1, \ldots, x_n)$,

(*) Not 1, because we are squaring about the *population* means here, not the sample means, and thus have an extra degree of freedom. Cf. Example 11.7 on this point.

the characteristic function of the z's is given by

$$\phi(t_1,\ldots,t_p) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \exp(it_1z_1+\ldots+it_pz_p)\,dF(x_1,\ldots,x_n) \qquad (11.86)$$

and the frequency function of the z's (if it exists) by

$$f(z_1,\ldots,z_p) = \frac{1}{(2\pi)^p}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \exp(-it_1z_1-\ldots-it_pz_p)\,\phi(t_1,\ldots,t_p)\,dt_1\ldots dt_p. \qquad (11.87)$$

Examples of the use of these results will occur in the sequel.

Steepest descents

11.13 Although, in this chapter, we have been concerned with the derivation of exact results for sampling distributions, it is convenient to refer here to an approximative method developed by Daniels (1954), which differs radically from the methods to be discussed in Chapter 12. The frequency function of a mean is expressed as an inversion of the nth power of the characteristic function. The resultant integral is evaluated approximately by choosing a path of integration through a saddlepoint of the integrand in such a way that the integrand is negligible outside the immediate neighbourhood of the saddlepoint. This, in point of fact, can be done, and the approximation which results is, in simpler cases at least, remarkably good. The method results in expansions of the frequency function of the Edgeworth type (**6.18**) but is more powerful than the Edgeworth expansion itself.

EXERCISES

11.1 Find the distribution of means of samples of n independent observations from

$$dF = e^{-x}dx, \quad 0 \leqslant x \leqslant \infty,$$

(a) by induction; (b) by the use of characteristic functions.

11.2 Derive by the method of characteristic functions the expression for the sampling distribution of the mean of samples from the population

$$dF = \frac{dx}{\pi(1+x^2)}, \quad -\infty \leqslant x \leqslant \infty.$$

11.3 Show that the distribution of the geometric mean g in samples of n from the Type III population

$$dF = \frac{e^{-x}x^{p-1}}{\Gamma(p)}dx, \quad 0 \leqslant x \leqslant \infty,$$

is

$$dF = \frac{ng^{np-1}}{\Gamma(n)\{\Gamma(p)\}^n}\sum_{j=0}^{\infty}(-1)^{n+nj+1}\left[\frac{d^{n-1}}{dt^{n-1}}\{\Gamma(t+1)\}^n g^{nt}\right]_{t=j}$$

(Kullback, 1934)

11.4 $x_1, \ldots, x_n$ are n independent normal variates with zero mean and unit variance. Show that $x_i^2\Big/\left(\sum_{j=1}^{n} x_j^2\right)$ is distributed independently of $\sum_{j=1}^{n} x_j^2$. Show also that if $\bar{x}$ is the mean of the x's $(x_i-\bar{x})^2/\Sigma(x_j-\bar{x})^2$ is independent of $\Sigma(x_j-\bar{x})^2$.

11.5 Use the result of the previous Exercise to show that if m_2, m_3 and m_4 are the sample mean moments of order 2, 3 and 4 respectively,

$$E\left(\frac{m_3}{m_2^{3/2}}\right)^k = \frac{Em_3^k}{Em_2^{3k/2}},$$

$$E\left(\frac{m_4}{m_2^2}\right)^k = \frac{Em_4^k}{Em_2^{2k}}.$$

11.6 If $x_1, \ldots, x_n$ are independent variates distributed uniformly in the range 0 to 1, show that for

$$Q = \prod_{j=1}^{n} (1-x_j)$$

the distribution function is given by

$$P\{Q \leqslant e^{-\frac{1}{2}t}\} = \int_t^\infty z^{n-1} e^{-\frac{1}{2}z}\, dz / \{2^n (n-1)!\}.$$

11.7 Show that the joint distribution of two order-statistics x_r, x_s $(r<s)$ is given by

$$dF(x_r, x_s) = \frac{n!}{(r-1)!\,(s-r-1)!\,(n-s)!} F_1^{r-1} (F_2-F_1)^{s-r-1} (1-F_2)^{n-s}\, dF_1\, dF_2$$

where F_1 is the distribution function of x with $x = x_r$ and F_2 is the corresponding function with $x = x_s$.

11.8 In the previous Exercise, integrate out x_s to obtain the distribution of x_r.

11.9 Show that the distribution of the geometric mean of n variables, one from each of the populations with frequency functions

$$\frac{x^{p-1}e^{-x}}{\Gamma(p)}, \frac{x^{p+1/n-1}e^{-x}}{\Gamma(p+1/n)}, \ldots, \frac{x^{p+(n-1)/n-1}e^{-x}}{\Gamma\{p+(n-1)/n\}}$$

is the same as the distribution of the arithmetic mean of n independent variables distributed in the first of these forms.

(Kullback, 1934)

11.10 If a frequency function is given as the sum of a number of terms of the Type A series

$$f(x) = \alpha(x)\left\{1 + \frac{a_3}{\sigma^3} H_3\left(\frac{x}{\sigma}\right) + \ldots + \frac{a_k}{\sigma^k} H_k\left(\frac{x}{\sigma}\right)\right\}$$

show that the sum S of n independent variates has a frequency function

$$f(S) = \alpha(S)\left\{1 + \frac{A_3}{\Sigma^3} H_3\left(\frac{S}{\Sigma}\right) + \ldots + \frac{A_k}{\Sigma^k} H_k\left(\frac{S}{\Sigma}\right)\right\}$$

where $\Sigma = \sigma\sqrt{n}$ and

$$A = \sum \frac{n!}{\nu_3!\,\nu_4! \ldots \nu_k!\,(N-\nu_3-\ldots-\nu_k)!} a_3^{\nu_3} \ldots a_k^{\nu_k},$$

the summation being taken over all values of the ν's for which

$$3\nu_3 + 4\nu_4 + \ldots + k\nu_k = j.$$

(Baker, 1930)

11.11 Show that the ratio of two independent normal variables has frequency function

$$f(v) = \frac{1}{\sqrt{(2\pi)}} \frac{m_2\sigma_1^2 + m_1\sigma_2^2 v}{(\sigma_1^2 + \sigma_2^2 v^2)^{\frac{3}{2}}} \exp\left\{-\tfrac{1}{2}\frac{(m_1 - m_2 v)^2}{\sigma_1^2 + \sigma_2^2 v^2}\right\}$$

where m_1, σ_1 are the mean and standard deviation of the first variate, m_2, σ_2 those of the second variate, and it is assumed that m_2 is so large compared with σ_2 that the range of the second variate is effectively positive.

Hence show that $\dfrac{m_1 - m_2 v}{(\sigma_1^2 + \sigma_2^2 v^2)^{\frac{1}{2}}}$ is normally distributed about zero mean with unit variance.

(Geary, 1930)

11.12 Show that the ratio of two independent variables distributed as

$$dF \propto e^{-\gamma_1(x-m_1)}(x-m_1)^{p_1-1}\,dx, \quad 0 < m_1 \leqslant x \leqslant \infty,$$
$$dF \propto e^{-\gamma_2(x-m_2)}(x-m_2)^{p_2-1}\,dx, \quad 0 < m_2 \leqslant x \leqslant \infty,$$

has a frequency function

$$f(v) = \frac{\gamma_1^{p_1} e^{\xi\gamma_1} m_2}{(p_1-1)!}\left\{\frac{\xi^{p_1-1}}{\left(1+\frac{\gamma_1 v}{\gamma_2}\right)^{p_2}} - \binom{p_1-1}{1}\left(\frac{v}{\gamma_2}\right)\frac{p_2\,\xi^{p_1-2}}{\left(1+\frac{\gamma_1 v}{\gamma_2}\right)^{p_2+1}} + \right.$$
$$\left. + \binom{p_1-2}{2}\left(\frac{v^2}{\gamma_2^2}\right)\frac{p_2(p_2+1)\,\xi^{p_1-3}}{\left(1+\frac{\gamma_1 v}{\gamma_2}\right)^{p_2+2}} - \dots\right\}$$
$$+ \frac{\gamma_1^{p_1} e^{\xi\gamma_1} p_2}{(p_1-1)!\,\gamma_2}\left\{\frac{\xi^{p_1-1}}{\left(1+\frac{\gamma_1 v}{\gamma_2}\right)^{p_2+1}} - \binom{p_1-1}{1}\left(\frac{v}{\gamma_2}\right)\frac{(p_2+1)\,\xi^{p_1-2}}{\left(1+\frac{\gamma_1 v}{\gamma_2}\right)^{p_2+2}} + \dots\right\}$$

where $\xi = m_1 - m_2 v$. (This includes Fisher's variance-ratio distribution as a particular case.)

11.13 Show that the ratio of two independent variates $v = x_1/x_2$, where x_1 is distributed normally with mean m_1 and variance σ^2 and x_2 like a standard deviation in normal samples, i.e. with distribution function given by

$$dF \propto e^{-\gamma(s-m_2)}(s-m_2)^{p-1}\,ds, \quad 0 \leqslant m_2 \leqslant s \leqslant \infty,$$

has a frequency function given by

$$f(v) = \frac{\gamma^p e^{-\frac{\xi^2}{2\sigma^2}}}{\sigma\sqrt{(2\pi)}\,\Gamma(p)}\left\{m_2 \sum_{j=0}^{\infty} \frac{v^j \xi^j \Gamma\left(p+\frac{j}{2}\right)}{j!\,\sigma^{2j}\left(\gamma + \frac{1}{2}\frac{v^2}{\sigma^2}\right)^{p+\frac{j}{2}}} + \sum_{j=0}^{\infty} \frac{v^j \xi^j \Gamma\left(p+\frac{j+1}{2}\right)}{j!\,\sigma^{2j}\left(\gamma + \frac{1}{2}\frac{v^2}{\sigma^2}\right)^{p+\frac{j+1}{2}}}\right\}$$

where $\xi = m_1 - m_2 v$.

(This includes " Student's " distribution as a particular case.)

11.14 Show that the distribution of the mean of samples of n from the Type II population

$$dF \propto x^{p-1}(1-x)^{p-1}\,dx, \quad p > 0, 0 \leqslant x \leqslant 1,$$

is given by

$$f(\bar{x}) = \tfrac{1}{2} n\,\pi^{\frac{1}{2}n-1}\{\Gamma(p)\}^n \int_{-\infty}^{\infty} \{\mathcal{J}_{p-\frac{1}{2}}(\tfrac{1}{2}t)\}^n \cos(n\bar{x}t)\,dt/t^{n(p-\frac{1}{2})}\}$$

where $\mathcal{J}_r(z)$ is the Bessel coefficient of order r in z.

(Irwin, 1927)

11.15 Show that the difference of two independent variates, z, each of which is distributed in the Type III form

$$dF = \frac{e^{-x}x^{p-1}}{\Gamma(p)}\,dx, \quad p > 0, \ 0 \leqslant x \leqslant \infty,$$

has the frequency function

$$f(z) = \frac{z^{p-\frac{1}{2}} K_{p-\frac{1}{2}}(z)}{2^{p-\frac{1}{2}} \Gamma(p) \Gamma(\frac{1}{2})},$$

where $K_r(x)$ is the Bessel function of second order and imaginary argument.

(K. Pearson *et al.*, 1932)

11.16 Show that the difference of two values drawn at random from the Poisson population whose general term is $e^{-\lambda} \lambda^r / r!$ is distributed in the form whose general term is $e^{-2\lambda} T_d(2\lambda)$, where d can take all integral values from $-\infty$ to ∞ and $T_d(2\lambda)$ is Bessel's modified function of the first kind of order d and argument 2λ. (Cf. Example 4.5.)

(Irwin, 1937)

11.17 Use Exercise 11.15 to find the distribution of the mean of n independent variates, each distributed in the form

$$dF = \tfrac{1}{2} e^{-|x|} dx, \quad -\infty \leqslant x \leqslant \infty.$$

11.18 Derive the joint characteristic function of mean and variance in samples of n observations from a normal distribution, and show that it factorizes into their marginal c.f.'s, and that therefore they are independently distributed.

11.19 If the c.f. of sample mean and variance in samples of n is $\phi(t_1, t_2)$ and that of the frequency distribution $f(x)$ is $\alpha(t)$, show that

$$\left[\frac{\partial}{\partial t_2} \phi(t_1, t_2)\right]_{t_2=0} = \frac{n-1}{n} i \left[\alpha^{n-1}(t_1/n) \int_{-\infty}^{\infty} x^2 e^{it_1 x/n} f(x)\, dx - \alpha^{n-2}(t_1/n) \left\{\int_{-\infty}^{\infty} x\, e^{it_1 x/n} f(x)\, dx\right\}^2\right].$$

Hence, if mean and variance are independent, show that

$$-\alpha \frac{d^2\alpha}{dt^2} + \left(\frac{d\alpha}{dt}\right)^2 = \sigma^2 \alpha^2$$

where σ^2 is the population variance. Hence show that the distribution must be normal.

[Lukacs (1942). Geary (1936) established the result for distributions possessing finite cumulants of all orders]

11.20 Samples of size m and n are drawn from a rectangular distribution $dF = dx/\theta$, $0 \leqslant x \leqslant \theta$ and their largest members are x_m and x_n respectively. Show that the distribution of x_m is

$$dF(x_m) = \{m x_m^{m-1}/\theta^m\}\, dx_m,$$

and hence that the distribution of the ratio x_m/x_n is given by

$$dG(u) = \frac{mn}{m+n} u^{m-1}\, du, \quad 0 \leqslant u \leqslant 1,$$

$$= \frac{mn}{m+n} \frac{du}{u^{n+1}}, \quad 1 \leqslant u \leqslant \infty.$$

(Murty, 1955)

11.21 k independent random samples, each of size n, are drawn from the rectangular parent $dF = dx$, $0 \leqslant x \leqslant 1$. The largest value in the rth sample is x_r. Show, by induction or otherwise, that the distribution of the product

$$v = x_1 x_2 \ldots x_k$$

is given by

$$dG(v) = \{n^k/(k-1)!\}\, v^{n-1} \{\log(1/v)\}^{k-1}\, dv.$$

(Rider, 1955)

11.22 Generalizing Example 11.21, show that if x and y are bivariate normally distributed with correlation ρ, the statistic

$$z = \frac{x-\mu_x}{\sigma_x} \Big/ \frac{y-\mu_y}{\sigma_y}$$

has the frequency function

$$g(z) = \frac{(1-\rho^2)^{\frac{1}{2}}\,dz}{\pi(1-2\rho z+z^2)}, \qquad -\infty \leqslant z \leqslant \infty.$$

(Fieller, 1932, *Biometrika*, **24**, 428)

11.23 In Example 11.21, show that if instead of normal variates we consider two independent variates distributed as in Exercise 2.12, their ratio is still distributed in the Cauchy form (11.85).

(Steck (1958), *Ann. Math. Statist.*, **29**, 604)

11.24 If x_1, x_2 ($x_2 \geqslant 0$, $E(x_2) = \mu$) have the joint f.f. $f(x_1, x_2)$ and c.f. $\phi(t_1, t_2)$, show that the f.f. of $u = x_1/x_2$ is $h(u) = \int_0^\infty f(uv, v)\,v\,dv$, generalizing (11.74). Now consider the joint f.f.

$$g(x_1, x_2) = \frac{x_2}{\mu} f(x_1, x_2).$$

Show that its c.f. is

$$\psi(t_1, t_2) = \frac{1}{i\mu} \frac{\partial \phi(t_1, t_2)}{\partial t_2},$$

and that the c.f. of the variate $w = x_1 - ux_2$ is $[\psi(t_1, t_2)]_{t_2 = -ut_1}$. By showing that the f.f. of w at $w = 0$ is equal to $h(u)/\mu$, use the Inversion Theorem to establish (11.78).

(Daniels, 1954)

11.25 If x is a standardized normal variate, show that the c.f. of $y = \frac{1}{x^2}$ is

$$\phi_y(t) = \exp\{-(-2it)^{\frac{1}{2}}\}$$

and hence that the c.f. of the mean $\bar{y}$ of a random sample of n y's is

$$\phi_{\bar{y}}(t) = \phi_y(nt) = \phi_{ny}(t),$$

so that the distribution of the mean of n observations is a multiple n of that of a single observation, becoming more widely dispersed as n increases.

CHAPTER 12

APPROXIMATIONS TO SAMPLING DISTRIBUTIONS—(1)

12.1 In the previous chapter we have considered methods of deriving sampling distributions in an exact form when the parent population is completely specified. Those methods are not applicable when the population is not completely known, and they may in any case lead to results which are difficult to apply in practice, e.g. by yielding an integral which has not been tabulated. In such cases we can frequently deal with the problem by finding an approximate form for the sampling distribution, particularly by ascertaining its lower moments, and then fitting a tractable type of curve such as one of the Pearson system.

A procedure of this kind has, in fact, already been considered in Chapter 10, wherein it was seen that approximate expressions could be derived for the first and second moments of sampling distributions in terms of the low moments of the parent. When the sampling distribution tends to normality this, in effect, solves our problem, for its first and second moments determine a normal distribution. The methods of this chapter are really developments of this idea. We shall discuss exact methods of finding the moments of sampling distributions in terms of population moments. Our results are important not only on their own account, but in giving an accurate method of judging the degree of approximation of the expressions for large n discussed in Chapter 10. In particular, we shall be able to take up some points which had to be left on one side in that chapter—e.g. the rapidity with which some functions of the moments such as $\sqrt{b_1}$ approach normality.

12.2 The statistics with whose distributions we are customarily concerned may be usefully classified into three groups. The largest group comprises statistics which are symmetric functions of the observations; that is to say, if the observations are $x_1, x_2, \ldots, x_n$, the statistic depends explicitly on every x and its value is unchanged if we interchange any two x's. Moreover, our statistics are nearly always algebraic. This class includes the arithmetic mean, moments, quantities such as b_1 and b_2 which are derived from moments and, in the bivariate domain, product-moments. The second group comprises statistics based on order-properties of the sample values—the median, the quantiles, the range, extreme values and so forth. The third group is best defined as including those statistics, such as the mode, which are not included in the first two groups.

The methods we shall develop in this and the following chapters apply to the first group only. We may recall that there are three different types of moment concerned in the investigation: (a) the moments of the parent population, (b) the moments of the sample and (c) the moments of the sampling distribution. They will be referred to as parent moments, sample moments (moment-statistics) and sampling moments respectively. Similarly we shall consider parent cumulants, sample cumulants and sampling cumulants.

12.3 In Chapter 10 we obtained the exact results

$$\left.\begin{aligned} E m_r' &= \mu_r' \\ \operatorname{var} m_r' &= E(m_r'-\mu_r')^2 = \frac{1}{n}(\mu_{2r}'-\mu_r'^2) \end{aligned}\right\} \tag{12.1}$$

and noted that formulae for sampling moments about the mean were more difficult to obtain. Although we shall later reject this approach in favour of another, it is instructive to consider what happens if we try to generalize the procedure of that chapter to our present problem. Suppose, for example, that we are interested in the sampling distribution of the variance. The above equations give us the first two sampling moments of the second moment about an arbitrary point. For the first sampling moment of the variance we have

$$\begin{aligned} E m_2 &= E\{\Sigma x^2/n-(\Sigma x/n)^2\} \\ &= E\left\{\frac{n-1}{n^2}\Sigma x^2-\frac{1}{n^2}\Sigma x_i x_j\right\}, \qquad i \neq j \\ &= \frac{n-1}{n}E x^2-\frac{n-1}{n}E x_i x_j. \end{aligned} \tag{12.2}$$

Since x_i and x_j are independent, $E x_i x_j$ is equal to $\mu_1'^2$. Equation (12.2) then gives us

$$E m_2 = \frac{n-1}{n}(\mu_2'-\mu_1^2) = \frac{n-1}{n}\mu_2. \tag{12.3}$$

This may be compared with the approximate expression given by the methods of Chapter 10, namely

$$E m_2 = \mu_2. \tag{12.4}$$

12.4 It will be clear that the same method can be used to derive the sampling moment of any order of a statistic which can be expressed as a symmetric function (rational and integral) of the observations. For instance, to find the fourth moment of the sample variance we expand $\{\Sigma x^2/n-(\Sigma x/n)^2\}^4$ in terms of sums of products of type $\Sigma x_i^\alpha x_j^\beta \ldots x_k^\gamma$ and, on taking expectations, replace them by $n(n-1)\ldots(n-t+1)\mu_\alpha'\mu_\beta'\ldots\mu_\gamma'$ where t is the number of different suffixes $i, j \ldots k$ in the sum. This gives us the desired sampling moment in terms of parent moments.

12.5 The method is straightforward enough. Its execution, however, leads to some tedious algebra and some cumbrous expressions, except in very simple cases. We shall therefore find it convenient to systematize the working in terms of the notation of symmetric functions.

Suppose that we have a set of x's, $x_1, x_2, \ldots, x_n$. When we write an expression such as $\Sigma x_i^2 x_j x_l^3 x_k$ we shall suppose that all the suffixes are different and that, subject to this, the summation takes place over all values of x. Thus, in the example just given, there are $n(n-1)(n-2)(n-3)$ terms in the summation. It is to be noted that in an expression such as $\Sigma x_i x_j$ every pair occurs twice, for example x_1x_2 and x_2x_1; whereas in $\Sigma x_i^2 x_j$, $x_1^2x_2$ and $x_2^2x_1$ occur, but each does not occur twice.

The *augmented* symmetric functions are defined by

$$[p_1^{\pi_1} p_2^{\pi_2} \dots p_s^{\pi_s}] = \Sigma x_i^{p_1} x_j^{p_1} \dots x_q^{p_2} x_r^{p_2} \dots x_u^{p_s} x_v^{p_s} \tag{12.5}$$

where there are π_1 powers p_1, π_2 powers p_2, and so on on the right of (12.5). For example

$$\Sigma x_i^2 x_j x_l^3 x_k = [1^2 2\,3]\,; \quad \Sigma x_i^2 x_j^2 x_k^2 = [2^3].$$

More usual functions are the *monomial* symmetric functions defined by

$$(p_1^{\pi_1} p_2^{\pi_2} \dots p_s^{\pi_s}) = [p_1^{\pi_1} p_2^{\pi_2} \dots p_s^{\pi_s}] / \{\pi_1! \pi_2! \dots \pi_s!\}. \tag{12.6}$$

Two particular cases are of special importance: the *unitary* functions

$$a_r = (1^r) = \Sigma x_i x_j \dots x_l / r! \tag{12.7}$$

and the one-part functions or power-sums

$$s_r = (r) = \Sigma x_i^r = [r]. \tag{12.8}$$

Tables exist giving these functions in terms of one another. The most useful are the tables giving the power-sums in terms of the augmented symmetrics and vice versa (David and Kendall, 1949). From (12.5) we have the fundamental result

$$\mathrm{E}[p_1^{\pi_1} p_2^{\pi_2} \dots p_s^{\pi_s}] = n(n-1) \dots (n-\rho+1)(\mu'_{p_1})^{\pi_1} (\mu'_{p_2})^{\pi_2} \dots (\mu'_{p_s})^{\pi_s} \tag{12.9}$$

where $\rho = \sum_{i=1}^{s} \pi_i$ and $p = \sum_{i=1}^{s} p_i \pi_i$ is the *weight* of the symmetric function. Appendix Table 10 gives the relationships up to weight 6.

Example 12.1

For the sample variance we have, in terms of the power sums,

$$m_2 = \frac{(2)}{n} - \frac{(1)^2}{n^2}.$$

From Appendix Table 10 or directly we have

$$(2) = [2]\,; \quad (1)^2 = [2] + [1^2].$$

Hence

$$\begin{aligned} m_2 &= \frac{[2]}{n} - \frac{[2]+[1^2]}{n^2} \\ &= \frac{n-1}{n^2}[2] - \frac{1}{n^2}[1^2]. \end{aligned} \tag{12.10}$$

Taking expectations and using (12.9) we have

$$\begin{aligned} \mathrm{E}\, m_2 &= \frac{n-1}{n^2} n\mu'_2 - \frac{1}{n^2} n(n-1)\mu'^2_1 \\ &= \frac{n-1}{n} \mu_2. \end{aligned} \tag{12.11}$$

Let us note here a very useful abbreviation of the working. The statistic m_2 is independent of the origin of calculation and hence its sampling moments cannot depend on μ'_1. Without loss of generality, therefore, we can take this parent mean to be zero, the other moments then becoming moments about the mean. This implies that we

can ignore any augmented symmetric function containing a unit. Thus from (12.10) we have at once

$$\mathrm{E}\, m_2 = \frac{n-1}{n}\mu_2.$$

In a similar way we find

$$m_2^2 = \frac{(2)^2}{n^2} - \frac{2(2)(1)^2}{n^3} + \frac{(1)^4}{n^4}. \tag{12.12}$$

From the tables,

$$\begin{aligned}(2)^2 &= [4]+[2^2]\\ (2)(1)^2 &= [4]+2[31]+[2^2]+[21^2]\\ (1)^4 &= [4]+4[31]+3[2^2]+6[21^2]+[1^4].\end{aligned}$$

We can ignore those augmented symmetrics containing a unit, and on substitution in (12.12) we find

$$\begin{aligned}m_2^2 &= \frac{[4]+[2^2]}{n^2} - 2\frac{[4]+[2^2]}{n^3} + \frac{[4]+3[2^2]}{n^4}\\ &= \frac{(n-1)^2}{n^4}[4] + \frac{n^2-2n+3}{n^4}[2^2],\end{aligned}$$

whence, immediately,

$$\mathrm{E}\, m_2^2 = \frac{(n-1)^2}{n^3}\mu_4 + \frac{(n-1)(n^2-2n+3)}{n^3}\mu_2^2. \tag{12.13}$$

We then find, using (12.11),

$$\begin{aligned}\mathrm{var}\, m_2 &= \mathrm{E}\, m_2^2 - (\mathrm{E}\, m_2)^2\\ &= \frac{(n-1)^2}{n^3}\mu_4 - \frac{(n-1)(n-3)}{n^3}\mu_2^2.\end{aligned} \tag{12.14}$$

For large n this becomes approximately

$$\mathrm{var}\, m_2 = \frac{\mu_4 - \mu_2^2}{n},$$

confirming the result we reached in (10.9) with $r = 2$.

If, in (12.14), we put $\kappa_4 = \mu_4 - 3\mu_2^2$, $\kappa_2 = \mu_2$ we find

$$\mathrm{var}\, m_2 = \left(\frac{n-1}{n}\right)^2 \left\{\frac{\kappa_4}{n} + \frac{2\kappa_2^2}{n-1}\right\}. \tag{12.15}$$

k-statistics

12.6 The algebraic complexity of the results obtained by this straightforward approach and the amount of work required to reach them, especially before the tables of symmetric functions were available, led to a search for simpler methods. In 1928 Sir Ronald Fisher revolutionized the subject in two ways : by proposing new symmetric functions to characterize the distributions, the so-called *k*-statistics ; and by showing how their sampling cumulants could be obtained by combinatorial methods.

We consider a family of statistics $k_1, k_2, \ldots, k_p, \ldots$ which are symmetric functions

of the observations and are such that the mean value of k_p is the pth cumulant κ_p :—

$$\mathrm{E}\,k_p = \kappa_p. \tag{12.16}$$

There is a possible source of confusion here due to notation. Whereas the moment-statistic m_p is the same function of the sample values as μ_p is of the parent values, the same relation does not hold between k_p and κ_p. Conversely, $\mathrm{E}\,m_p$ is not equal to μ_p.

12.7 Note first of all that k_p is uniquely determined by this definition; for if there were two functions k_p and k'_p obeying (12.16) their difference $k_p - k'_p$ would have a zero mean value. But this difference is itself a symmetric function and can therefore be expressed as the sum of terms Σx^p, $\Sigma x_j x_k^{p-1}$, etc., and hence its mean value is a series of terms each of which is a product of moments. The vanishing of this series would imply a relationship among the moments, which is impossible except perhaps for particular parent populations. Hence $k_p - k'_p$ must vanish identically and thus $k_p = k'_p$.

Secondly, note that the k's are, like the mean-moments, independent of the origin of measurement, except for k_1, which is equal to the mean itself. In fact, we have by Taylor's theorem

$$k_p(x_1+h, x_2+h, \ldots, x_n+h)$$
$$= k_p(x_1, x_2, \ldots, x_n) + \frac{h}{1!} D k_p(x_1, x_2, \ldots, x_n) + \frac{h^2}{2!} D^2 k_p(x_1, x_2, \ldots, x_n) + \ldots \tag{12.17}$$

where

$$D = \frac{\partial}{\partial x_1} + \frac{\partial}{\partial x_2} + \ldots + \frac{\partial}{\partial x_n}.$$

Taking mean values, and remembering that κ_p itself is independent of the origin, except for κ_1, we have

$$\kappa_p = \kappa_p + \frac{h}{1!} \mathrm{E}(D k_p) + \ldots \tag{12.18}$$

Thus $\mathrm{E}(D k_p)$ and other terms on the right vanish separately, for (12.18) is an identity in h. In virtue of the remark above, this implies that $D k_p = 0$, $D^2 k_p = 0$, and so on; and hence, from (12.17),

$$k_p(x_1+h, x_2+h, \ldots, x_n+h) = k_p(x_1, x_2, \ldots, x_n),$$

i.e. k_p is independent of h. The exception to this rule is k_1 which has as its mean value $\kappa_1 = \mu'_1$ and thus

$$k_1 = \frac{1}{n} \Sigma x. \tag{12.19}$$

12.8 We now proceed to find explicit expressions for the k-statistics in terms of the observations $x_1, \ldots, x_n$. By definition k_p is of degree p in these observations (for κ_p is of order p in the moments, that is, the sum of the orders of the moments comprising any term in κ_p is p). We may then write

$$k_p = \Sigma\Sigma (x_1^{p_1} x_2^{p_1} \ldots x_{\pi_1}^{p_1} x_{\pi_1+1}^{p_2} \ldots x_{\pi_1+\pi_2}^{p_2} \ldots x_{\pi_1+\ldots+\pi_s}^{p_s}) A(p_1^{\pi_1} \ldots p_s^{\pi_s}) \tag{12.20}$$

where the second summation extends over all the ways of assigning the $\pi_1 + \pi_2 + \ldots + \pi_s$

subscripts (including permutations) from the n available and the first summation extends over all partitions of the number p, $(p_1^{\pi_1} p_2^{\pi_2} \dots p_s^{\pi_s})$. $A(p_1^{\pi_1} \dots p_s^{\pi_s})$ is a number depending on the partition.

We have
$$p_1\pi_1 + p_2\pi_2 + \dots + p_s\pi_s = p \tag{12.21}$$
and define ρ by
$$\pi_1 + \pi_2 + \dots + \pi_s = \rho. \tag{12.22}$$

We assume $n \geqslant p$. On taking mean values of (12.20) we have, since the x's are independent,
$$\kappa_p = \Sigma \{(\mu_{p_1}^{\pi_1} \mu_{p_2}^{\pi_2} \dots \mu_{p_s}^{\pi_s}) AB\}, \tag{12.23}$$
where B is the number of ways of picking out the ρ subscripts from n, permutations allowed, and is therefore equal to $n(n-1)\dots(n-\rho+1) = n^{[\rho]}$.

Now from equation (3.39), we have
$$\kappa_p = p! \Sigma\Sigma \left(\frac{\mu_{p_1}}{p_1!}\right)^{\pi_1} \dots \left(\frac{\mu_{p_s}}{p_s!}\right)^{\pi_s} \frac{(-1)^{\rho-1}(\rho-1)!}{\pi_1! \dots \pi_s!}, \tag{12.24}$$
the summation extending over all partitions subject to (12.21) and (12.22). On identifying corresponding terms in (12.23) and (12.24) we find the values of the A's and on substituting in (12.20) obtain finally
$$k_p = \frac{p! \Sigma (-1)^{\rho-1}(\rho-1)!}{n^{[\rho]}} \Sigma \frac{x_1^{p_1} \dots x_\rho^{p_s}}{(p_1!)^{\pi_1} \dots (p_s!)^{\pi_s} \pi_1! \dots \pi_s!}, \tag{12.25}$$
the explicit expression of k_p in terms of the x's.

We may notice an important simplification of this expression which is crucial in a discussion of the sampling properties of the k's. Apart from factors in ρ and n a typical term in (12.25) may be written
$$p! \left(\frac{x_1^{p_1}}{p_1!} \frac{x_2^{p_1}}{p_1!} \dots \frac{x_{\pi_1}^{p_1}}{p_1!} \dots \frac{x_\rho^{p_s}}{p_s!}\right) \cdot \frac{1}{\pi_1! \dots \pi_s!}$$
where, it is to be remembered, permutations of the subscripts are allowed. There will be a term of this type corresponding to every partition of ρ into π's and of p into p's. Consequently we may write
$$k_p = \Sigma \frac{(-1)^{\rho-1}(\rho-1)!}{n^{[\rho]}} \Sigma (x_{\gamma_1} x_{\gamma_2} \dots x_{\gamma_p}), \tag{12.26}$$
where there is a term in the second summation corresponding to every possible way of assigning the subscripts. In this assignment, subscripts are regarded as distinct entities. For example, if from the n subscripts we choose p_1 to be 1, p_1 to be 2, ... p_2 to be π_1+1, and so on, there will be as many different terms as there are ways of choosing p_1 from the 1's, and so on, i.e.
$$\frac{p!}{(p_1!)^{\pi_1} \dots (p_s!)^{\pi_s} \pi_1! \dots \pi_s!}. \tag{12.27}$$

In fact, (12.25) is a condensed form of (12.26) in which all the terms leading to the same x-product are added together, their number being given by (12.27).

Expression of *k*-statistics in terms of symmetric products and sums

12.9 We can write down the k's in terms of the augmented symmetric functions at once from the expressions for cumulants in terms of moments. For example, we have

$$\kappa_3 = \mu'_3 - 3\mu'_2\mu'_1 + 2\mu'^3_1.$$

Hence, comparing (12.24) and (12.25), we have

$$k_3 = \frac{[3]}{n} - \frac{3\,[21]}{n(n-1)} + \frac{2\,[1^3]}{n(n-1)(n-2)}.$$

Substituting for the augmented symmetrics in terms of power sums,

$$[3] = (3), [21] = -(3) + (2)(1), [1^3] = 2(3) - 3(2)(1) + (1)^3$$

we find

$$k_3 = \frac{1}{n(n-1)(n-2)}\{n^2(3) - 3n(2)(1) + 2(1)^3\}.$$

Writing $s_r = (r)$ we may put this in the form

$$k_3 = \frac{1}{n(n-1)(n-2)}(n^2 s_3 - 3n\,s_2 s_1 + 2s_1^3).$$

12.10 The first eight k-statistics in terms of the power sums are as follows :—

$$\left.\begin{aligned}
k_1 &= \frac{1}{n}s_1 \\
k_2 &= \frac{1}{n^{[2]}}(ns_2 - s_1^2) \\
k_3 &= \frac{1}{n^{[3]}}(n^2 s_3 - 3n\,s_2 s_1 + 2s_1^3) \\
k_4 &= \frac{1}{n^{[4]}}\{(n^3+n^2)s_4 - 4(n^2+n)s_3 s_1 - 3(n^2-n)s_2^2 + 12n\,s_2 s_1^2 - 6s_1^4\} \\
k_5 &= \frac{1}{n^{[5]}}\{(n^4+5n^3)s_5 - 5(n^3+5n^2)s_4 s_1 - 10(n^3-n^2)s_3 s_2 + 20(n^2+2n)s_3 s_1^2 \\
&\qquad + 30(n^2-n)s_2^2 s_1 - 60n\,s_2 s_1^3 + 24s_1^5\} \\
k_6 &= \frac{1}{n^{[6]}}\{(n^5+16n^4+11n^3-4n^2)s_6 - 6(n^4+16n^3+11n^2-4n)s_5 s_1 \\
&\qquad - 15n(n-1)^2(n+4)s_4 s_2 - 10(n^4-2n^3+5n^2-4n)s_3^2 \\
&\qquad + 30(n^3+9n^2+2n)s_4 s_1^2 + 120(n^3-n)s_3 s_2 s_1 + 30(n^3-3n^2+2n)s_2^3 \\
&\qquad - 120(n^2+3n)s_3 s_1^3 - 270(n^2-n)s_2^2 s_1^2 + 360n\,s_2 s_1^4 - 120s_1^6\} \\
k_7 &= \frac{1}{n^{[7]}}\{(n^6+42n^5+119n^4-42n^3)s_7 - 7(n^5+42n^4+119n^3-42n^2)s_6 s_1 \\
&\qquad - 21(n^5+12n^4-31n^3+18n^2)s_5 s_2 - 35(n^5+5n^3-6n^2)s_4 s_3 \\
&\qquad + 42(n^4+27n^3+44n^2-12n)s_5 s_1^2 + 210(n^4+6n^3-13n^2+6n)s_4 s_2 s_1 \\
&\qquad + 140(n^4+5n^2-6n)s_3^2 s_1 + 210(n^4-3n^3+2n^2)s_3 s_2^2 \\
&\qquad - 210(n^3+13n^2+6n)s_4 s_1^3 - 1260(n^3+n^2-2n)s_3 s_2 s_1^2 \\
&\qquad - 630(n^3-3n^2+2n)s_2^3 s_1 + 840(n^2+4n)s_3 s_1^4 + 2520(n^2-n)s_2^2 s_1^3 \\
&\qquad - 2520n\,s_2 s_1^5 + 720s_1^7\}
\end{aligned}\right\} \quad (12.28)$$

$$
\left.\begin{aligned}
k_8 = \frac{1}{n^{[8]}}\{&(n^7+99n^6+757n^5+141n^4-398n^3+120n^2)s_8-8(n^6+99n^5+757n^4\\
&+141n^3-398n^2+120n)s_7s_1-28(n^6+37n^5-39n^4-157n^3\\
&+278n^2-120n)s_6s_2-56(n^6+9n^5-23n^4+111n^3-218n^2+120n)s_5s_3\\
&-35(n^6+n^5+33n^4-121n^3+206n^2-120n)s_4^2+56(n^5+68n^4+359n^3\\
&-8n^2-60n)s_6s_1^2+336(n^5+23n^4-31n^3-23n^2+30n)s_5s_2s_1\\
&+560(n^5+5n^4+5n^3+5n^2-6n)s_4s_3s_1+420(n^5+2n^4-25n^3\\
&+46n^2-24n)s_4s_2^2+560(n^5-4n^4+11n^3-20n^2+12n)s_3^2s_2\\
&-336(n^4+38n^3+99n^2-18n)s_5s_1^3-2520(n^4+10n^3-17n^2+6n)s_4s_2s_1^2\\
&-1680(n^4+2n^3+7n^2-10n)s_3^2s_1^2-5040(n^4-2n^3-n^2+2n)s_3s_2^2s_1\\
&-630(n^4-6n^3+11n^2-6n)s_2^4+1680(n^3+17n^2+12n)s_4s_1^4\\
&+13{,}440(n^3+2n^2-3n)s_3s_2s_1^3+10{,}080(n^3-3n^2+2n)s_2^3s_1^2\\
&-6720(n^2+5n)s_3s_1^5-25{,}200(n^2-n)s_2^2s_1^4+20{,}160n\,s_2s_1^6-5040s_1^8\}
\end{aligned}\right\} \quad (12.28)
$$

In particular, we have

$$
\left.\begin{aligned}
k_1 &= m_1'\\
k_2 &= \frac{n}{n-1}m_2\\
k_3 &= \frac{n^2}{(n-1)(n-2)}m_3\\
k_4 &= \frac{n^2}{(n-1)(n-2)(n-3)}\{(n+1)m_4-3(n-1)m_2^2\}
\end{aligned}\right\} \quad (12.29)
$$

expressing the k's in terms of the moment-statistics.

Values of k_9 and k_{10} are given explicitly by Zia Ud-Din (1954).

12.11 There is a well-known theorem of symmetric functions which states that any rational integral algebraic symmetric function of $x_1, \ldots, x_n$ can be expressed uniquely, rationally, integrally and algebraically in terms of the symmetric sums s_r. It can thus be so expressed in terms of the k's, for from equations such as (12.28) the s's can be so expressed in terms of the k's. Thus an investigation of the sampling constants of any symmetric function expressible in terms of rational integral algebraic symmetric functions can be translated into an investigation concerning the k's.

The reader who is prepared to take the algebra for granted may prefer to pass over the rest of this chapter and the next rather lightly, noting the main results without following the proofs. We are about to embark on a combinatorial method for deriving systematically the sampling cumulants of k-statistics in terms of parent cumulants. We shall find that the results are substantially simpler than equivalent results stated in terms of moments.

Sampling cumulants of k-statistics

12.12 The problem of determining the sampling-moments or the sampling cumulants of k-statistics is that of finding mean values of powers and products of

those statistics. To any number a with partition $(a_1^{\alpha_1} a_2^{\alpha_2} \dots a_s^{\alpha_s})$ there will correspond a moment

$$\mu'(a_1^{\alpha_1} \dots a_s^{\alpha_s}) = \mathrm{E}(k_{a_1}^{\alpha_1} \dots k_{a_s}^{\alpha_s}) \tag{12.30}$$

and a cumulant $\kappa(a_1^{\alpha_1} \dots a_s^{\alpha_s})$ related to the moments by the identity (cf. (3.28))

$$\Sigma \left\{ \kappa(a_1^{\alpha_1} \dots a_s^{\alpha_s}) \frac{t_{a_1}^{\alpha_1}}{\alpha_1!} \dots \frac{t_{a_s}^{\alpha_s}}{\alpha_s!} \right\} = \log \left\{ \Sigma \mu'(b_1^{\beta_1} \dots b_m^{\beta_m}) \frac{t_{b_1}^{\beta_1}}{\beta_1!} \dots \frac{t_{b_m}^{\beta_m}}{\beta_m!} \right\}. \tag{12.31}$$

For example, the fourth cumulant of k_2 will be expressible in terms of the fourth moment of k_2 about the origin and moments of lower order. These quantities will be written $\kappa(2^4)$ and $\mu'(2^4)$, in accordance with (12.30). Again, the cumulant $\kappa(32)$ corresponds to the moment $\mu'(32)$, the mean value of $k_3 k_2$. Generally, in the simultaneous distribution of the k's there will be a separate formula of degree a for every partition of a.

Now the product $k_{a_1}^{\alpha_1} \dots k_{a_s}^{\alpha_s}$ is homogeneous and of total degree a in the x's. Hence, when mean values are taken $\mu'(a_1^{\alpha_1} \dots a_s^{\alpha_s})$ will be homogeneous and of total order a in the parent μ's. Since the κ's themselves are of homogeneous order in the μ's it follows that $\kappa(a_1^{\alpha_1} \dots a_s^{\alpha_s})$ is of homogeneous order in the κ's. Hence we get the first rule for the sampling of k-statistics :—

Rule 1. $\kappa(a_1^{\alpha_1} \dots a_s^{\alpha_s})$ consists of the sum of terms each of which, except for constants, is a product of parent κ's of order a.

For instance, $\kappa(2^4)$ is of total order 8 and is therefore the sum of terms in κ_8, $\kappa_6 \kappa_2$, $\kappa_5 \kappa_3$, κ_4^2, $\kappa_4 \kappa_2^2$, $\kappa_3^2 \kappa_2$ and κ_2^4. Similarly $\kappa(32)$ will contain a term in κ_5 and one in $\kappa_3 \kappa_2$ and no others. As seen in the next rule, no terms in κ_1 appear.

Rule 2. No term in $\kappa(a_1^{\alpha_1} \dots a_s^{\alpha_s})$ contains κ_1, except $\kappa(1)$ itself.

This follows as in Example 12.1. The k-statistics are independent of the origin and hence their sampling distribution cannot depend on the variable quantity κ_1. The exception occurs when we are dealing with the only statistic which is dependent on the origin, namely k_1, and here $\kappa(1) = \kappa_1$ as is evident from the definitions.

12.13 We now enunciate and illustrate the rules by which the terms in $\kappa(a_1^{\alpha_1} \dots a_s^{\alpha_s})$ can be found. As the proof of the validity of the rules is difficult to grasp until their nature has been comprehended we defer a proof until the next chapter.

To find the term in $\kappa_{b_1}^{\beta_1} \dots \kappa_{b_m}^{\beta_m}$ in $\kappa(a_1^{\alpha_1} \dots a_s^{\alpha_s})$ consider the two-way array

$$\begin{array}{c|c} & b_1 \\ & b_1 \\ & \vdots \\ & b_2 \\ & \vdots \\ \hline a_1 \quad a_1 \dots a_2 \dots & a \end{array} \tag{12.32}$$

where there is a row corresponding to every κ in the term $\kappa_{b_1}^{\beta_1} \ldots \kappa_{b_m}^{\beta_m}$ and a column corresponding to every part in $\kappa(a_1^{\alpha_1} \ldots a_s^{\alpha_s})$. Consider the various ways in which the body of the table can be completed by the insertion of numbers whose row and column sums are the respective b and a numbers; e.g. if we are seeking the coefficient of $\kappa_6 \kappa_2^2$ in $\kappa(4^2 2)$ we shall consider such arrays as

2	2	2	6		2	3	1	6		3	3	.	6	
1	1	.	2		1	1	.	2		1	.	1	2	
1	1	.	2		1	.	1	2		.	1	1	2	(12.33)
4	4	2	10		4	4	2	10		4	4	2	10	

Then the rules by which these arrays give the coefficients of $\kappa_{b_1}^{\beta_1} \ldots \kappa_{b_m}^{\beta_m}$ are as follows:

Rule 3. Every array in which the numbers in the body of the array fall into two or more blocks, each confined to separate rows or columns, is to be ignored.

For instance, in the foregoing example

4	2	.	6
.	2	.	2
.	.	2	2
4	4	2	10

is to be ignored, since the 2×2 block in the top left-hand corner has no row or column number in common with the entry in the bottom right-hand corner.

Rule 4. Subject to the ignoration of terms enjoined by Rule 3, to the coefficient of $\kappa_{b_1}^{\beta_1} \ldots \kappa_{b_m}^{\beta_m}$ in $\kappa(a_1^{\alpha_1} \ldots a_s^{\alpha_s})$ there will be a contribution corresponding to each way of completing the array (12.32). Such of these as do not vanish are composed of a numerical coefficient multiplied by a function of n.

Rule 5. The numerical coefficient is the number of ways in which the column totals, considered as composed of distinct individuals, can be allocated to form the array concerned, divided by $\beta_1! \beta_2! \ldots \beta_m!$.

Rule 6. The function of n, called the pattern function, depends only on the configuration of zeros in the array, not on the actual numbers composing it or on the row and column totals. The function is given by considering the separations of the rows into distinct groups or separates.

(i) With a separation into one separate, there is associated the number n; with a separation into two separates, $n(n-1)$; . . . , with a separation into q separates, $n(n-1) \ldots (n-q+1)$.

(ii) In each separation we count the number of separates in which a particular column is represented by a non-zero entry. If in ρ separates, we assign the factor

$$\frac{(-1)^{\rho-1}(\rho-1)!}{n(n-1)\ldots(n-\rho+1)}$$ to that column.

This is done for each column and the factors multiplied together.

(iii) For each separation, the number in (i) is multiplied by the product in (ii).

(iv) The result, summed over all separations, gives the pattern function.

Rule 7. Any array containing a row which consists of a single non-zero entry has a vanishing pattern function and is to be ignored.

Rule 8. Any array containing a column which consists of a single non-zero entry has a pattern function $1/n$ times that of the array obtained by omitting that column.

Rule 9. Any array the non-zero elements of which consist of two groups connected only by a single column has a vanishing pattern function and is to be ignored.

Example 12.2

As an illustration of these rules (which are not as difficult as they look), suppose we seek for the coefficient of $\kappa_6\kappa_2^2$ in $\kappa(4^2 2)$. If the reader will write down the thirty or so possible arrays with column totals 4, 4, 2 and row totals 6, 2, 2, he will find that the only ones which do not vanish are those of (12.33) and permutations of rows and columns with the same sum, namely

2	2	2	6		2	3	1	6		3	2	1	6		2	3	1	6
1	1	.	2		1	1	.	2		1	1	.	2		1	.	1	2
1	1	.	2		1	.	1	2		.	1	1	2		1	1	.	2
4	4	2	10		4	4	2	10		4	4	2	10		4	4	2	10
		(a)					(b)					(c)					(d)		

3	2	1	6		3	3	.	6		3	3	.	6						
.	1	1	2		1	.	1	2		.	1	1	2						
1	1	.	2		.	1	1	2		1	.	1	2						
4	4	2	10		4	4	2	10		4	4	2	10						
		(e)					(f)					(g)							

(12.34)

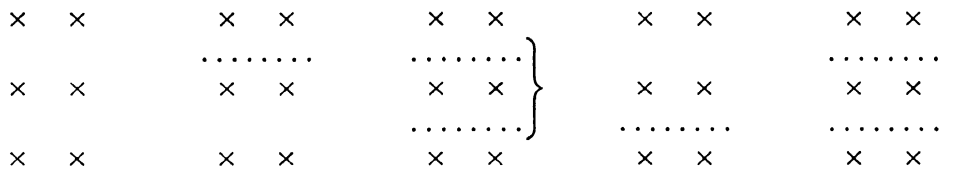
With practice the reader will find it unnecessary to write down arrays such as (c), (d) and (e), which are merely obtained from (b) by permuting rows and columns, but for clarity at this stage they have been set out in full. There is one trap here to be particularly noticed. In array (b) the two columns summing to 4 and the two rows summing to 2 are different, and their permutations result in 4 different arrays. But in array (f), though the rows and columns are different, there are only 2 different arrays.

Each of these arrays contributes to the coefficient required. Consider first of all that from (a). The numerical coefficient is $\left(\frac{4!}{2!1!1!}\right)\left(\frac{4!}{2!1!1!}\right)\cdot\frac{1}{2!} = 72$. The first factor in brackets is the number of ways of allocating 4 individuals in the partition 2, 1, 1, similarly for the second, and we divide by 2! since there are 2 members of the row totals the same, this being the only β factor.

Under Rule 8, the pattern function is $1/n$ times that of

$$\begin{array}{cc} \times & \times \\ \times & \times \\ \times & \times \end{array}$$

The separations, five in number, are as follows :—



The first is, of course, the original pattern array (one separate). The next three consist each of two separates, obtained by taking alone the first, second and third row. The last consists of three separates, one for each row.

The contributions respectively under Rule 6 will be found to be

$$n.\left(\frac{1}{n}\right)\left(\frac{1}{n}\right) = \frac{1}{n}$$

$$3n(n-1)\left\{\frac{-1}{n(n-1)}\right\}\left\{\frac{-1}{n(n-1)}\right\} = \frac{3}{n(n-1)}$$

$$n(n-1)(n-2)\left\{\frac{(-1)^2 2!}{n(n-1)(n-2)}\right\}\left\{\frac{(-1)^2 2!}{n(n-1)(n-2)}\right\} = \frac{4}{n(n-1)(n-2)}.$$

The sum of these is $\frac{n}{(n-1)(n-2)}$ and hence the contribution from array (a) in (12.34) is $\frac{72}{(n-1)(n-2)}$.

Now for arrays (b) to (e), which have all the same numerical factor and the same pattern function and can therefore be considered together. For any one the numerical factor is

$$\left(\frac{4!}{2!1!1!}\right)\left(\frac{4!}{3!1!}\right)\left(\frac{2!}{1!1!}\right)\cdot\frac{1}{2!} = 48$$

and that of the four together is thus 192.

Under Rule 6 the pattern function will depend on the configuration

$$\begin{matrix} \times & \times & \times \\ \times & \times & . \\ \times & . & \times \end{matrix}$$

where $\times$ stands for a non-zero entry and a period for a zero entry. There are five separations of this, one of one separate, three of two separates, and one of three separates. The contribution from the first is

$$n\frac{1}{n}\frac{1}{n}\frac{1}{n} = \frac{1}{n^2}$$

for each column has a non-zero entry in the separate. The contribution from the three separations given respectively by isolating the first, second and third row will be found to be

$$n(n-1)\left[\frac{-1}{n^3(n-1)^3} + \frac{1}{n^3(n-1)^2} + \frac{1}{n^3(n-1)^2}\right] = \frac{2n-3}{n^2(n-1)^2}.$$

The contribution from the separation of three separates is

$$n(n-1)(n-2)\left[\frac{2!}{n(n-1)(n-2)}\cdot\frac{-1}{n(n-1)}\cdot\frac{-1}{n(n-1)}\right]=\frac{2}{n^2(n-1)^2}.$$

The pattern function is the sum of these three contributions and is thus $\frac{1}{(n-1)^2}$.

The contribution from arrays (f) and (g) in (12.34) will be found to be $\frac{32}{(n-1)^2}$.

Hence, adding all the contributions together, we find that the coefficient of $\kappa_6\kappa_2^2$ in $\kappa(4^2 2)$ is

$$\frac{72}{(n-1)(n-2)}+\frac{192}{(n-1)^2}+\frac{32}{(n-1)^2}=\frac{8(37n-65)}{(n-1)^2(n-2)},$$

as shown in equation (12.66) below.

12.14 *Rule 10.* The expression for any $\kappa(a_1^{\alpha_1}\ldots)$ which contains a unit part may be obtained from that without the part by (1) dividing throughout by n and (2) increasing the suffix of one of the κ's by unity in every possible way.

For example, it may be shown that

$$\kappa(2^2)=\frac{\kappa_4}{n}+\frac{2\kappa_2^2}{n-1}.$$

Hence

$$\kappa(2^2 1)=\frac{\kappa_5}{n^2}+\frac{4\kappa_3\kappa_2}{n(n-1)}$$

$$\kappa(2^2 1^2)=\frac{\kappa_6}{n^3}+\frac{4\kappa_3^2}{n^2(n-1)}+\frac{4\kappa_4\kappa_2}{n^2(n-1)},$$

and so on.

12.15 The reader cannot be blamed for doubting whether this rather elaborate combinatorial procedure represents much of an advance on the straightforward algebraical approach considered earlier in the chapter. He should try a few of the more complicated cases both ways. In point of fact, it is so easy to make algebraical slips with either method that most formulae in current use have been checked by being calculated by both methods. The combinatorial method, however, is something more than an algebraical short cut. We shall find in the sequel that several results of importance are derivable directly from it, whereas the straightforward evaluation of expectations yields them only by non-straightforward manœuvres.

Example 12.3

For comparison with Example 12.1 let us find the variance of m_2 by the combinatorial method.

From (12.29) we have

$$k_2=\frac{n}{n-1}m_2.$$

Hence
$$\operatorname{var} m_2 = \left(\frac{n-1}{n}\right)^2 \operatorname{var} k_2$$
$$= \left(\frac{n-1}{n}\right)^2 \kappa(2^2).$$

$\kappa(2^2)$ consists of two terms, one in κ_4 and one in κ_2^2. The only array contributing to the first is

2	2	4
2	2	4

with a numerical factor unity and a pattern function $1/n$. The arrays giving the second are of type

		2
		2
2	2	4

If any entry in this were a 2 the row in which it appeared would contain only a single entry and hence the array would vanish. The only contributing array is therefore

1	1	2
1	1	2
2	2	4

The numerical coefficient is $\left(\frac{2!}{1!1!}\right)^2 \cdot \frac{1}{2!} = 2$. The pattern function will be found to be $1/(n-1)$. Hence

$$\kappa(2^2) = \frac{\kappa_4}{n} + \frac{2\kappa_2^2}{n-1}$$

as given in (12.35), and

$$\operatorname{var} m_2 = \left(\frac{n-1}{n}\right)^2 \kappa(2^2)$$
$$= \left(\frac{n-1}{n}\right)^2 \left\{\frac{1}{n}(\mu_4 - 3\mu_2^2) + \frac{2}{n-1}\mu_2^2\right\}$$
$$= \frac{(n-1)^2}{n^3}\mu_4 + \frac{(3-n)(n-1)}{n^3}\mu_2^2,$$

agreeing with (12.14).

Example 12.4

To find the third moment of k_2 we require $\kappa(2^3)$. This will be the sum of factors in κ_6, $\kappa_4\kappa_2$, κ_3^2 and κ_2^3.

The coefficient of the first is $1/n^2$. For the second we have to consider the array

1	1	2	4
1	1	.	2
2	2	2	6

all others vanishing except the two equivalent partitions obtained when the column with the single entry appears in the first or second place. The numerical factor is then

$$3.\left(\frac{2!}{1!1!}\right)^2 = 12.$$

The pattern function is $1/n$ times that of

$$\begin{matrix} \times & \times \\ \times & \times \end{matrix}$$

i.e. is $1/\{n(n-1)\}$. The coefficient of $\kappa_4\kappa_2$ is then $12/\{n(n-1)\}$.

For the term in κ_3^2 the only contributory array is

$$\begin{array}{ccc|c} 1 & 1 & 1 & 3 \\ 1 & 1 & 1 & 3 \\ \hline 2 & 2 & 2 & 6 \end{array}$$

with a factor $\left(\frac{2!}{1!1!}\right)^3 \cdot \frac{1}{2!} = 4$ and pattern function $\frac{n-2}{n(n-1)^2}$.

For the last term we have to consider the array

$$\begin{array}{ccc|c} 1 & 1 & . & 2 \\ . & . & 1 & 2 \\ 1 & 1 & 1 & 2 \\ \hline 2 & 2 & 2 & 6 \end{array}$$

with a numerical coefficient 8 and a pattern function $\frac{1}{(n-1)^2}$. Collecting terms together we get

$$\kappa(2^3) = \frac{\kappa_6}{n^2} + \frac{12\kappa_4\kappa_2}{n(n-1)} + \frac{4(n-2)}{n(n-1)^2}\kappa_3^2 + \frac{8}{(n-1)^2}\kappa_2^3.$$

This is also the value of the third moment $\mu(2^3)$ measured about the mean of the sampling distribution κ_2. We see that if the parent is normal the third moment reduces to $\frac{8\kappa_2^3}{(n-1)^2}$, i.e. is of order n^{-2}, indicating a rapid tendency towards symmetry.

Example 12.5

Few things illustrate the usefulness of expressing the formulae in terms of cumulants and the power of the combinatorial method better than the simplification imported when the parent population is normal. In this case only terms in κ_2 survive, all higher cumulants vanishing.

As an illustration let us prove that $\kappa(pq) = 0$ for normal samples unless $p = q$.

The only term which can appear in $\kappa(pq)$ is $\kappa_2^{\frac{1}{2}(p+q)}$ and evidently, if $p+q$ is odd, this itself is impossible and $\kappa(pq)$ must vanish. If $p+q$ is even we have to consider

the array

		2
		2
		.
		.
		2
p	q	$p+q$

Now if any entry in this array is 2 the array vanishes since the row concerned will contain only one entry. The reverse can only happen if all the entries are unity, in which case the sums p and q must be equal. This establishes the result.

Example 12.6

If $\kappa(a_1^{\pi_1} \ldots a_s^{\pi_s})$ contains r parts, every term in it is of order $n^{-(r-1)}$. For example, every term in $\kappa(3^2 2^2)$ is of order n^{-3}.

By Rule 7, every array with non-vanishing pattern function (p.f.) has at least two non-zero entries (n.z.e.) in each row. By Rule 8, we may confine ourselves to arrays with at least two n.z.e. in each column. We have to show that the p.f. of every such r-column array is of order $n^{-(r-1)}$. Every array has a single-separate separation; by Rule 6, this contributes $n\left(\frac{1}{n}\right)^r = n^{-(r-1)}$. We now need only show that no separation with more than one separate has a contribution of higher order. Consider any 2-separate separation, S_2; if the p.f. is not to vanish by Rule 9, at least 2 columns must have a n.z.e. in each separate, and by Rule 6 the contribution of S_2 is of order $n^2\left(\frac{1}{n^2}\right)^2\left(\frac{1}{n}\right)^{r-2}$ or less, i.e. it is $o(n^{-(r-1)})$. Any 3-separate separation, S_3, may be derived by a sub-separation of an S_2, and using Rule 3 it follows that the contribution of S_3 is of order $n^3\left(\frac{1}{n^2}\right)^3\left(\frac{1}{n}\right)^{r-3}$, i.e. $o(n^{-(r-1)})$. The same argument, step-by-step, shows that no separation can contribute a term of order as high as $n^{-(r-1)}$. This completes the proof.

Example 12.7

We may use the properties of sampling cumulants to prove a characteristic property of the normal distribution which was discussed analytically in Example 11.3 and geometrically in Example 11.7: namely, that the mean and variance are independent in normal samples.

Let us recall the definition of the joint characteristic function of two variates. If and only if that function factorizes (as in **4.16**) into two univariate c.f.'s, they are independent. Regarding the three c.f.'s as moment-generating functions and taking logarithms we have, from (3.74),

$$\sum_{r,s=1}^{\infty} \kappa_{rs} \frac{(it_1)^r}{r!}\frac{(it_2)^s}{s!} = \sum_{r=1}^{\infty} \kappa_{r0}\frac{(it_1)^r}{r!} + \sum_{s=1}^{\infty} \kappa_{0s}\frac{(it_2)^s}{s!}$$

which implies $\kappa_{rs} = 0$ if $rs \neq 0$. Provided that cumulants of all orders exist this is a necessary and sufficient condition for independence.

Now consider the joint distribution of mean and variance, or, equivalently, of k_1 and k_2. Since $\kappa_p = 0$, $p > 2$, we have from Rule 10

$$\kappa(2^s 1^r) = 0 \qquad (\text{any } r, s \neq 0).$$

Hence k_1 and k_2 are independent.

We may now establish a result due to Geary (1936), that the converse is true, i.e. only in normal variation will mean and variance be independent. For, by Rule 10,

$$\kappa(21^r) = \frac{1}{n^r}\kappa_{r+2}, \qquad r > 0.$$

If k_1 and k_2 are independent this is zero for all $r > 0$. This implies normality on the part of the parent.

It is rather remarkable that we have not had to use the relations $\kappa(2^s 1^r) = 0$ for $s > 1$ to establish the converse; and that the converse requires independence for only one value of n.

Several later writers, e.g. Lukacs (1942), have proved Geary's theorem under less restrictive conditions about the existence of cumulants.

12.16 The following formulae give results for statistics of degree not greater than 10, with some of the 12th degree. They hold for n not less than the order of the k-statistic considered.

Second k-statistic

$$\kappa(2^2) = \frac{\kappa_4}{n} + \frac{2\kappa_2^2}{n-1} \tag{12.35}$$

$$\kappa(2^3) = \frac{\kappa_6}{n^2} + \frac{12\kappa_4\kappa_2}{n(n-1)} + \frac{4(n-2)}{n(n-1)^2}\kappa_3^2 + \frac{8}{(n-1)^2}\kappa_2^3 \tag{12.36}$$

$$\begin{aligned}\kappa(2^4) = {} & \frac{\kappa_8}{n^3} + \frac{24}{n^2(n-1)}\kappa_6\kappa_2 + \frac{32(n-2)}{n^2(n-1)^2}\kappa_5\kappa_3 + \frac{8(4n^2-9n+6)}{n^2(n-1)^3}\kappa_4^2 \\ & + \frac{144}{n(n-1)^2}\kappa_4\kappa_2^2 + \frac{96(n-2)}{n(n-1)^3}\kappa_3^2\kappa_2 + \frac{48}{(n-1)^3}\kappa_2^4\end{aligned} \tag{12.37}$$

$$\begin{aligned}\kappa(2^5) = {} & \frac{\kappa_{10}}{n^4} + \frac{40\kappa_8\kappa_2}{n^3(n-1)} + \frac{80(n-2)}{n^3(n-1)^2}\kappa_7\kappa_3 + \frac{40(5n^2-12n+9)}{n^3(n-1)^3}\kappa_6\kappa_4 \\ & + \frac{16(n-2)(6n^2-12n+7}{n^3(n-1)^4}\kappa_5^2 + \frac{480}{n^2(n-1)^2}\kappa_6\kappa_2^2 + \frac{1280(n-2)}{n^2(n-1)^3}\kappa_5\kappa_3\kappa_2 \\ & + \frac{320(4n^2-9n+6)}{n^2(n-1)^4}\kappa_4^2\kappa_2 + \frac{480(2n^2-7n+6)}{n^2(n-1)^4}\kappa_4\kappa_3^2 + \frac{1920}{n(n-1)^3}\kappa_4\kappa_2^3 \\ & + \frac{1920(n-2)}{n(n-1)^4}\kappa_3^2\kappa_2^2 + \frac{384}{(n-1)^4}\kappa_2^5\end{aligned} \tag{12.38}$$

$$\begin{aligned}\kappa(2^6) = {} & \frac{1}{n^5}\kappa_{12} + \frac{60}{n^4(n-1)}\kappa_{10}\kappa_2 + \frac{160(n-2)}{n^4(n-1)^2}\kappa_9\kappa_3 + \frac{240(2n^2-5n+4)}{n^4(n-1)^3}\kappa_8\kappa_4 \\ & + \frac{96(n-2)(7n^2-14n+9)}{n^4(n-1)^4}\kappa_7\kappa_5 + \frac{4(113n^4-520n^3+950n^2-800n+265)}{n^4(n-1)^5}\kappa_6^2 \\ & + \frac{1200}{n^3(n-1)^2}\kappa_8\kappa_2^2 + \frac{4800(n-2)}{n^3(n-1)^3}\kappa_7\kappa_3\kappa_2 + \frac{2400(5n^2-12n+9)}{n^3(n-1)^4}\kappa_6\kappa_4\kappa_2\end{aligned}$$

$$+\frac{160(n-2)(31n-53)}{n^3(n-1)^4}\kappa_6\kappa_3^2+\frac{960(n-2)(6n^2-12n+7)}{n^3(n-1)^5}\kappa_5^2\kappa_2$$
$$+\frac{1920(n-2)(9n^2-23n+16)}{n^3(n-1)^5}\kappa_5\kappa_4\kappa_3+\frac{480(11n^3-41n^2+59n-31)}{n^3(n-1)^5}\kappa_4^3$$
$$+\frac{9600}{n^2(n-1)^3}\kappa_6\kappa_2^3+\frac{38400(n-2)}{n^2(n-1)^4}\kappa_5\kappa_3\kappa_2^2+\frac{9600(4n^2-9n+6)}{n^2(n-1)^5}\kappa_4^2\kappa_2^2$$
$$+\frac{28800(2n^2-7n+6)}{n^2(n-1)^5}\kappa_4\kappa_3^2\kappa_2+\frac{960(n-2)(5n-12)}{n^2(n-1)^5}\kappa_3^4+\frac{28800}{n(n-1)^4}\kappa_4\kappa_2^4$$
$$+\frac{38400(n-2)}{n(n-1)^5}\kappa_3^2\kappa_2^3+\frac{3840}{(n-1)^5}\kappa_2^6 \tag{12.39}$$

Third k-statistic

$$\kappa(3^2)=\frac{1}{n}\kappa_6+\frac{9}{n-1}\kappa_4\kappa_2+\frac{9}{n-1}\kappa_3^2+\frac{6n}{(n-1)(n-2)}\kappa_2^3 \tag{12.40}$$

$$\kappa(3^3)=\frac{1}{n^2}\kappa_9+\frac{27}{n(n-1)}\kappa_7\kappa_2+\frac{27(3n-4)}{n(n-1)^2}\kappa_6\kappa_3+\frac{27(4n-7)}{n(n-1)^2}\kappa_5\kappa_4$$
$$+\frac{54(4n-7)}{(n-1)^2(n-2)}\kappa_5\kappa_2^2+\frac{162(5n-12)}{(n-1)^2(n-2)}\kappa_4\kappa_3\kappa_2+\frac{36(7n^2-30n+34)}{(n-1)^2(n-2)^2}\kappa_3^3$$
$$+\frac{108n(5n-12)}{(n-1)^2(n-2)^2}\kappa_3\kappa_2^3 \tag{12.41}$$

$$\kappa(3^4)=\frac{1}{n^3}\kappa_{12}+\frac{54}{n^2(n-1)}\kappa_{10}\kappa_2+\frac{108(2n-3)}{n^2(n-1)^2}\kappa_9\kappa_3+\frac{27(17n^2-49n+35)}{n^2(n-1)^3}\kappa_8\kappa_4$$
$$+\frac{108(7n^2-20n+16)}{n^2(n-1)^3}\kappa_7\kappa_5+\frac{27(17n^2-47n+39)}{n^2(n-1)^3}\kappa_6^2+\frac{27(37n-70)}{n(n-1)^2(n-2)}\kappa_8\kappa_2^2$$
$$+\frac{324(19n^2-67n+54)}{n(n-1)^3(n-2)}\kappa_7\kappa_3\kappa_2+\frac{162(65n^2-245n+234)}{n(n-1)^3(n-2)}\kappa_6\kappa_4\kappa_2$$
$$+\frac{108(82n^3-481n^2+958n-640)}{n(n-1)^3(n-2)^2}\kappa_6\kappa_3^2+\frac{108(59n^2-220n+224)}{n(n-1)^3(n-2)}\kappa_5^2\kappa_2$$
$$+\frac{324(75n^3-473n^2+1016n-756)}{n(n-1)^3(n-2)^2}\kappa_5\kappa_4\kappa_3$$
$$+\frac{27(173n^4-1503n^3+4962n^2-7380n+4200)}{n(n-1)^3(n-2)^3}\kappa_4^3$$
$$+\frac{108(71n^2-263n+234)}{(n-1)^3(n-2)^2}\kappa_6\kappa_2^3+\frac{648(79n^2-343n+378)}{(n-1)^3(n-2)^2}\kappa_5\kappa_3\kappa_2^2$$
$$+\frac{486(63n^2-290n+352)}{(n-1)^3(n-2)^2}\kappa_4^2\kappa_2^2+\frac{972(99n^3-688n^2+1612n-1280)}{(n-1)^3(n-2)^3}\kappa_4\kappa_3^2\kappa_2$$
$$+\frac{162(87n^3-594n^2+1420n-1176)}{(n-1)^3(n-2)^3}\kappa_3^4+\frac{972n(23n^2-103n+118)}{(n-1)^3(n-2)^3}\kappa_4\kappa_2^4$$
$$+\frac{648n(103n^2-510n+640)}{(n-1)^3(n-2)^3}\kappa_3^2\kappa_2^3+\frac{648n^2(5n-12)}{(n-1)^3(n-2)^3}\kappa_2^6 \tag{12.42}$$

Fourth k-statistic

$$\kappa(4^2) = \frac{1}{n}\kappa_8 + \frac{16}{n-1}\kappa_6\kappa_2 + \frac{48}{n-1}\kappa_5\kappa_3 + \frac{34}{n-1}\kappa_4^2 + \frac{72n}{(n-1)(n-2)}\kappa_4\kappa_2^2$$
$$+ \frac{144n}{(n-1)(n-2)}\kappa_3^2\kappa_2 + \frac{24n(n+1)}{(n-1)(n-2)(n-3)}\kappa_2^4 \qquad (12.43)$$

$$\kappa(4^3) = \frac{1}{n^2}\kappa_{12} + \frac{48}{n(n-1)}\kappa_{10}\kappa_2 + \frac{16(13n-17)}{n(n-1)^2}\kappa_9\kappa_3 + \frac{12(41n-65)}{n(n-1)^2}\kappa_8\kappa_4$$
$$+ \frac{48(16n-29)}{n(n-1)^2}\kappa_7\kappa_5 + \frac{12(37n-70)}{n(n-1)^2}\kappa_6^2 + \frac{72(11n-19)}{(n-1)^2(n-2)}\kappa_8\kappa_2^2$$
$$+ \frac{288(19n-41)}{(n-1)^2(n-2)}\kappa_7\kappa_3\kappa_2 + \frac{48(203n-523)}{(n-1)^2(n-2)}\kappa_6\kappa_4\kappa_2$$
$$+ \frac{144(56n^2-257n+302)}{(n-1)^2(n-2)^2}\kappa_6\kappa_3^2 + \frac{1440(4n-11)}{(n-1)^2(n-2)}\kappa_5^2\kappa_2$$
$$+ \frac{1152(22n^2-106n+133)}{(n-1)^2(n-2)^2}\kappa_5\kappa_4\kappa_3 + \frac{8(709n^2-3430n+4456)}{(n-1)^2(n-2)^2}\kappa_4^3$$
$$+ \frac{288(19n^3-98n^2+125n+2)}{(n-1)^2(n-2)^2(n-3)}\kappa_6\kappa_2^3 + \frac{1728(24n^3-140n^2+200n+4)}{(n-1)^2(n-2)^2(n-3)}\kappa_5\kappa_3\kappa_2^2$$
$$+ \frac{432(61n^3-371n^2+552n+12)}{(n-1)^2(n-2)^2(n-3)}\kappa_4^2\kappa_2^2 + \frac{864(103n^3-629n^2+948n+24)}{(n-1)^2(n-2)^2(n-3)}\kappa_4\kappa_3^2\kappa_2$$
$$+ \frac{288(41n^4-384n^3+1209n^2-1282n-36)}{(n-1)^2(n-2)^2(n-3)^2}\kappa_3^4 + \frac{288n(53n^2-179n-52)}{(n-1)^2(n-2)^2(n-3)}\kappa_4\kappa_2^4$$
$$+ \frac{1728n(29n^3-196n^2+317n+62)}{(n-1)^2(n-2)^2(n-3)^2}\kappa_3^2\kappa_2^3 + \frac{1728n(n+1)(n^2-5n+2)}{(n-1)^2(n-2)^2(n-3)^2}\kappa_2^6 \qquad (12.44)$$

Fifth k-statistic

$$\kappa(5^2) = \frac{1}{n}\kappa_{10} + \frac{25}{n-1}\kappa_8\kappa_2 + \frac{100}{n-1}\kappa_7\kappa_3 + \frac{200}{n-1}\kappa_6\kappa_4 + \frac{125}{n-1}\kappa_5^2$$
$$+ \frac{200n}{(n-1)(n-2)}\kappa_6\kappa_2^2 + \frac{1200n}{(n-1)(n-2)}\kappa_5\kappa_3\kappa_2 + \frac{850n}{(n-1)(n-2)}\kappa_4^2\kappa_2$$
$$+ \frac{1500n}{(n-1)(n-2)}\kappa_4\kappa + \frac{600n(n+1)}{(n-1)(n-2)(n-3)}\kappa_4\kappa_2^3$$
$$+ \frac{1800n(n+1)}{(n-1)(n-2)(n-3)}\kappa_3^2\kappa_2^2 + \frac{120n^2(n+5)}{(n-1)(n-2)(n-3)(n-4)}\kappa_2^5 \qquad (12.45)$$

Sixth k-statistic

$$\kappa(6^2) = \frac{1}{n}\kappa_{12} + \frac{1}{n-1}(36\kappa_{10}\kappa_2 + 180\kappa_9\kappa_3 + 465\kappa_8\kappa_4 + 780\kappa_7\kappa_5 + 461\kappa_6^2)$$
$$+ \frac{n}{(n-1)(n-2)}(450\kappa_8\kappa_2^2 + 3600\kappa_7\kappa_3\kappa_2 + 7200\kappa_6\kappa_4\kappa_2 + 6300\kappa_6\kappa_3^2$$
$$+ 4500\kappa_5^2\kappa_2 + 21600\kappa_5\kappa_4\kappa_3 + 4950\kappa_4^3)$$

$$+\frac{n(n+1)}{(n-1)(n-2)(n-3)}(2400\kappa_6\kappa_2^3+21600\kappa_5\kappa_3\kappa_2^2+15300\kappa_4^2\kappa_2^2 +54000\kappa_4\kappa_3^2\kappa_2+8100\kappa_3^4)$$
$$+\frac{n^2(n+5)}{(n-1)(n-2)(n-3)(n-4)}(5400\kappa_4\kappa_2^4+21600\kappa_3^2\kappa_2^3)$$
$$+\frac{n(n+1)(n^2+15n-4)}{(n-1)(n-2)(n-3)(n-4)(n-5)}720\kappa_2^6 \tag{12.46}$$

Product–cumulant formulae

$$\kappa(32)=\frac{1}{n}\kappa_5+\frac{6}{n-1}\kappa_3\kappa_2 \tag{12.47}$$

$$\kappa(42)=\frac{1}{n}\kappa_6+\frac{8}{n-1}\kappa_4\kappa_2+\frac{6}{n-1}\kappa_3^2 \tag{12.48}$$

$$\kappa(52)=\frac{1}{n}\kappa_7+\frac{10}{n-1}\kappa_5\kappa_2+\frac{20}{n-1}\kappa_4\kappa_3 \tag{12.49}$$

$$\kappa(62)=\frac{1}{n}\kappa_8+\frac{12}{n-1}\kappa_6\kappa_2+\frac{30}{n-1}\kappa_5\kappa_3+\frac{20}{n-1}\kappa_4^2 \tag{12.50}$$

$$\kappa(72)=\frac{1}{n}\kappa_9+\frac{14}{n-1}\kappa_7\kappa_2+\frac{42}{n-1}\kappa_6\kappa_3+\frac{70}{n-1}\kappa_5\kappa_4 \tag{12.51}$$

$$\kappa(82)=\frac{1}{n}\kappa_{10}+\frac{16}{n-1}\kappa_8\kappa_2+\frac{56}{n-1}\kappa_7\kappa_3+\frac{112}{n-1}\kappa_6\kappa_4+\frac{70}{n-1}\kappa_5^2 \tag{12.52}$$

$$\kappa(43)=\frac{1}{n}\kappa_7+\frac{12}{n-1}\kappa_5\kappa_2+\frac{30}{n-1}\kappa_4\kappa_3+\frac{36n}{(n-1)(n-2)}\kappa_3\kappa_2^2 \tag{12.53}$$

$$\kappa(53)=\frac{1}{n}\kappa_8+\frac{1}{n-1}(15\kappa_6\kappa_2+45\kappa_5\kappa_3+30\kappa_4^2)+\frac{n}{(n-1)(n-2)}(60\kappa_4\kappa_2^2+90\kappa_3^2\kappa_2) \tag{12.54}$$

$$\kappa(63)=\frac{1}{n}\kappa_9+\frac{1}{n-1}(18\kappa_7\kappa_2+63\kappa_6\kappa_3+105\kappa_5\kappa_4) +\frac{n}{(n-1)(n-2)}(90\kappa_5\kappa_2^2+360\kappa_4\kappa_3\kappa_2+90\kappa_3^3) \tag{12.55}$$

$$\kappa(73)=\frac{1}{n}\kappa_{10}+\frac{1}{n-1}(21\kappa_8\kappa_2+84\kappa_7\kappa_3+168\kappa_6\kappa_4+105\kappa_5^2) +\frac{n}{(n-1)(n-2)}(126\kappa_6\kappa_2^2+630\kappa_5\kappa_3\kappa_2+420\kappa_4^2\kappa_2+630\kappa_4\kappa_3^2) \tag{12.56}$$

$$\kappa(54)=\frac{1}{n}\kappa_9+\frac{1}{n-1}(20\kappa_7\kappa_2+70\kappa_6\kappa_3+120\kappa_5\kappa_4) +\frac{n}{(n-1)(n-2)}(120\kappa_5\kappa_2^2+600\kappa_4\kappa_3\kappa_2+180\kappa_3^3) +\frac{n(n+1)}{(n-1)(n-2)(n-3)}240\kappa_3\kappa_2^3 \tag{12.57}$$

$$\kappa(64) = \frac{1}{n}\kappa_{10} + \frac{1}{n-1}(24\kappa_8\kappa_2 + 96\kappa_7\kappa_3 + 194\kappa_6\kappa_4 + 120\kappa_5^2)$$
$$+ \frac{n}{(n-1)(n-2)}(180\kappa_6\kappa_2^2 + 1080\kappa_5\kappa_3\kappa_2 + 720\kappa_4^2\kappa_2 + 1260\kappa_4\kappa_3^2)$$
$$+ \frac{n(n+1)}{(n-1)(n-2)(n-3)}(480\kappa_4\kappa_2^3 + 1080\kappa_3^2\kappa_2^2) \tag{12.58}$$

$$\kappa(32^2) = \frac{1}{n^2}\kappa_7 + \frac{16}{n(n-1)}\kappa_5\kappa_2 + \frac{12(2n-3)}{n(n-1)^2}\kappa_4\kappa_3 + \frac{48}{(n-1)^2}\kappa_2^2\kappa_3 \tag{12.59}$$

$$\kappa(42^2) = \frac{1}{n^2}\kappa_8 + \frac{20}{n(n-1)}\kappa_6\kappa_2 + \frac{8(5n-7)}{n(n-1)^2}\kappa_5\kappa_3 + \frac{4(7n-10)}{n(n-1)^2}\kappa_4^2$$
$$+ \frac{80}{(n-1)^2}\kappa_4\kappa_2^2 + \frac{120}{(n-1)^2}\kappa_3^2\kappa_2 \tag{12.60}$$

$$\kappa(52^2) = \frac{1}{n^2}\kappa_9 + \frac{24}{n(n-1)}\kappa_7\kappa_2 + \frac{20(3n-4)}{n(n-1)^2}\kappa_6\kappa_3 + \frac{20(5n-7)}{n(n-1)^2}\kappa_5\kappa_4$$
$$+ \frac{120}{(n-1)^2}\kappa_5\kappa_2^2 + \frac{480}{(n-1)^2}\kappa_4\kappa_3\kappa_2 + \frac{120}{(n-1)^2}\kappa_3^3 \tag{12.61}$$

$$\kappa(62^2) = \frac{1}{n^2}\kappa_{10} + \frac{28}{n(n-1)}\kappa_8\kappa_2 + \frac{12(7n-9)}{n(n-1)^2}\kappa_7\kappa_3 + \frac{4(41n-56)}{n(n-1)^2}\kappa_6\kappa_4$$
$$+ \frac{20(5n-7)}{n(n-1)^2}\kappa_5^2 + \frac{168}{(n-1)^2}\kappa_6\kappa_2^2 + \frac{840}{(n-1)^2}\kappa_5\kappa_3\kappa_2 + \frac{560}{(n-1)^2}\kappa_4^2\kappa_2$$
$$+ \frac{840}{(n-1)^2}\kappa_4\kappa_3^2 \tag{12.62}$$

$$\kappa(3^22) = \frac{1}{n^2}\kappa_8 + \frac{21}{n(n-1)}\kappa_6\kappa_2 + \frac{6(8n-11)}{n(n-1)^2}\kappa_5\kappa_3 + \frac{9(3n-5)}{n(n-1)^2}\kappa_4^2$$
$$+ \frac{18(6n-11)}{(n-1)^2(n-2)}\kappa_4\kappa_2^2 + \frac{18(9n-20)}{(n-1)^2(n-2)}\kappa_3^2\kappa_2 + \frac{36n}{(n-1)^2(n-2)}\kappa_2^4 \tag{12.63}$$

$$\kappa(432) = \frac{1}{n^2}\kappa_9 + \frac{26}{n(n-1)}\kappa_7\kappa_2 + \frac{24(3n-4)}{n(n-1)^2}\kappa_6\kappa_3 + \frac{10(11n-17)}{n(n-1)^2}\kappa_5\kappa_4$$
$$+ \frac{36(5n-9)}{(n-1)^2(n-2)}\kappa_5\kappa_2^2 + \frac{12(61n-128)}{(n-1)^2(n-2)}\kappa_4\kappa_3\kappa_2 + \frac{36(5n-12)}{(n-1)^2(n-2)}\kappa_3^3$$
$$+ \frac{360n}{(n-1)^2(n-2)}\kappa_3\kappa_2^3 \tag{12.64}$$

$$\kappa(532) = \frac{1}{n^2}\kappa_{10} + \frac{31}{n(n-1)}\kappa_8\kappa_2 + \frac{101n-131}{n(n-1)^2}\kappa_7\kappa_3 + \frac{5(37n-55)}{n(n-1)^2}\kappa_6\kappa_4$$
$$+ \frac{5(23n-35)}{n(n-1)^2}\kappa_5^2 + \frac{30(9n-16)}{(n-1)^2(n-2)}\kappa_6\kappa_2^2 + \frac{30(45n-92)}{(n-1)^2(n-2)}\kappa_5\kappa_3\kappa_2$$
$$+ \frac{60(15n-31)}{(n-1)^2(n-2)}\kappa_4^2\kappa_2 + \frac{30(45n-103)}{(n-1)^2(n-2)}\kappa_4\kappa_3^2 + \frac{720n}{(n-1)^2(n-2)}\kappa_4\kappa_2^3$$
$$+ \frac{1620n}{(n-1)^2(n-2)}\kappa_3^2\kappa_2^2 \tag{12.65}$$

$$\kappa(4^22) = \frac{1}{n^2}\kappa_{10} + \frac{32}{n(n-1)}\kappa_8\kappa_2 + \frac{8(13n-37)}{n(n-1)^2}\kappa_7\kappa_3 + \frac{4(49n-73)}{n(n-1)^2}\kappa_6\kappa_4$$
$$+ \frac{4(29n-46)}{n(n-1)^2}\kappa_5^2 + \frac{8(37n-65)}{(n-1)^2(n-2)}\kappa_6\kappa_2^2 + \frac{1536}{(n-1)^2}\kappa_5\kappa_3\kappa_2$$
$$+ \frac{144(7n-15)}{(n-1)^2(n-2)}\kappa_4^2\kappa_2 + \frac{72(21n-50)}{(n-1)^2(n-2)}\kappa_4\kappa_3^2 + \frac{96(10n^2-27n-1)}{(n-1)^2(n-2)(n-3)}\kappa_4\kappa_2^3$$
$$+ \frac{144(17n^2-53n-2)}{(n-1)^2(n-2)(n-3)}\kappa_3^2\kappa_2^2 + \frac{192n(n+1)}{(n-1)^2(n-2)(n-3)}\kappa_2^5 \qquad (12.66)$$

$$\kappa(43^2) = \frac{1}{n^2}\kappa_{10} + \frac{33}{n(n-1)}\kappa_8\kappa_2 + \frac{6(19n-25)}{n(n-1)^2}\kappa_7\kappa_3 + \frac{3(65n-107)}{n(n-1)^2}\kappa_6\kappa_4$$
$$+ \frac{6(19n-34)}{n(n-1)^2}\kappa_5^2 + \frac{18(19n-33)}{(n-1)^2(n-2)}\kappa_6\kappa_2^2 + \frac{72(23n-52)}{(n-1)^2(n-2)}\kappa_5\kappa_3\kappa_2$$
$$+ \frac{54(19n-48)}{(n-1)^2(n-2)}\kappa_4^2\kappa_2 + \frac{54(33n^2-148n+172)}{(n-1)^2(n-2)^2}\kappa_4\kappa_3^2$$
$$+ \frac{72n(17n-40)}{(n-1)^2(n-2)^2}\kappa_4\kappa_2^3 + \frac{108n(27n-70)}{(n-1)^2(n-2)^2}\kappa_3^2\kappa_2^2 + \frac{216n^2}{(n-1)^2(n-2)^2}\kappa_2^5 \qquad (12.67)$$

$$\kappa(32^3) = \frac{1}{n^3}\kappa_9 + \frac{30}{n^2(n-1)}\kappa_7\kappa_2 + \frac{2(31n-53)}{n^2(n-1)^2}\kappa_6\kappa_3 + \frac{12(9n^2-23n+16)}{n^2(n-1)^3}\kappa_5\kappa_4$$
$$+ \frac{240}{n(n-1)^2}\kappa_5\kappa_2^2 + \frac{360(2n-3)}{n(n-1)^3}\kappa_4\kappa_3\kappa_2 + \frac{24(5n-12)}{n(n-1)^3}\kappa_3^3 + \frac{480}{(n-1)^3}\kappa_3\kappa_2^3 \qquad (12.68)$$

$$\kappa(42^3) = \frac{1}{n^3}\kappa_{10} + \frac{36}{n^2(n-1)}\kappa_8\kappa_2 + \frac{4(23n-37)}{n^2(n-1)^2}\kappa_7\kappa_3 + \frac{4(47n^2-120n+81)}{n^2(n-1)^3}\kappa_6\kappa_4$$
$$+ \frac{12(9n^2-24n+17)}{n^2(n-1)^3}\kappa_5^2 + \frac{360}{n(n-1)^2}\kappa_6\kappa_2^2 + \frac{288(5n-7)}{n(n-1)^3}\kappa_5\kappa_3\kappa_2$$
$$+ \frac{144(7n-10)}{n(n-1)^3}\kappa_4^2\kappa_2 + \frac{24(49n-95)}{n(n-1)^3}\kappa_4\kappa_3^2 + \frac{960}{(n-1)^3}\kappa_4\kappa_2^3 + \frac{2160}{(n-1)^3}\kappa_3^2\kappa_2^2 \qquad (12.69)$$

$$\kappa(3^22^2) = \frac{1}{n^3}\kappa_{10} + \frac{37}{n^2(n-1)}\kappa_8\kappa_2 + \frac{6(17n-27)}{n^2(n-1)^2}\kappa_7\kappa_3 + \frac{3(61n^2-166n+117)}{n^2(n-1)^3}\kappa_6\kappa_4$$
$$+ \frac{2(59n^2-154n+113)}{n^2(n-1)^3}\kappa_5^2 + \frac{6(67n-131)}{n(n-1)^2(n-2)}\kappa_6\kappa_2^2$$
$$+ \frac{24(71n^2-246n+202)}{n(n-1)^3(n-2)}\kappa_5\kappa_3\kappa_2 + \frac{36(29n^2-103n+93)}{n(n-1)^3(n-2)}\kappa_4^2\kappa_2$$
$$+ \frac{36(38n^2-155n+160)}{n(n-1)^3(n-2)}\kappa_4\kappa_3^2 + \frac{72(14n-23)}{(n-1)^3(n-2)}\kappa_4\kappa_2^3$$
$$+ \frac{144(19n-44)}{(n-1)^3(n-2)}\kappa_3^2\kappa_2^2 + \frac{288n}{(n-1)^3(n-2)}\kappa_2^5 \qquad (12.70)$$

12.17 Additional formulae for the case of a normal parent population are as

follows. There are two general formulae :—

$$\kappa(2^r) = \frac{2^{r-1}(r-1)!}{(n-1)^{r-1}}\kappa_2^r \tag{12.71}$$

$$\kappa(p^q 2^r) = \frac{2^r(r+\frac{1}{2}pq-1)!}{(\frac{1}{2}pq-1)!(n-1)^r}\kappa_2^r\kappa(p^q) \tag{12.72}$$

and the following specific formulae of degree 12 and upwards (those of lower degree, of course, being derivable from equations (12.34) to (12.70) by putting all κ's higher than the second equal to zero).

$$\kappa(3^2 2^4) = \frac{34{,}560n}{(n-1)^5(n-2)}\kappa_2^7 \tag{12.73}$$

$$\kappa(3^4 2) = \frac{7776n^2(5n-12)}{(n-1)^4(n-2)^3}\kappa_2^7 \tag{12.74}$$

$$\kappa(3^4 2^2) = \frac{108{,}864n^2(5n-12)}{(n-1)^5(n-2)^3}\kappa_2^8 \tag{12.75}$$

$$\kappa(3^4 2^3) = \frac{1{,}741{,}824n^2(5n-12)}{(n-1)^6(n-2)^3}\kappa_2^9 \tag{12.76}$$

$$\kappa(3^6) = \frac{466{,}560n^3(22n^2-111n+142)}{(n-1)^5(n-2)^5}\kappa_2^9 \tag{12.77}$$

$$\kappa(3^6 2) = \frac{18}{n-1}\kappa_2\kappa(3^6) \tag{12.78}$$

$$\kappa(3^6 2^2) = \frac{360}{(n-1)^2}\kappa_2^2\kappa(3^6) \tag{12.79}$$

$$\kappa(4^2 2^2) = \frac{1920n(n+1)}{(n-1)^3(n-2)(n-3)}\kappa_2^6 \tag{12.80}$$

$$\kappa(4^2 2^3) = \frac{23{,}040n(n+1)}{(n-1)^4(n-2)(n-3)}\kappa_2^7 \tag{12.81}$$

$$\kappa(4^2 2^4) = \frac{322{,}560n(n+1)}{(n-1)^5(n-2)(n-3)}\kappa_2^8 \tag{12.82}$$

$$\kappa(4^3 2) = \frac{20{,}736n(n+1)(n^2-5n+2)}{(n-1)^3(n-2)^2(n-3)^2}\kappa_2^7 \tag{12.83}$$

$$\kappa(4^3 2^2) = \frac{290{,}304n(n+1)(n^2-5n+2)}{(n-1)^4(n-2)^2(n-3)^2}\kappa_2^8 \tag{12.84}$$

$$\kappa(4^3 2^3) = \frac{4{,}644{,}864n(n+1)(n^2-5n+2)}{(n-1)^5(n-2)^2(n-3)^2}\kappa_2^9 \tag{12.85}$$

$$\kappa(4^4) = \frac{6912n(n+1)}{(n-1)^3(n-2)^3(n-3)^3}\{53n^4-428n^3+1025n^2-474n+180\}\kappa_2^8 \tag{12.86}$$

$$\kappa(4^4 2) = \frac{16}{n-1}\kappa_2\kappa(4^4) \tag{12.87}$$

$$\kappa(4^4 2^2) = \frac{288}{(n-1)^2}\kappa_2^2\kappa(4^4) \tag{12.88}$$

$$\kappa(4^5) = \frac{484.12^5}{n^4}\kappa_2^{10} \text{ approximately} \tag{12.89}$$

It should be noted that (12.73) to (12.89) are all expressions of even degree. Those of odd degree are all zero, because $\kappa_r = 0$ for $r>2$ in normal variation. In fact, (12.72) implies that $\kappa(p^q 2^r) = 0$ if pq is odd or if $q = 1$ and $p>2$. Methods of proof of (12.71) and (12.72) are suggested in Exercises 12.9 and 12.10. Hsu and Lawley (1939) gave exact results for $\kappa(4^5)$ and $\kappa(4^6)$.

12.18 As an illustration of the way in which the sampling formulae can be used to approximate to a sampling distribution, let us consider the distribution of $\sqrt{b_1}$ in samples from a normal population. We have, in terms of the sample moments,

$$\sqrt{b_1} = \frac{m_3}{m_2^{3/2}} = \frac{n-2}{\sqrt{\{n(n-1)\}}}\frac{k_3}{k_2^{3/2}}.$$

For a normal distribution the variance of k_3, $\kappa(3^2)$ is, by (12.40), equal to

$$\frac{6n\kappa_2^3}{(n-1)(n-2)}.$$

We therefore consider the statistic

$$x = \sqrt{\frac{(n-1)(n-2)}{6n}}k_3 k_2^{-3/2} \tag{12.90}$$

$$= \frac{n-1}{\sqrt{\{6(n-2)\}}}\sqrt{b_1} \tag{12.91}$$

which will, to order n^{-1}, have unit variance. We have

$$x = \sqrt{\frac{(n-1)(n-2)}{6n}}\frac{k_3}{\kappa_2^{3/2}}\left\{1+\frac{k_2-\kappa_2}{\kappa_2}\right\}^{-3/2} \tag{12.92}$$

Since the population is symmetrical, the mean value of x is zero. We then have, expanding (12.92),

$$x^2 = \frac{(n-1)(n-2)}{6n}\cdot\frac{1}{\kappa_2^3}\left\{k_3^2 - \frac{3}{\kappa_2}k_3^2(k_2-\kappa_2) + \frac{6}{\kappa_2^2}k_3^2(k_2-\kappa_2)^2 - \frac{10}{\kappa_2^3}k_3^2(k_2-\kappa_2)^3 + \frac{15}{\kappa_2^4}k_3^2(k_2-\kappa_2)^4 - \frac{21}{\kappa_2^5}k_3^2(k_2-\kappa_2)^5 + \frac{28}{\kappa_2^6}k_3^2(k_2-\kappa_2)^6 + \ldots\right\}. \tag{12.93}$$

The variance may be obtained by taking mean values of both sides, and since κ_2 is the mean value of k_2 we have

$$\text{var } x = \frac{(n-1)(n-2)}{6n}\cdot\frac{1}{\kappa_2^3}\left\{\mu(3^2) - \frac{3}{\kappa_2}\mu(3^2 2) + \frac{6}{\kappa_2^2}\mu(3^2 2^2) - \frac{10}{\kappa_2^3}\mu(3^2 2^3) + \frac{15}{\kappa_2^4}\mu(3^2 2^4) - \frac{21}{\kappa_2^5}\mu(3^2 2^5) + \frac{28}{\kappa_2^6}\mu(3^2 2^6) + \ldots\right\}. \tag{12.94}$$

We now express the product μ's in terms of product κ's by using equation (12.31) and identifying coefficients. For a normal distribution $\kappa(32^r) = 0$ and we will take our approximation to order n^{-4}, so that κ's of five parts or more may be neglected.

We then find

$$\begin{aligned}\operatorname{var} x = \frac{(n-1)(n-2)}{6n}\cdot\frac{1}{\kappa_2^3}\Big[&\kappa(3^2)-\frac{3}{\kappa_2}\kappa(3^2 2)+\frac{6}{\kappa_2^2}\{\kappa(3^2 2^2)+\kappa(3^2)\kappa(2^2)\}\\&-\frac{10}{\kappa_2^3}\{\kappa(3^2 2^3)+3\kappa(3^2 2)\kappa(2^2)+\kappa(3^2)\kappa(2^3)\}+\frac{15}{\kappa_2^4}\{6\kappa(3^2 2^2)\kappa(2^2)\\&+4\kappa(3^2 2)\kappa(2^3)+\kappa(3^2)\kappa(2^4)+3\kappa(3^2)\kappa^2(2^2)\}-\frac{21}{\kappa_2^5}\{15\kappa(3^2 2)\kappa^2(2^2)\\&+10\kappa(3^2)\kappa(2^3)\kappa(2^2)\}+\frac{28}{\kappa_2^6}.15\kappa(3^2)\kappa^3(2^2)\Big].\end{aligned} \tag{12.95}$$

Substituting the values of equations (12.35) to (12.89) we find, after some purely algebraic reduction,

$$\begin{aligned}\operatorname{var} x &= 1-\frac{6}{n-1}+\frac{28}{(n-1)^2}-\frac{120}{(n-1)^3}+\ldots\\&= 1-\frac{6}{n}+\frac{22}{n^2}-\frac{70}{n^3}+\ldots\end{aligned} \tag{12.96}$$

In a similar way it may be shown that

$$\mu_4(x) = 3-\frac{1056}{n^2}+\frac{24{,}132}{n^3}-\ldots \tag{12.97}$$

$\mu_3(x)$ is zero, for the distribution is symmetrical.

In point of fact, the exact results for this case are known (cf. Exercise 12.9) and hence can be used to check the algebra here.

It appears from (12.96) and (12.97) that as $n\to\infty$ the second moment of x tends to unity and the fourth moment to 3, which is in conformity with tendency to normality. But the tendency is by no means very rapid. When $n = 100$ the variance is approximately 0·942 and in assuming x to be distributed with unit variance we should commit an error of about 6 per cent.

There are two ways of improving on the first approximation that x is normally distributed with unit variance. In the first place we may consider a transformation to a new variate ξ, chosen so that ξ is normally distributed to order n^{-2}. Secondly, we may fit a Pearson curve to the distribution of x, using the values of moments given by (12.96) and (12.97). The appropriate curve is the Type VII

$$dF \propto \left(1+\frac{x^2}{a^2}\right)^{-m} dx. \tag{12.98}$$

The first line was adopted by Fisher (1928), who obtained the following transformation :

$$\xi = x\left(1+\frac{3}{n}+\frac{91}{4n^2}\right)-\frac{3}{2n}\left(1-\frac{111}{2n}\right)(x^3-3x)-\frac{33}{8n^2}(x^5-10x^3+15x). \tag{12.99}$$

The second was adopted by E. S. Pearson (1930), who tabulated the 1 per cent. and 5 per cent. significance points of (12.98), that is to say the values of the variable x for various values of n such that 99 per cent. and 95 per cent. of the total frequency of the sampling distribution falls within a range of $\pm x$ on each side of the mean.

12.19 By such methods approximations can be obtained to the sampling distributions of any statistics which are expressible as symmetric functions. The reader may care to refer to investigations on these lines by F. N. David (1949*a*, *b*) into the sampling moments of the coefficient of variation, the logarithm of the variance-ratio (Fisher's z) and the variance-ratio itself. Difficulties associated with the expansion of denominators in ratio-statistics may sometimes be overcome by the use of the following device.

Consider a statistic t which is in the form of a ratio a/b, where b is always positive. Then

$$\begin{aligned} F(t_0) &= \mathrm{P}\{t \leqslant t_0\} = \mathrm{P}\{a/b \leqslant t_0\} \\ &= \mathrm{P}\{a \leqslant b t_0\} = \mathrm{P}\{a - b t_0 \leqslant 0\}. \end{aligned} \tag{12.100}$$

We then consider the statistic $a - b t_0$, which does not suffer from the disadvantage of being a ratio. For example, suppose our statistic is "Student's" t, given by

$$t = \frac{(\bar{x}-\mu)\sqrt{(n-1)}}{s},$$

where μ is the parent mean and $s^2 = \Sigma(x-\bar{x})^2/n$. The distribution is symmetrical and hence

$$\begin{aligned} \mathrm{P}\{|t| \leqslant |t_0|\} &= \mathrm{P}\{t^2 \leqslant t_0^2\} \\ &= \mathrm{P}\{(\bar{x}-\mu)^2(n-1) - s^2 t_0^2 \leqslant 0\}. \end{aligned} \tag{12.101}$$

We then consider a statistic

$$u = (n-1)(\bar{x}-\mu)^2 - t_0^2 s^2 \tag{12.102}$$

and can obtain as many moments or cumulants as we like for given μ and t_0^2.

In this particular instance, for normal variation we know that $\bar{x}$ and s are independent and can find the cumulants of u directly. For the characteristic function of u, writing the imaginary quantity it as θ, we have, from (11.39) and (11.40),

$$\begin{aligned} \phi(t) &\propto \int\int \exp[\{(n-1)(\bar{x}-\mu)^2 - t_0^2 s^2\}\theta] \exp\{-\tfrac{1}{2}n(\bar{x}-\mu)^2\}\, d\bar{x} \exp\{-\tfrac{1}{2}ns^2\} s^{n-2}\, ds \\ &= \left\{1 - \frac{2(n-1)}{n}\theta\right\}^{-\frac{1}{2}} \left\{1 + \frac{2t_0^2\theta}{n}\right\}^{-\frac{1}{2}(n-1)}. \end{aligned} \tag{12.103}$$

Hence we find, for u,

$$\kappa_1 = \frac{n-1}{n}(1-t_0^2) \tag{12.104}$$

$$\kappa_2 = \frac{2(n-1)}{n^2}(n-1+t_0^4). \tag{12.105}$$

As n tends to infinity ϕ tends to $(1-2\theta)^{-\frac{1}{2}}\exp(-\theta t_0^2)$, which apart from the second term is the c.f. of a Type III or χ^2 distribution with one degree of freedom.

Now we require only the probability that $u \leqslant 0$, which is therefore the probability that a deviate with mean (12.104) and variance (12.105) is less than zero. For $n = 20$, $t_0 = 2$ we have $\kappa_1 = -2{\cdot}85$, $\kappa_2 = 3{\cdot}42$. This is equivalent to a standardized deviate of $-2{\cdot}85/\sqrt{3{\cdot}42} = -1{\cdot}563$. The probability that this is exceeded in a normal

distribution is 0·941, which is accordingly our approximation to the probability that t will not exceed $t_0 = 2$ in absolute value. The true value is 0·940.

Although we have considered a normal case, the method is particularly useful in judging the effect of departures from normality on certain statistics. Reference may be made to David and Johnson (1951) for some work on these lines concerning the variance-ratio.

Sampling from finite populations

12.20 In sampling without replacement from finite populations the algebra of expectations becomes much more complex in virtue of the fact that successive drawings are no longer independent. Let us follow through the work of **12.3** in the finite case to illustrate the point. Equation (12.2) still holds, i.e.

$$\mathrm{E}\,m_2 = \frac{n-1}{n}\mathrm{E}\,x^2 - \frac{n-1}{n}\mathrm{E}\,x_i x_j, \qquad (i \neq j). \tag{12.106}$$

If capital M's are written to denote the moments of the parent population of N members, we now have, by definition,

$$\begin{aligned} \mathrm{E}\,x^2 &= M_2' \\ \mathrm{E}\,x_i x_j &= \sum_{i,j=1}^{N} x_i x_j / \{N(N-1)\} \qquad (i \neq j) \\ &= \frac{\left(\sum_1^N x_i\right)^2 - \sum_1^N x_i^2}{N(N-1)} = \frac{N^2 M_1'^2 - N M_2'}{N(N-1)}. \end{aligned}$$

Substituting in (12.106) we then find

$$\begin{aligned} \mathrm{E}\,m_2 &= \frac{n-1}{n}\left\{\frac{N}{N-1}(M_2' - M_1'^2)\right\} \\ &= \frac{n-1}{n}\cdot\frac{N}{N-1}M_2. \end{aligned} \tag{12.107}$$

Again, with sufficient pertinacity, we can find all the moments of symmetric-function statistics by this kind of approach, but the algebra rapidly becomes unmanageable. A simpler method is as follows.

12.21 Let us first of all note that the expectation of any augmented symmetric function for a sample bears a very simple relation to the corresponding function for the parent. Denoting the two functions by the subscripts n and N respectively, we have

$$\mathrm{E}\,[p_1^{\pi_1}\ldots p_s^{\pi_s}]_n / n^{[\rho]} = [p_1^{\pi_1}\ldots p_s^{\pi_s}]_N / N^{[\rho]} \tag{12.108}$$

where $\rho = \sum_i \pi_i$. It follows that if K_p is the pth k-statistic for the parent, we have

$$\mathrm{E}_N k_p = K_p \tag{12.109}$$

where E_N means an expectation in the set of N values forming the parent. This attractive property of k-statistics is one of their most useful features. We have, for

example, $k_2 = nm_2/(n-1)$ and hence

$$\mathrm{E}_N \frac{nm_2}{n-1} = K_2 = \frac{NM_2}{N-1}$$

from which (12.107) follows immediately.

Secondly, let us imagine the population of N as itself a sample from an infinite population with cumulants κ_p. If E refers to expectation in this infinite population we shall have

$$\mathrm{E}\,k_p = \mathrm{E}\,K_p = \kappa_p.$$

If then we arrive at any stage with some symmetric function whose expectation E may be expressed *linearly* in terms of κ's, say

$$\mathrm{E}(f) = \Sigma a_j \kappa_j, \tag{12.110}$$

we may write down the analogous relation

$$\mathrm{E}_N(f) = \Sigma a_j K_j. \tag{12.111}$$

For if this were not true, $\mathrm{E}_N(f)$ could be expressed as some other function of K's whose expectation E would be the same as that of $\Sigma a_j K_j$, and this would imply a relationship among the K's. This is sometimes known as the Irwin-Kendall principle from the authors who introduced it (1944).

Example 12.8. Moments of the mean in the finite case

The mean-statistic m_1 is equal to k_1. Thus

$$\mathrm{E}_N(k_1) = K_1. \tag{12.112}$$

From the known result for the variance of the mean we have

$$\mathrm{E}(k_1^2) = \frac{\kappa_2}{n} + \kappa_1^2, \quad \mathrm{E}(K_1^2) = \frac{\kappa_2}{N} + \kappa_1^2.$$

Hence

$$\mathrm{E}(k_1 - K_1)^2 = \mathrm{E}\,k_1^2 - \mathrm{E}\,K_1^2 = \kappa_2\left(\frac{1}{n} - \frac{1}{N}\right).$$

It then follows at once that

$$\mathrm{E}_N(k_1 - K_1)^2 = K^2\left(\frac{1}{n} - \frac{1}{N}\right), \tag{12.113}$$

giving us the variance of the mean. If we prefer to express this in terms of parent moments we have

$$\mathrm{E}_N(k_1 - K_1)^2 = \frac{N-n}{(N-1)n} M_2. \tag{12.114}$$

For the third moment we have

$$\mathrm{E}(k_1 - \kappa_1)^3 = \mathrm{E}(k_1 - K_1 + K_1 - \kappa_1)^3$$

$$= \mathrm{E}(k_1 - K_1)^3 + 3\mathrm{E}(k_1 - K_1)^2(K_1 - \kappa_1) + 3\mathrm{E}(k_1 - K_1)(K_1 - \kappa_1)^2 + \mathrm{E}(K_1 - \kappa_1)^3. \tag{12.115}$$

The third term on the right vanishes, since the expectation of $k_1 - K_1$ in the finite population vanishes. Also we have

$$\mathrm{E}(k_1 - \kappa_1)^3 = \frac{\kappa_3}{n^2}, \quad \mathrm{E}(K_1 - \kappa_1)^3 = \frac{\kappa_3}{N^2}.$$

Hence, from (12.115) we find

$$\mathrm{E}(k_1-K_1)^3 = \kappa_3\left(\frac{1}{n^2}-\frac{1}{N^2}\right)-3\mathrm{E}(K_1-\kappa_1)(k_1-K_1)^2.$$

Writing

$$\alpha_r = \frac{1}{n^r}-\frac{1}{N^r} \tag{12.116}$$

we then find

$$\begin{aligned}\mathrm{E}(k_1-K_1)^3 &= \alpha_2\kappa_3-3\mathrm{E}(K_1-\kappa_1)\,\mathrm{E}_N(k_1-K_1)^2\\ &= \alpha_2\kappa_3-3\mathrm{E}(K_1-\kappa_1)\alpha_1 K_2\\ &= \alpha_2\kappa_3-3\alpha_1\mathrm{E}(K_1-\kappa_1)(K_2-\kappa_2)\\ &= \alpha_2\kappa_3-3\alpha_1\kappa_3/N\\ &= \left(\alpha_2-\frac{3\alpha_1}{N}\right)\kappa_3.\end{aligned}$$

Hence, for the third moment,

$$\mathrm{E}_N(k_1-K_1)^3 = \left(\alpha_2-\frac{3\alpha_1}{N}\right)K_3. \tag{12.117}$$

Substituting for α_2, α_1 and

$$K_3 = \frac{N^2}{(N-1)(N-2)}M_3$$

we find

$$\mathrm{E}_N(m_1-M_1)^3 = \frac{(N-n)(N-2n)}{n^2(N-1)(N-2)}M_3. \tag{12.118}$$

Likewise we find

$$\begin{aligned}\mathrm{E}_N(k_1-K_1)^4 = K_4\left\{\alpha_3-4\alpha_2/N+6\alpha_1/N^2-\frac{3(N-1)}{N(N+1)}(\alpha_2-2\alpha_1/N)\right\}\\ +3K_2^2\frac{N-1}{N+1}(\alpha_2-2\alpha_1/N)\end{aligned} \tag{12.119}$$

which is equivalent to

$$\begin{aligned}\mathrm{E}_N(m_1-M_1)^4 = \frac{N-n}{n^3(N-1)(N-2)(N-3)}\{(N^2-6nN+N+6n^2)M_4\\ +3N(n-1)(N-n-1)M_2^2\}.\end{aligned} \tag{12.120}$$

Example 12.9

We have

$$\mathrm{E}(k_r-\kappa_r)(k_1-\kappa_1) = \mathrm{E}(k_r-K_r)(k_1-K_1)+\mathrm{E}(K_r-\kappa_r)(K_1-\kappa_1)$$

and hence

$$\begin{aligned}\mathrm{E}(k_r-K_r)(k_1-K_1) &= \frac{1}{n}\kappa_{r+1}-\frac{1}{N}\kappa_{r+1}\\ &= \alpha_1\kappa_{r+1},\end{aligned}$$

from which

$$\mathrm{E}_N(k_r-K_r)(k_1-K_1) = \alpha_1 K_{r+1}. \tag{12.121}$$

Thus, if the finite population is symmetrical, any k-statistic of even order is uncorrelated with k_1, for K_{r+1} then vanishes.

12.22 In developing our next point, we encounter a difficulty of notation. We have already used the symbols κ, k and K. In the next chapter, when we proceed to multivariate extensions, we shall require corresponding quantities with several suffixes, e.g. κ_r, k_{rs}, K_{rs}. We now require a notation for a quantity whose mean value is the product of κ's, e.g. $\kappa_r \kappa_s$. Previous writers (Tukey (1950), Wishart (1952), Kendal (1952)) notwithstanding, we shall write these quantities with the letter l and thus we define them by

$$\mathrm{E}\, l_{rs\ldots u} = \kappa_r \kappa_s \ldots \kappa_u. \tag{12.122}$$

It will, as before, follow that if $L_{rs\ldots u}$ is the corresponding statistic for a finite population

$$\mathrm{E}_N l_{rs\ldots u} = L_{rs\ldots u}. \tag{12.123}$$

The use of these functions derives from the fact that we can express them in terms of augmented symmetrics once and for all and hence derive *non-linear* functions of them as linear functions to which the argument of **12.21** can be applied. The necessary table is given by Abdel-Aty (1954) up to weight 12. Wishart (1952) gave expressions up to order 8 for these multiple k-statistics in terms of products of simple k-statistics. For example we have, from his table,

$$k_2 k_1 = l_2 l_1 = l_3/n + l_{21} = k_3/n + l_{21} \tag{12.124}$$

and it follows at once that

$$\mathrm{E}(k_2 k_1) = \kappa_3/n + \kappa_2 \kappa_1, \tag{12.125}$$

and also that

$$\mathrm{E}_N(k_2 k_1) = K_3/n + K_2 K_1. \tag{12.126}$$

Similarly we have

$$k_1^4 = k_4/n^3 + 4l_{31}/n^2 + 3l_{22}/n^2 + 6l_{211}/n + l_{1111} \tag{12.127}$$

whence

$$\mathrm{E}(k_1^4) = \kappa_4/n^3 + 4\kappa_3 \kappa_1/n^2 + 3\kappa_2^2/n^2 + 6\kappa_2 \kappa_1^2/n + \kappa_1^4, \tag{12.128}$$

again with a parallel expression for $\mathrm{E}_N(k_1^4)$.

Example 12.10

One of the most useful practical applications of these results is that they enable us to write down unbiassed estimators of products of cumulants. We found in Examples 12.1 and 12.3 that

$$\mathrm{var}\, m_2 = \frac{(n-1)^2}{n^2}\left\{\frac{\kappa_4}{n} + \frac{2\kappa_2^2}{n-1}\right\}. \tag{12.129}$$

In some practical applications of this result we may require to calculate the quantity on the right, but if we have only a sample available we are ignorant of κ_4 and κ_2 and have to estimate them. Thus we require a quantity whose mean value is the function on the right in (12.129). In terms of the multiple k-statistics this is written down at sight as

$$\frac{(n-1)^2}{n^2}\left\{\frac{k_4}{n} + \frac{2l_{22}}{n-1}\right\}. \tag{12.130}$$

From Wishart's table we then read off

$$l_{22} = \frac{n-1}{n+1}\left(k_2^2 - \frac{k_4}{n}\right)$$

and, on substitution in (12.130), for the unbiassed estimator we find

$$\frac{(n-1)^2}{n^2(n+1)}\left\{\frac{(n-1)k_4}{n} + 2k_2^2\right\}. \tag{12.131}$$

Example 12.11

To find an unbiassed estimator of $\mu_3\mu_1'^2$.

This is equal to $\kappa_3\kappa_1^2$ and the estimator is l_{311}. From Appendix Table 11 we write down the estimator directly in terms of augmented symmetric functions as

$$l_{311} = \frac{2}{n^{[5]}}[1^5] - \frac{3}{n^{[4]}}[21^3] + \frac{[31^2]}{n^{[3]}}.$$

From Appendix Table 10 we can then substitute for the augmented symmetrics in terms of the power-sums (r). We find

$$l_{311} = \frac{2}{n^{[5]}}\{24(5) - 30(4)(1) - 20(3)(2) + 20(3)(1)^2 + 15(2)^2(1) - 10(2)(1)^3 + (1)^5\}$$

$$- \frac{3}{n^{[4]}}\{-6(5) + 6(4)(1) + 5(3)(2) - 3(3)(1)^2 - 3(2)^2(1) + (2)(1)^3\}$$

$$+ \frac{1}{n^{[3]}}\{2(5) - 2(4)(1) - (3)(2) + (3)(1)^2\}$$

which reduces, on putting $(r) = s_r$, to

$$\frac{1}{n^{[5]}}\{2n(n+2)s_5 - 2(n^2+2n+6)s_4s_1 - (n^2+8n-8)s_3s_2 + (n^2+2n+16)s_3s_1^3$$
$$+ (9n-16)s_2^2s_1 - (3n+8)s_2s_1^3 + 2s_1^5\}$$

The l-statistics are called " polykays " by some American authors, but we feel that there are limits to linguistic miscegenation which should not be exceeded.

EXERCISES

12.1 By using (12.25) directly, write down k_3 and k_4 in terms of the augmented symmetric functions and hence verify (12.29).

12.2 Assuming that

$$k_3 = a_0s_3 + a_1s_2s_1 + a_2s_1^3$$

obtain the coefficients a by taking expectations of both sides and hence derive k_3 of (12.28).

12.3 Use the approach of Example 12.1 to obtain $\kappa(2^3)$ in terms of the parent cumulants.

12.4 Show that the pattern functions of the following patterns :—

$$\begin{array}{ccc} \times & \times & \times \\ \times & . & \times \\ \times & \times & . \end{array} \qquad \begin{array}{cccc} \times & \times & \times & \times \\ . & . & \times & \times \\ \times & \times & . & . \end{array}$$

are $\dfrac{1}{(n-1)^2}$ and $\dfrac{1}{n(n-1)^2}$ respectively.

(Fisher, 1928)

12.5 Show that the pattern function of the pattern

$$\begin{array}{cccccc} \times & \times & \times & . & . & . \\ \times & \times & \times & . & . & . \end{array}$$

with p columns is $\dfrac{1}{n^{p-1}}\left\{1-\dfrac{(-1)^{p-1}}{(n-1)^{p-1}}\right\}$. (Fisher, 1928)

12.6 Verify the formulae of equations (12.37) and (12.43).

12.7 If a pattern contains a column with three entries ; if the patterns obtained by suppressing this column and (1) amalgamating the three rows, (2) amalgamating the pairs of three rows, and (3) leaving the rows unamalgamated are A, B_1, B_2, B_3 and C respectively, show that the pattern function for the original pattern is

$$\frac{n}{(n-1)(n-2)}A-\frac{1}{(n-1)(n-2)}(B_1+B_2+B_3)+\frac{2}{n(n-1)(n-2)}C.$$

Deduce that the function of the pattern

$$\begin{array}{ccc} \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \\ \times & \times & . \end{array}$$

is $\dfrac{n^3-8n^2+17n+2}{(n-1)^2(n-2)^2(n-3)}$.

(Fisher and Wishart, 1931)

12.8 Use the combinatorial method to verify that, to order n^{-1},

$$\text{var } m_4 = (\mu_8-\mu_4^2+16\mu_3^2\mu_2-8\mu_5\mu_3)/n.$$

12.9 Show that in normal samples k_2 is independent of $k_p/k_2^{\frac{1}{2}p}$ (cf. Exercise 11.4). Hence show that

$$\text{var}\,(k_3/k_2^{3/2}) = \frac{6n(n-1)}{(n-2)(n+1)(n+3)}$$

$$\mu_4(k_3/k_2^{3/2}) = \frac{108n^2(n-1)^2(n^2+27n-70)}{(n-2)^3(n+1)(n+3)(n+5)(n+7)(n+9)}$$

and verify equations (12.96) and (12.97).

12.10 Defining y by the relation

$$y = \sqrt{\frac{(n-1)(n-2)(n-3)}{24n(n+1)}}\frac{k_4}{k_2^2},$$

x

show by the method of **12.18** that the moments of the distribution of y in samples from a normal population are

$$\mu_1 = 0$$

$$\mu_2 = 1 - \frac{12}{n} + \frac{88}{n^2} - \frac{532}{n^3} \cdots$$

$$\mu_3 = 6\sqrt{\frac{6}{n}}\left(1 - \frac{65}{2n} + \frac{4811}{8n^2} - \frac{136{,}605}{16n^3} \cdots\right)$$

$$\mu_4 = 3 + \frac{468}{n} - \frac{32{,}196}{n^2} + \frac{1{,}118{,}388}{n^3} \cdots$$

(E. S. Pearson, 1930)

12.11 Show that, in samples from a finite population,

$$\mathrm{E}_N(k_2 - K_2)^2 = \frac{(N-n)(Nn-n-N-1)}{n(n-1)N(N+1)} K_4 + \frac{2(N-n)}{(n-1)(N+1)} K_2^2$$

which is equivalent to

$$\mathrm{E}_N\left(m_2 - \frac{n-1}{n}\frac{N}{N-1}M_2\right)^2 = \frac{(n-1)N(N-n)}{n^3(N-1)^2(N-2)(N-3)}$$
$$\{(nN-N-n-1)(N-1)M_4 - (nN^2-3N^2+6N-3n-3)M_2^2\}$$

12.12 Show likewise that

$$\mathrm{E}_N(k_3 - K_3)(k_2 - K_2) = \frac{(N-n)\{(n-1)(N-1)-6\}}{n(n-1)N(N+5)} K_5 + \frac{6(N-n)}{(n-1)(N+5)} K_3 K_2.$$

12.13 Show that in samples from a symmetrical finite population k_{2r} is uncorrelated with any k-statistic of odd order.

12.14 Show that if $M(z)$ is the moment-generating function of x, the m.g.f., say $M(\xi)$, of a function $y = f(x)$, may be written

$$M(\xi) = \left[\exp\left\{\xi f\left(\frac{d}{dz}\right)\right\} M(z)\right]_{z=0}$$

Hence show that the generating function of the moments of the k-statistics,

$$\Sigma\left\{\mu(p_1^{\pi_1} \ldots p_s^{\pi_s}) \frac{t_{p_1}^{\pi_1}}{\pi_1!} \cdots \frac{t_{p_s}^{\pi_s}}{\pi_s!}\right\}$$

is given by

$$\left[\exp\{t_1\mathbf{K}_1 + t_2\mathbf{K}_2 + t_3\mathbf{K}_3 + \ldots\} \exp\left\{\kappa_1 s_1 + \frac{\kappa_2 s_2}{2!} + \ldots\right\}\right]_{x=0}$$

where $\mathbf{K}_r$ is the same function of the operators $\dfrac{\partial}{\partial x}$ as k_r is of the observations x and $s_r = \Sigma(x^r)$.

Deduce that

$$\mathbf{K}_p(s_p) = p!$$
$$\mathbf{K}_p(s_{p_1} s_{p_2} \ldots) = 0$$

where $(p_1 p_2 \ldots)$ is any partition of p.

(Fisher, 1928)

12.15 Show that

$$E\{k_p(x_1 + h, x_2, \ldots, x_n)\} = \kappa_p + \frac{1}{n}h^p$$

and hence that

$$\mathbf{S}_p k_p = p!,$$
$$\mathbf{S}_q k_p = 0, \qquad q \neq p,$$

where $\mathbf{S}_p$ is the same function of the operators $\frac{\partial}{\partial x}$ as s_p is of the observations x.

(Kendall, 1940*a*)

12.16 Show that the generating function of the moments of the k-statistics is given by

$$\left[\exp\left\{\frac{\mathbf{S}_1\kappa_1}{1!}+\frac{\mathbf{S}_2\kappa_2}{2!}+\ldots\right\}\exp\{k_1t_1+k_2t_2+\ldots\}\right]_{x=0}$$

and hence derive the result of the previous exercise.

(Kendall, 1942)

12.17 Use Exercise 12.16 to show that, in the expression of k_p in terms of the symmetric sums s, the sum of the coefficients is $1/n$.

12.18 Show that an unbiassed estimator of $\mu_1'^4$ is

$$\frac{1}{n^{[4]}}\{s_1^4-6s_2s_1^2+3s_2^2+8s_3s_1-6s_4\}.$$

12.19 Show that the symmetric function $l_{rs\ldots}$ may be written as

$$\Sigma'\frac{(-1)^{\Sigma(\rho-1)}\,\Pi\,(\rho-1)!}{\Pi\,\{\pi_1!\,\pi_2!\ldots\}}\;\frac{r!\,s!\ldots}{\Pi\,\{(p_1!)^{\pi_1}(p_2!)^{\pi_2}\ldots\}}\;\frac{\{\Pi\,(p_1^{\pi_1}p_2^{\pi_2}\ldots)\}}{n^{(\Sigma\rho)}}$$

where $(p_1^{\pi_1}p_2^{\pi_2}\ldots)$ is a partition of r, $\Sigma p\pi = r$, having ρ parts; there are similar partitions of $s, \ldots$; the signs Π and Σ denote products and sums over all $r, s, \ldots$; and Σ' involves summation over all partitions.

Hence show that

$$l_{222} = -\frac{[1^6]}{n^{[6]}}+\frac{3\,[21^4]}{n^{[5]}}-\frac{3\,[2^21^2]}{n^{[4]}}+\frac{[2^3]}{n^{[3]}}$$

as given in Appendix Table 11.

12.20 By the method of Example 12.7 show that if k_1 and k_p are uncorrelated, for all $p>1$, the parent distribution is normal; and that if k_1 is independent of any $k_p (p>1)$ the parent is likewise normal. (Use Exercise 4.5.)

CHAPTER 13

APPROXIMATIONS TO SAMPLING DISTRIBUTIONS—(2)

13.1 In this chapter we shall discuss two main topics. The first is the extension to bivariate and multivariate statistics of the methods developed in the last chapter for the univariate case. The second is the structure of k-statistics and cumulants, with particular reference to the proof of the combinatorial rules for developing their sampling properties.

Bivariate k-statistics

13.2 Given any bipartite number pp' we shall have for any partition $\{(p_1p_1')^{\pi_1}(p_2p_2')^{\pi_2}\dots\}$ and the bivariate cumulant $\kappa_{pp'}$ a k-statistic $k_{pp'}$ whose mean value is $\kappa_{pp'}$. Explicitly

$$k_{pp'} = p!p'!\sum \frac{(-1)^{\rho-1}(\rho-1)!}{n^{[\rho]}} \sum \frac{(x_1^{p_1}y_1^{p_1'}x_2^{p_1}y_2^{p_1'}\dots x_\rho^{p_\rho}y_\rho^{p'_\rho})}{(p_1!)^{\pi_1}(p_1'!)^{\pi_1}\dots\pi_1!\pi_2!\dots}. \tag{13.1}$$

In particular, corresponding to (12.28) we have

$$\left.\begin{aligned}
k_{11} &= \frac{1}{n^{[2]}}(ns_{11}-s_{10}s_{01})\\
k_{21} &= \frac{1}{n^{[3]}}(n^2s_{21}-2ns_{10}s_{11}-ns_{20}s_{01}+2s_{10}^2s_{01})\\
k_{31} &= \frac{1}{n^{[4]}}\{n^2(n+1)s_{31}-n(n+1)s_{30}s_{01}-3n(n-1)s_{11}s_{20}\\
&\quad -3n(n+1)s_{21}s_{10}+6ns_{11}s_{10}^2+6ns_{20}s_{10}s_{01}-6s_{01}s_{10}^3\}\\
k_{22} &= \frac{n}{(n-1)(n-2)(n-3)}\Big\{(n+1)s_{22}-\frac{2(n+1)}{n}s_{21}s_{01}-\frac{2(n+1)}{n}s_{12}s_{10}\\
&\quad -\frac{(n-1)}{n}s_{20}s_{02}-\frac{2(n-1)}{n}s_{11}^2+\frac{8}{n}s_{11}s_{10}s_{01}+\frac{2}{n}s_{02}s_{10}^2+\frac{2}{n}s_{20}s_{01}^2\\
&\quad -\frac{6}{n^2}s_{10}^2s_{01}^2\Big\}
\end{aligned}\right\} \tag{13.2}$$

In generalization of the mean value functions of the k's we may write, for example,

$$\mathrm{E}(k_{20}k_{11}) = \mu'\begin{pmatrix}2 & 1\\ 0 & 1\end{pmatrix}$$

$$\mathrm{E}(k_{20}k_{11}k_{02}) = \mu'\begin{pmatrix}2 & 1 & 0\\ 0 & 1 & 2\end{pmatrix}$$

with corresponding κ's. The latter may be expressed in terms of the cumulants of the bivariate distribution as in the univariate; and the coefficients will now depend on partitions of bipartite numbers. Our rules still apply (and in particular the pattern functions appropriate to particular arrays are the same); but the numerical coefficients

associated with completed arrays are modified, for we now have to consider the number of ways of allocating two different sorts of individuals in a two-way partition of a bipartite number. An example will make the modification clear.

Suppose we wish to find the coefficient of $\kappa_{33}\kappa_{11}^2$ in $\kappa\begin{pmatrix}2 & 2 & 1\\ 2 & 2 & 1\end{pmatrix}$. The total degree is 10 and, the orders of the product being 6, 2, 2, we have to consider arrays of type

			6
			2
			2
4	4	2	10

i.e. those we discussed in Chapter 12. The pattern functions are those we have already found. For the numerical coefficients we have to regard the column totals as consisting of the two types of object in number (2, 2), (2, 2), and (1, 1) and the row totals (3, 3), (1, 1) and (1, 1). For instance, the array

2	2	2	6
1	1	.	2
1	1	.	2
4	4	2	10

might be written either as

(1, 1)	(1, 1)	(1, 1)	(3, 3)
(0, 1)	(1, 0)	.	(1, 1)
(1, 0)	(0, 1)	.	(1, 1)
(2, 2)	(2, 2)	(1, 1)	(5, 5)

(13.3)

or as

(2, 0)	(0, 2)	(1, 1)	(3, 3)
(0, 1)	(1, 0)	.	(1, 1)
(0, 1)	(1, 0)	.	(1, 1)
(2, 2)	(2, 2)	(1, 1)	(5, 5)

(13.4)

each of which will make a contribution to the numerical coefficient. It will be found that no other arrays are possible except those obtained by permuting the first two columns.

The numerical coefficient in (13.3) and the permuted array together is

$$2\left(\frac{2!}{1!1!}\right)\left(\frac{2!}{1!1!}\right)\left(\frac{2!}{1!1!}\right)\left(\frac{2!}{1!1!}\right)\cdot\frac{1}{2!} = 16.$$

That in (13.4) and the permuted array is

$$2\left(\frac{2!}{2!}\right)\left(\frac{2!}{1!1!}\right)\left(\frac{2!}{1!1!}\right)\left(\frac{2!}{2!}\right)\cdot\frac{1}{2!} = 4.$$

The total contribution is thus 20. The pattern function is $\dfrac{1}{(n-1)(n-2)}$.

In the same way it will be found that for the partitions

2	3	1	6		3	3	.	6
1	1	.	2		1	.	1	2
1	.	1	2		.	1	1	2
4	4	2	10		4	4	2	10

the coefficients are 48 and 8. Thus the desired coefficient of $\kappa_{33}\kappa_{11}^2$ is

$$\frac{20}{(n-1)(n-2)}+\frac{48}{(n-1)^2}+\frac{8}{(n-1)^2}=\frac{4(19n-33)}{(n-1)^2(n-2)}.$$

Example 13.1

To find an exact expression for the covariance of the estimates of variance of two correlated variables, i.e.

$$\kappa\begin{pmatrix}2 & 0\\ 0 & 2\end{pmatrix}.$$

This will clearly consist of three terms, in κ_{22}, $\kappa_{02}\kappa_{20}$ and κ_{11}^2. For the first we have the partition

(2, 0)	(0, 2)	(2, 2)
(2, 0)	(0, 2)	(2, 2)

with pattern function $\frac{1}{n}$ and numerical coefficient unity. For the second no contribution exists, the only arrangement being

(2, 0)	.	(2, 0)
.	(0, 2)	(0, 2)
(2, 0)	(0, 2)	(2, 2)

which has a vanishing pattern function. For the third term we have

(1, 0)	(0, 1)	(1, 1)
(1, 0)	(0, 1)	(1, 1)
(2, 0)	(0, 2)	(2, 2)

the pattern function for which is $\frac{1}{(n-1)}$ and numerical coefficient 2. Hence

$$\kappa\begin{pmatrix}2 & 0\\ 0 & 2\end{pmatrix}=\frac{1}{n}\kappa_{22}+\frac{2}{n-1}\kappa_{11}^2.$$

We note the analogy between this expression and the variance of k_{20}, which may be written (cf. (12.35))

$$\kappa\begin{pmatrix}2 & 2\\ 0 & 0\end{pmatrix}=\frac{\kappa_{40}}{n}+\frac{2\kappa_{20}^2}{n-1}.$$

In particular, if the population is normal all κ's except those of the second order

vanish. We then find

$$\text{correlation } (k_{20}, k_{02}) = \frac{\kappa_{11}^2}{\kappa_{20}\kappa_{02}} = \rho^2,$$

where ρ is the correlation parameter of the bivariate normal form.

13.3 It is possible to derive the bivariate formulae from the univariate formulae by a symbolic process given by Kendall (1940*c*). This was used by Cook (1951) to derive the bivariate formulae up to those of order 4 in each variate. Reference should be made to her paper for the details and the formulae and to a later paper for some applications. We quote here some of the simpler results.

$$\kappa\begin{pmatrix}2 & 1\\ 0 & 1\end{pmatrix} = \frac{1}{n}\kappa_{31} + \frac{2}{n-1}\kappa_{20}\kappa_{11} \tag{13.5}$$

$$\kappa\begin{pmatrix}2 & 0\\ 0 & 2\end{pmatrix} = \frac{1}{n}\kappa_{22} + \frac{2}{n-1}\kappa_{11}^2 \tag{13.6}$$

$$\kappa\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix} = \frac{1}{n}\kappa_{22} + \frac{1}{n-1}\kappa_{20}\kappa_{02} + \frac{1}{n-1}\kappa_{11}^2 \tag{13.7}$$

$$\kappa\begin{pmatrix}3 & 3\\ 0 & 0\end{pmatrix} = \frac{1}{n}\kappa_{60} + \frac{9}{n-1}\kappa_{40}\kappa_{20} + \frac{9}{n-1}\kappa_{30}^2 + \frac{6n}{(n-1)(n-2)}\kappa_{20}^3 \tag{13.8}$$

(this is, of course, a univariate result).

$$\kappa\begin{pmatrix}3 & 2\\ 0 & 1\end{pmatrix} = \frac{1}{n}\kappa_{51} + \frac{6}{n-1}\kappa_{31}\kappa_{20} + \frac{3}{n-1}\kappa_{40}\kappa_{11} + \frac{9}{n-1}\kappa_{30}\kappa_{21} + \frac{6n}{(n-1)(n-2)}\kappa_{20}^2\kappa_{11} \tag{13.9}$$

$$\kappa\begin{pmatrix}3 & 1\\ 0 & 2\end{pmatrix} = \frac{1}{n}\kappa_{42} + \frac{3}{n-1}\kappa_{22}\kappa_{20} + \frac{6}{n-1}\kappa_{31}\kappa_{11} + \frac{3}{n-1}\kappa_{30}\kappa_{12} + \frac{6}{n-1}\kappa_{21}^2 + \frac{6n}{(n-1)(n-2)}\kappa_{20}\kappa_{11}^2 \tag{13.10}$$

$$\kappa\begin{pmatrix}3 & 0\\ 0 & 3\end{pmatrix} = \frac{1}{n}\kappa_{33} + \frac{9}{n-1}\kappa_{22}\kappa_{11} + \frac{9}{n-1}\kappa_{12}\kappa_{21} + \frac{6n}{(n-1)(n-2)}\kappa_{11}^3 \tag{13.11}$$

$$\kappa\begin{pmatrix}2 & 2\\ 1 & 1\end{pmatrix} = \frac{1}{n}\kappa_{42} + \frac{4}{n-1}\kappa_{22}\kappa_{20} + \frac{4}{n-1}\kappa_{31}\kappa_{11} + \frac{1}{n-1}\kappa_{40}\kappa_{02} + \frac{4}{n-1}\kappa_{30}\kappa_{12} + \frac{5}{n-1}\kappa_{21}^2 + \frac{4n}{(n-1)(n-2)}\kappa_{20}\kappa_{11}^2 + \frac{2n}{(n-1)(n-2)}\kappa_{20}^2\kappa_{02} \tag{13.12}$$

$$\kappa\begin{pmatrix}2 & 1\\ 1 & 2\end{pmatrix} = \frac{1}{n}\kappa_{33} + \frac{5}{n-1}\kappa_{22}\kappa_{11} + \frac{2}{n-1}\kappa_{13}\kappa_{20} + \frac{2}{n-1}\kappa_{31}\kappa_{02} + \frac{1}{n-1}\kappa_{30}\kappa_{03} + \frac{8}{n-1}\kappa_{21}\kappa_{12} + \frac{4n}{(n-1)(n-2)}\kappa_{11}\kappa_{20}\kappa_{02} + \frac{2n}{(n-1)(n-2)}\kappa_{11}^3 \tag{13.13}$$

There is obviously a strong family resemblance among these formulae and we shall see later how they can be written in a concise form.

Example 13.2

The methods may similarly be extended to deal with bivariate finite populations. Formulae in this field are not required often enough to make it worth while listing

them, but it may be useful to illustrate the method of deriving them. Let us then derive the variance of k_{11} in samples of n from a finite population of size N. We have, after the manner of **12.21**,

$$\mathrm{E}(k_{11}-K_{11})^2 = \mathrm{E}(k_{11}^2)-\mathrm{E}(K_{11}^2). \tag{13.14}$$

From (13.7) we have

$$\mathrm{E}(k_{11}^2) = \mu'\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix} = \kappa\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}+\kappa_{11}^2$$

$$= \frac{1}{n}\kappa_{22}+\frac{1}{n-1}\kappa_{20}\kappa_{02}+\frac{1}{n-1}\kappa_{11}^2+\kappa_{11}^2, \tag{13.15}$$

$$\mathrm{E}(K_{11}^2) = \frac{1}{N}\kappa_{22}+\frac{1}{N-1}\kappa_{20}\kappa_{02}+\frac{1}{N-1}\kappa_{11}^2+\kappa_{11}^2. \tag{13.16}$$

Hence (13.14) becomes

$$\operatorname{var} k_{11} = \left(\frac{1}{n}-\frac{1}{N}\right)\kappa_{22}+\left(\frac{1}{n-1}-\frac{1}{N-1}\right)\kappa_{20}\kappa_{02}+\left(\frac{1}{n-1}-\frac{1}{N-1}\right)\kappa_{11}^2,$$

which reduces to

$$\frac{\operatorname{var} k_{11}}{N-n}-\frac{1}{nN}\kappa_{22} = \frac{1}{(n-1)(N-1)}\kappa_{20}\kappa_{02}+\frac{1}{(n-1)(N-1)}\kappa_{11}^2. \tag{13.17}$$

We now require to eliminate the non-linear terms on the right. From (13.16) we have

$$\mathrm{E}(K_{11}^2)-\frac{1}{N}\kappa_{22} = \frac{1}{N-1}\kappa_{20}\kappa_{02}+\frac{N}{N-1}\kappa_{11}^2. \tag{13.18}$$

From (13.6) in terms of K's, we likewise find

$$\mathrm{E}(K_{20}K_{02})-\frac{1}{N}\kappa_{22} = \kappa_{20}\kappa_{02}+\frac{2}{N-1}\kappa_{11}^2. \tag{13.19}$$

Hence, from (13.17), (13.18) and (13.19),

$$\begin{vmatrix} \dfrac{\operatorname{var} k_{11}}{N-n}-\dfrac{1}{nN}\kappa_{22} & \dfrac{1}{(n-1)(N-1)} & \dfrac{1}{(n-1)(N-1)} \\ \mathrm{E}(K_{11}^2)-\dfrac{1}{N}\kappa_{22} & \dfrac{1}{N-1} & \dfrac{N}{N-1} \\ \mathrm{E}(K_{20}K_{02})-\dfrac{1}{N}\kappa_{22} & 1 & \dfrac{2}{N-1} \end{vmatrix} = 0. \tag{13.20}$$

This reduces to the expression

$$\operatorname{var} k_{11} = \frac{N-n}{(n-1)(N+1)(N-2)}\left\{\frac{N-2}{Nn}(Nn-N-n-1)K_{22}\right.$$

$$\left.+(N-1)K_{20}K_{02}+(N-3)K_{11}^2\right\}. \tag{13.21}$$

In deriving this we have, in accordance with the Irwin-Kendall principle of **12.21**, replaced the expectations over the infinite population by those over the finite population of N; terms such as $\mathrm{E}(K_{11}^2)$ being then replaced by their constant value such as K_{11}^2. Equation (13.21) should be compared with the result of Exercise 12.11.

Proof of the combinatorial rules

13.4 We now proceed to prove the validity of the rules enunciated and exemplified above. Rules 1 and 2 have already been proved.

As a preliminary let us define an operator ∂_p such that

$$\left.\begin{aligned} \partial_p \mu'_r &= r(r-1)\dots(r-p+1)\mu'_{r-p}, \quad r>p \\ \partial_p \mu'_p &= p! \\ \partial_p \mu'_r &= 0, \qquad r<p, \end{aligned}\right\} \tag{13.22}$$

and

$$\partial_p(AB) = (\partial_p A)B + A(\partial_p B), \tag{13.23}$$

so that ∂ acting on a product is distributive.

In virtue of (13.23) we have

$$\begin{aligned} \partial_p(\mu'_r)^m &= m(\mu'_r)^{m-1}\partial_p \mu'_r \\ &= \frac{\partial}{\partial \mu'_r}(\mu'_r)^m \partial_p \mu'_r. \end{aligned}$$

It follows that if f is a polynomial function in the μ's

$$\partial_p f = \frac{\partial f}{\partial \mu'_1}\partial_p \mu'_1 + \frac{\partial f}{\partial \mu'_2}\partial_p \mu'_2 + \dots \tag{13.24}$$

and this also holds if f can be expanded in a series of polynomials in the μ's.

Now consider the expression defining the cumulants in terms of the moments, (3.30),

$$\exp\left(\kappa_1 t + \dots + \kappa_p \frac{t^p}{p!} + \dots\right) = 1 + \mu'_1 t + \dots \frac{\mu'_p t^p}{p!} + \dots$$

On operating on both sides by ∂_p there results

$$\exp\left(\kappa_1 t + \dots + \kappa_p \frac{t^p}{p!} + \dots\right)\left(\partial_p \kappa_1 t + \dots + \partial_p \kappa_p \frac{t^p}{p!} + \dots\right) = t^p + \mu'_1 t^{p+1} + \dots$$

$$= t^p\left(1 + \mu'_1 t + \dots + \mu'_p \frac{t^p}{p!} + \dots\right),$$

and hence

$$\partial_p \kappa_1 t + \dots + \partial_p \kappa_p \frac{t^p}{p!} + \dots = t^p.$$

This is an identity in t and hence

$$\left.\begin{aligned} \partial_p \kappa_p &= p! \\ \partial_q \kappa_p &= 0, \qquad q \neq p \end{aligned}\right\}. \tag{13.25}$$

For example,

$$\begin{aligned} \kappa_4 &= \mu'_4 - 4\mu'_3\mu'_1 - 3\mu'^2_2 + 12\mu'_2\mu'^2_1 - 6\mu'^4_1. \\ \partial_1 \kappa_4 &= 4\mu'_3 - 4\mu'_3 - 12\mu'_2\mu'_1 - 12\mu'_2\mu'_1 + 24\mu'^3_1 + 24\mu'_2\mu'_1 - 24\mu'^3_1 \\ &= 0. \\ \partial_2 \kappa_4 &= 12\mu'_2 - 24\mu'^2_1 - 12\mu'_2 + 24\mu'^2_1 \\ &= 0. \\ \partial_3 \kappa_4 &= 24\mu'_1 - 24\mu'_1 \\ &= 0. \\ \partial_4 \kappa_4 &= 4!. \end{aligned}$$

13.5 Now in accordance with Rule 1, which we have already established, $\kappa(a_1^{\alpha_1}\ldots a_s^{\alpha_s})$ and hence $\mu(a_1^{\alpha_1}\ldots a_s^{\alpha_s})$ may be expressed in terms of parent κ's by an equation of the form

$$\mu(a_1^{\alpha_1}a_2^{\alpha_2}\ldots) = \Sigma\{A(\kappa_{b_1}^{\beta_1}\kappa_{b_2}^{\beta_2}\ldots)\} \tag{13.26}$$

where A is a factor which it is our object to find. To save printing we omit the primes on the μ's for the rest of this proof. Operate on both sides of (13.26) by $(\partial_{b_1}^{\beta_1}\partial_{b_2}^{\beta_2}\ldots)$. Every term on the right is annihilated except that in $(\kappa_{b_1}^{\beta_1}\kappa_{b_2}^{\beta_2}\ldots)$ and we have

$$A.(b_1!)^{\beta_1}(b_2!)^{\beta_2}\ldots\beta_1!\beta_2!\ldots = (\partial_{b_1}^{\beta_1}\partial_{b_2}^{\beta_2}\ldots)\mu(a_1^{\alpha_1}a_2^{\alpha_2}\ldots). \tag{13.27}$$

We now consider an operator θ_p, analogous to ∂_p, which, when acting on a power of x (of any suffix), reduces the exponent by p and multiplies by $r(r-1)\ldots(r-p+1)$; and we will suppose the operator to be distributive.(*) Regarding $\mu(a_1^{\alpha_1}a_2^{\alpha_2}\ldots)$ as the mean value of $(k_{a_1}^{\alpha_1}k_{a_2}^{\alpha_2}\ldots)$ we see that the result of operating by the ∂'s on the mean value expressed in terms of μ's is the same as that given by taking the mean value of the operation of the θ's. But this latter operation results in a constant, which is equal to its mean value; and we thus have

$$A = \frac{(\theta_{b_1}^{\beta_1}\theta_{b_2}^{\beta_2}\ldots)}{(b_1!)^{\beta_1}(b_2!)^{\beta_2}\ldots\beta_1!\beta_2!\ldots}(k_{a_1}^{\alpha_1}k_{a_2}^{\alpha_2}\ldots). \tag{13.28}$$

Our rules are concerned with the evaluation of this operation.

13.6 Consider now a completed array of type (12.32). A little reflection will show that there is one such array for every term in (13.28) which does not vanish by operation, and that every term in (13.28) will have its corresponding completed array. The numbers in the body of the array are the powers of x occurring in the k-product; added horizontally they compose the orders of the operators; added vertically they compose the orders of the corresponding k's. A completed array is, so to speak, a chart of part of the operation; and the whole operation is the sum of all possible completed arrays.

The operation (13.28) gives us the coefficients in $\mu(a_1^{\alpha_1}a_2^{\alpha_2}\ldots)$, but we wish to find those in the corresponding $\kappa(a_1^{\alpha_1}a_2^{\alpha_2}\ldots)$. The necessary allowance is made by Rule 3, which we now prove; that is, the coefficient of $(\kappa_{b_1}^{\beta_1}\kappa_{b_2}^{\beta_2}\ldots)$ in $\kappa(a_1^{\alpha_1}a_2^{\alpha_2}\ldots)$ is given by all completed arrays, *ignoring those which are resolvable into separate blocks each confined to separate rows and columns.*

Referring to equation (12.31), expressing the relation between multivariate moments and cumulants, we see that $\kappa(a_1^{\alpha_1}a_2^{\alpha_2}\ldots)$ is the sum of terms composed of products of one, two, three . . . multivariate moments. The first term is $\mu(a_1^{\alpha_1}a_2^{\alpha_2}\ldots)$ itself. Consider a two-part term such as $\mu(a_1^{\alpha_1'}a_2^{\alpha_2'}\ldots)\mu(a_1^{\alpha_1''}a_2^{\alpha_2''}\ldots)$, where $\alpha_1'+\alpha_1'' = \alpha_1$, etc. Its coefficient in the expansion on the right-hand side of (12.31) is

$$-\tfrac{1}{2}\cdot\frac{2!}{1!1!}\frac{t_{a_1}^{\alpha_1'}}{\alpha_1'!\alpha_1''!}\frac{t_{a_2}^{\alpha_2'}}{\alpha_2'!\alpha_2''!}\cdots$$

(*) θ_p may be regarded as equivalent to $\left(\frac{\partial^p}{\partial x_1^p}+\frac{\partial^p}{\partial x_2^p}+\ldots+\frac{\partial^p}{\partial x_n^p}\right)$.

and hence the coefficient with which it appears in the formula for $\kappa(a_1^{\alpha_1}a_2^{\alpha_2}\dots)$ is

$$-\frac{\alpha_1!}{\alpha_1'!\alpha_1''!}\frac{\alpha_2!}{\alpha_2'!\alpha_2''!}\dots = -\binom{\alpha_1}{\alpha_1'}\binom{\alpha_2}{\alpha_2'}\dots \tag{13.29}$$

Now $\mu(a_1^{\alpha'_1}a_2^{\alpha'_2}\dots)$ will itself have an array of type (12.32) with column totals $(a_1^{\alpha'_1}a_2^{\alpha'_2}\dots)$ and row totals, say $(b_1^{\beta'_1}b_2^{\beta'_2}\dots)$; and similarly for $\mu(a_1^{\alpha''_1}a_2^{\alpha''_2}\dots)$. Provided that $\beta_1'+\beta_1'' = \beta_1$ these arrays will correspond to terms in the κ's which, when multiplied, will give a term in $(\kappa_{b_1}^{\beta_1}\kappa_{b_2}^{\beta_2}\dots)$. Thus the product of these terms may be considered as an array of type (12.32) with column totals $(a_1^{\alpha_1}a_2^{\alpha_2}\dots)$ and row totals $(b_1^{\beta_1}b_2^{\beta_2}\dots)$ and with the body of the table resolvable into two separate blocks. Since there are α_1 columns of total a_1, there will be $\binom{\alpha_1}{\alpha_1'}\binom{\alpha_2}{\alpha_2'}$ products of this type in the expression which gives $\mu(a_1^{\alpha_1}a_2^{\alpha_2}\dots)$. This factor is the same as (13.29) but of opposite sign. Hence, if we ignore the separate two-part blocks in the array for μ we shall have allowed for the products of two moments which must be subtracted from μ to give κ.

Now some of these separate blocks will themselves be separable into two blocks, and in subtracting them all from $\mu(a_1^{\alpha_1}a_2^{\alpha_2}\dots)$ we subtract too much. For example, if there are three separate blocks, L, M, N, we shall, by considering L and $(M+N)$ as two blocks, have subtracted L, M, N. We shall have done the same by considering M and $(L+N)$, and N and $(L+M)$ as two blocks. That is, we have subtracted $2L$, $2M$, $2N$ too much. We must restore these blocks to the array for μ. Such additions, summed over all blocks of three, will be found to equal the terms in the expansion of (12.31) which result from the product of three moments.

In restoring these blocks we restore too many of the cases where there are four separate blocks. These must be subtracted again, and correspond to the negative term in (12.31) involving the product of four moments. Proceeding in this way we establish Rule 3.(*)

13.7 Now we proceed to Rules 4, 5 and 6, which are the fundamental rules of the whole process. Consider again the array of type (12.32), to fix the ideas, say,

2	3	1	6
1	1	.	2
1	.	1	2
4	4	2	10

(13.30)

This array will represent a number of terms in the operation, each of which consists of the operation of θ_6 on a term $x^2.x^3.x$ (the first row), θ_2 on $x.x$ (the second row), and so on. Provided that the suffixes of the x's in any row are alike, every suffix of the x's will provide a term, for k_p contains terms with every distribution of powers (adding to p) and suffixes. There will, for instance, be terms of the following kind :—

$$\begin{array}{ccc} x_1^2 & x_1^3 & x_1 \\ x_2 & x_2 & . \\ x_3 & . & x_3 \end{array} \qquad \begin{array}{ccc} x_1^2 & x_1^3 & x_1 \\ x_1 & x_1 & . \\ x_2, & . & x_2, \end{array} \qquad \begin{array}{ccc} x_1^2 & x_1^3 & x_1 \\ x_1 & x_1 & . \\ x_1 & . & x_1. \end{array}$$

(*) Kaplan (1952) has adduced a statistical argument in support of Rule 3.

In fact, for any completed array, we have terms in which

(i) all the x's have the same suffix (n in number, one for each suffix),
(ii) all the x's but one row have the same suffix ($n(n-1)$ in number),
(iii) all the x's but two rows have the same suffix and the remaining two are the same ($n(n-1)$ in number),

and so on. These cases correspond to the various separations dealt with in Rule 5.

Now in case (i) the term in any column arises from the term in x^p in k_p and (apart from numerical factors which are considered presently) is n^{-1}, from equation (12.25). Hence any column which contains an entry contributes a factor n^{-1} and the total function of n arising from case (i) is the product of n and of (n^{-1}) to the power of the number of columns containing a non-zero entry.

Similarly in cases (ii) and (iii) the n-function for each separation is the product of $n(n-1)$ and, for each column, a factor in n^{-1} or $\frac{-1}{n(n-1)}$ according as the column contains non-zero entries in one or in both parts of the separation; and so on.

This explains the origin of the pattern function as described in Rule 6. But in order to establish that rule completely (and incidentally to establish Rules 4 and 5) we have to show that the numerical coefficients arising from each separation are the same. When this is done the validity of Rule 6 is demonstrated, for the separate contributions in n may be added together to give the pattern function and the whole multiplied by the numerical coefficient.

θ_1 may be considered as the operation of picking out an x from the operand in all possible ways and replacing it by unity. Similarly $\frac{\theta_p}{p!}$ may be regarded as picking out p x's with the same suffix and replacing them by unity. It is thus evident that operating on a k-product by a θ-product $\frac{\theta_{b_1}\theta_{b_2}\dots}{b_1!\,b_2!\dots}$ of the same degree will yield a result which is the number of ways in which sets of x's can be picked out of the k-product so that each set contains b_1 of one suffix, b_2 of a second suffix (which may be the same as the first), and so on.

Now consider the operation (13.28) in which the k's are expressed in the simplified form (12.26). The operations θ being distributive, we shall emerge from the operation with a sum of terms comprising all the possible ways in which the individual x's can be picked out of the k-product such that the row and column totals of the two-way array are satisfied. Consider the sets corresponding to a particular array, such as (13.30). The contribution to the total will consist of the ways of picking out individuals such that

(i) from the individuals in the first k_4 are chosen four in the partition (2, 1, 1),
(ii) from the second k_4 are chosen four in the partition (3, 1),
(iii) from the k_2 are chosen two in the partition (1, 1),
(iv) these are associated in all possible ways such that individuals in a row arise from the same suffix.

On consideration it will be seen that the total number of ways of doing this is the

number of ways of allocating the individuals from column totals as required by Rule 5 ; *and this is true whether sets of rows have the same suffix or not.*

Rules 5 and 6, and hence Rule 4, follow at once.

13.8 The remaining rules are ancillary.

Rule 7 follows from Rule 2. In fact, the pattern function is independent of the numbers composing the array, and the pattern with a row containing one element can therefore form the skeleton of an array in which that element is unity ; and this would entail the appearance of κ_1, which by Rule 2 is impossible.

Rule 8 follows from Rule 6. The column containing the single element appears in just one separate of all the separations, and the contributions to the pattern function are thus all multiplied by n^{-1} owing to its presence.

Rule 10 follows from Rule 8. The addition of a unit part is equivalent to the addition of an extra column containing unity. This multiplies all pattern functions by n^{-1}, leaves numerical coefficients unchanged and increases the suffix of every κ according to the row in which the unit appears.

13.9 There only remains to prove Rule 9. Note that any pattern function can be evaluated linearly in terms of the functions of the pattern obtained by omitting one of the columns. For example, consider the right-hand column of

$$\begin{array}{cccc} . & \times & \times & \times \\ \times & . & \times & \times \\ \times & \times & . & . \\ \times & \times & . & . \end{array} \qquad (13.31)$$

and the contributions to the pattern function from it. The 15 separations which are possible with four rows can be divided into two classes, that in which the two rows in the fourth column lie in the same separate and that in which they do not. In separations of the first type the contributions from the first three columns will be the contributions of all separations of

$$\begin{array}{ccc} \times & \times & \times \\ \times & \times & . \\ \times & \times & . \end{array} \qquad (13.32)$$

in which the first two rows are amalgamated. Considering the function of the first three rows

$$\begin{array}{ccc} . & \times & \times \\ \times & . & \times \\ \times & \times & . \\ \times & \times & . \end{array} \qquad (13.33)$$

in which amalgamation has not taken place, we see that the contribution consists of all contributions which do not occur in the first. Calling the first A and the second B, we see that the contribution is

$$\frac{1}{n}A - \frac{1}{n(n-1)}(B-A) = \frac{1}{n-1}A - \frac{1}{n(n-1)}B,$$

i.e. a linear function of the derived patterns A and B. The proof of the general result follows exactly the same lines.

Now if a pattern may be divided into two groups connected only by a single column we can reduce it step by step by omitting the other columns. We end up with this single column, and the pattern function of this column must vanish; for the column total a corresponds to k_a, whose mean value the one-column array expresses, and since by definition this mean value is κ_a no composite terms such as would be given by two rows or more can appear.

13.10 The adaptation of the argument to the bivariate case is fairly straightforward. The only rule requiring examination is the one giving the numerical coefficients attaching to the pattern function; and on examination it should be clear, from the structure of the bivariate k-statistics, that the procedure sketched in **13.2** applies. The use of an annihilating operator

$$\partial_{pq}\mu_{rs} = r^{[p]}s^{[q]}\mu_{r-p,s-q} \tag{13.34}$$

gives the required result.

We may also state without proof that similar methods apply to multivariate statistics of higher order than the bivariate. For trivariate statistics, for example, we should have to consider partitions into three-part members. Only the simpler results in such cases are required in practice and they can usually be written down from the univariate or bivariate cases from considerations of the type which we develop in the following paragraphs.

Summary of results

13.11 Let us review the results of this and the previous chapter. We have seen that, by the use of simple algebraic methods, we can obtain as many moments or cumulants as we please of any statistic which is expressible as a symmetric function of the observations. We have also noted that the working is greatly simplified by the use of k-statistics and cumulants.

It is reasonable to inquire how this simplification comes about. What are the structural properties of k-statistics and cumulants which give them these advantages, and do they represent the limit of achievement in sampling simplicity? The answer to the last question seems to be affirmative and it is instructive to consider why this should be so.

13.12 The definition of univariate cumulants in terms of moments by the identity in t

$$\exp\{\Sigma\kappa_p t^p/p!\} = \Sigma\mu_p t^p/p! \tag{13.35}$$

tends to obscure the essential structure of the cumulant-moment relationship. Let us, instead, consider the relation between multivariate cumulants and moments, as defined formally by (12.31). To fix the ideas, suppose we have four variates and require their product-moment μ'_{1111}. From (12.31) we find, on examining the coefficient in $t_1t_2t_3t_4$,

$$\begin{aligned}\mu'_{1111} = {} & \kappa_{1111}+(\kappa_{1110}\kappa_{0001}+\kappa_{1101}\kappa_{0010}+\kappa_{1011}\kappa_{0100}+\kappa_{0111}\kappa_{1000})\\ & +(\kappa_{1100}\kappa_{0011}+\kappa_{1010}\kappa_{0101}+\kappa_{0110}\kappa_{1001})\\ & +(\kappa_{1100}\kappa_{0010}\kappa_{0001}+\kappa_{1010}\kappa_{0100}\kappa_{0001}+\kappa_{1001}\kappa_{0100}\kappa_{0010}+\kappa_{0110}\kappa_{1000}\kappa_{0001}\\ & +\kappa_{0101}\kappa_{1000}\kappa_{0010}+\kappa_{0011}\kappa_{1000}\kappa_{0100})\\ & +\kappa_{1000}\kappa_{0100}\kappa_{0010}\kappa_{0001}. \qquad (13.36)\end{aligned}$$

Eventually for any set of four variates x_i, x_j, x_k, x_l we could write

$$\mu'_{ijkl} = \kappa_{ijkl}+(\kappa_{ijk}\kappa_l+\kappa_{ijl}\kappa_k+\ldots \text{ etc.}).$$

Such formulae exhibit the essential simplicity of the relationship. In the expression of μ's in terms of κ's every possibility on the right occurs exactly once. When we invert the relationship to give κ's in terms of μ's we have, analogously to (13.36),

$$\begin{aligned}\kappa_{1111} = {} & \mu'_{1111}-(\mu'_{1110}\mu'_{0001}+\mu'_{1101}\mu'_{0010}+\mu'_{1011}\mu'_{0100}+\mu'_{0111}\mu'_{1000})\\ & -(\mu'_{1100}\mu'_{0011}+\mu'_{1010}\mu'_{0101}+\mu'_{0110}\mu'_{1001})\\ & +2(\mu'_{1100}\mu'_{0010}\mu'_{0001}+ \text{ etc.})-6\mu'_{1000}\mu'_{0100}\mu'_{0010}\mu'_{0001}. \qquad (13.37)\end{aligned}$$

This type of formula differs from (13.36) only in the factor $(\rho-1)!(-1)^{\rho-1}$ attached to sets which contain ρ parts.

13.13 In particular, these formulae remain valid if two or more of the four variates are identical. If, for example, the first and second are identical, expressions such as κ_{1111} are written κ_{211} and become trivariate; and if in addition the third and fourth are identical the cumulant becomes κ_{22} and is bivariate. In the extreme case when all are identical it becomes κ_4. Thus we can obtain all the relations of order 4 for any number of variates by condensing (13.36) or (13.37) by amalgamating suffixes. Certain terms which are distinct in the originals then become the same and can be added. For example, from (13.36) we find

$$\begin{aligned}\mu'_{22} = {} & \kappa_{22}+2\kappa_{21}\kappa_{01}+2\kappa_{12}\kappa_{10}+\kappa_{20}\kappa_{02}+2\kappa_{11}^2+\kappa_{20}\kappa_{01}^2+\kappa_{02}\kappa_{10}^2+4\kappa_{11}\kappa_{10}\kappa_{01}\\ & +\kappa_{10}^2\kappa_{01}^2. \qquad (13.38)\end{aligned}$$

If we now add the suffixes we reach the known univariate form

$$\mu'_4 = \kappa_4+4\kappa_3\kappa_1+3\kappa_2^2+6\kappa_2\kappa_1^2+\kappa_1^4. \qquad (13.39)$$

The origin of the numerical coefficients in (13.39) is now clear. The symbolic methods of deriving multivariate from univariate formulae (e.g. **3.28**) are merely convenient shorthand ways of working the process of amalgamation in reverse, a procedure which is possible in virtue of the symmetry of the terms in (13.36) among the suffixes.

13.14 It thus appears that the formal definitions of type (12.31) and (13.35) involving the exponential function derive their utility from the enumerative algebraic property of the exponential as illustrated in (13.36); not, for instance, from the analytical properties of the exponential such as its being equal to its own derivative or its being the inverse of the logarithmic function. The numerical coefficients appearing in our relations between cumulants and moments are convenient summaries of relations which are structurally much simpler than they appear. Similar considerations apply to k-statistics. Apart from factors in $(-1)^{\rho-1}(\rho-1)!/n^{[\rho]}$ they are, as

we have noted (12.8), sums of products of the contributory variates x taken in all possible ways once and only once.

13.15 We can now summarize our formulae. For instance (13.36) can be written

$$\kappa_{ijkl} = \mu'_{ijkl} + \overset{4}{\Sigma}\mu'_i\mu'_{jkl} + \overset{3}{\Sigma}\mu'_{ij}\mu'_{kl} + \overset{6}{\Sigma}\mu'_i\mu'_j\mu'_{kl} + \mu'_i\mu'_j\mu_k\mu'_l, \tag{13.40}$$

where the summations occur over all ways of grouping the subscripts, which are shown in the numbers above the summation signs. (There is a slight change of notation here : μ'_i refers to the mean value of the ith variate, not the mean of the ith power of some variate x.)

Likewise, if $s_{ij\ldots l}$ refers to the products $x_i x_j \ldots x$ summed over the sample, we have for the k-statistics

$$\left.\begin{aligned} k_i &= s_i/n \\ k_{ij} &= (n s_{ij} - s_i s_j)/n^{[2]} \\ k_{ijk} &= (n^2 s_{ijk} - n\overset{3}{\Sigma} s_i s_{jk} + 2 s_i s_j s_k)/n^{[3]} \end{aligned}\right\} \tag{13.41}$$

and so on.

For example, if the ith and jth variates are identical we find, in the more familiar notation,

$$k_2 = (n s_2 - s_1^2)/\{n(n-1)\} \tag{13.42}$$

whereas the bivariate k_{11} is given by

$$k_{11} = (n s_{11} - s_{10} s_{01})/\{n(n-1)\}. \tag{13.43}$$

If we require the bivariate k_{21} we amalgamate i and j and obtain from the third formula of (13.41)

$$k_{21} = \{n^2 s_{21} - 2n s_{10} s_{11} - n s_{20} s_{01} + 2 s_{10}^2 s_{01}\}/n^{[3]}$$

as given in (13.2).

13.16 This notation, which was suggested by Kaplan (1952) as a tensor notation, enables us to summarize the multivariate formulae in little more space than is required for the univariate results. For instance, for four variates we have for the covariance of two pairs

$$\begin{aligned} \kappa(ab, ij) &= \mathrm{E}\{(k_{ab} - \kappa_{ab})(k_{ij} - \kappa_{ij})\} \\ &= \frac{1}{n}\kappa_{abij} + \frac{1}{n-1}(\kappa_{ai}\kappa_{bj} + \kappa_{aj}\kappa_{bi}). \end{aligned} \tag{13.44}$$

This embodies seven types of formulae. If all variates are the same (a, b, i, j referring to one variate) we have

$$\operatorname{var} k_2 = \kappa(2^2) = \frac{1}{n}\kappa_4 + \frac{2}{n-1}\kappa_2^2. \tag{13.45}$$

If variate a is the same as i and variate b the same as j, we have

$$\operatorname{var} k_{11} = \kappa\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \frac{1}{n}\kappa_{22} + \frac{1}{n-1}\kappa_{20}\kappa_{02} + \frac{1}{n-1}\kappa_{11}^2. \tag{13.46}$$

If a and b are the same, i and j the same, we find

$$\operatorname{cov}(k_{20}, k_{02}) = \kappa\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} = \frac{1}{n}\kappa_{22} + \frac{2}{n-1}\kappa_{11}^2. \tag{13.47}$$

If a, b and i are the same and j is different,

$$\operatorname{cov}(k_{20}, k_{11}) = \kappa\begin{pmatrix}2 & 1\\ 0 & 1\end{pmatrix} = \frac{1}{n}\kappa_{31} + \frac{2}{n-1}\kappa_{11}\kappa_{20}. \tag{13.48}$$

If a and b are the same, i, j are different,

$$\operatorname{cov}(k_{200}, k_{011}) = \kappa\begin{pmatrix}2 & 0\\ 0 & 1\\ 0 & 1\end{pmatrix} = \frac{1}{n}\kappa_{211} + \frac{2}{n-1}\kappa_{110}\kappa_{101}. \tag{13.49}$$

If a and i are the same, b, j are different,

$$\operatorname{cov}(k_{110}, k_{101}) = \kappa\begin{pmatrix}1 & 1\\ 1 & 0\\ 0 & 1\end{pmatrix} = \frac{1}{n}\kappa_{211} + \frac{1}{n-1}(\kappa_{200}\kappa_{011} + \kappa_{110}\kappa_{101}). \tag{13.50}$$

Finally, if all four are different,

$$\operatorname{cov}(k_{1100}, k_{0011}) = \kappa\begin{pmatrix}1 & 0\\ 1 & 0\\ 0 & 1\\ 0 & 1\end{pmatrix} = \frac{1}{n}\kappa_{1111} + \frac{1}{n-1}(\kappa_{1010}\kappa_{0101} + \kappa_{1001}\kappa_{0110}). \tag{13.51}$$

13.17 The following formulae summarize further multivariate results.

$$\kappa(ab, ijk) = \kappa_{abijk}/n + \overset{6}{\Sigma}\kappa_{ai}\kappa_{bjk}/(n-1). \tag{13.52}$$

$$\kappa(ab, ijkl) = \kappa_{abijkl}/n + (\overset{8}{\Sigma}\kappa_{ai}\kappa_{bjkl} + \overset{6}{\Sigma}\kappa_{aij}\kappa_{bkl})/(n-1). \tag{13.53}$$

$$\begin{aligned}\kappa(ab, ij, pq) = {} & \kappa_{abijpq}/n^2 + \overset{12}{\Sigma}\kappa_{abip}\kappa_{jq}/\{n(n-1)\} + \overset{4}{\Sigma}\kappa_{aip}\kappa_{bjq}(n-2)/\{n(n-1)^2\} \\ & + \overset{8}{\Sigma}\kappa_{ai}\kappa_{bp}\kappa_{jq}/(n-1)^2.\end{aligned} \tag{13.54}$$

$$\begin{aligned}\kappa(abc, ijk) = {} & \kappa_{abcijk}/n + (\overset{9}{\Sigma}\kappa_{ai}\kappa_{bcjk} + \overset{9}{\Sigma}\kappa_{abi}\kappa_{cjk})/(n-1) \\ & + \overset{6}{\Sigma}\kappa_{ai}\kappa_{bj}\kappa_{ck}n/\{(n-1)(n-2)\}.\end{aligned} \tag{13.55}$$

$$\begin{aligned}\kappa(abc, ijkl) = {} & \kappa_{abcijkl}/n + \overset{12}{\Sigma}\kappa_{ai}\kappa_{bcjkl}/(n-1) + (\overset{12}{\Sigma}\kappa_{abi}\kappa_{cjkl} + \overset{18}{\Sigma}\kappa_{aij}\kappa_{bckl})/(n-1) \\ & + \overset{36}{\Sigma}\kappa_{ai}\kappa_{bj}\kappa_{ckl}n/\{(n-1)(n-2)\}.\end{aligned} \tag{13.56}$$

$$\begin{aligned}\kappa(ab, ij, pq, uv) = {} & \kappa_{abijpquv}/n + \overset{24}{\Sigma}\kappa_{ai}\kappa_{bjpquv}/\{n^2(n-1)\} + \overset{32}{\Sigma}\kappa_{aip}\kappa_{bjquv}(n-2)/\{n^2(n-1)^2\} \\ & + \overset{8}{\Sigma}\kappa_{aipu}\kappa_{bjqv}(n^2-3n+3)/\{n^2(n-1)^3\} + 3\overset{24}{\Sigma}\kappa_{abpu}\kappa_{ijqv}/\{n^2(n-1)\} \\ & + (\overset{96}{\Sigma}\kappa_{ai}\kappa_{bp}\kappa_{jquv} + \overset{48}{\Sigma}\kappa_{ai}\kappa_{pu}\kappa_{bjqv})/\{n(n-1)^2\} \\ & + \overset{96}{\Sigma}\kappa_{ai}\kappa_{bpu}\kappa_{jqv}(n-2)/\{n(n-1)^3\} + \overset{48}{\Sigma}\kappa_{bi}\kappa_{jp}\kappa_{qu}\kappa_{va}/(n-1)^3.\end{aligned} \tag{13.57}$$

$$\kappa(abcd,ijkl) = \kappa_{abcdijkl}/n + \overset{16}{\Sigma}\kappa_{ai}\kappa_{bcdjkl}/(n-1) + \overset{48}{\Sigma}\kappa_{abi}\kappa_{cdjkl}/(n-1)$$
$$+ \overset{72}{\Sigma}\kappa_{ai}\kappa_{bj}\kappa_{cdkl}n/\{(n-1)(n-2)\} + (\overset{16}{\Sigma}\kappa_{aijk}\kappa_{bcdl}$$
$$+ \overset{18}{\Sigma}\kappa_{abij}\kappa_{cdkl})/(n-1) + \overset{144}{\Sigma}\kappa_{ai}\kappa_{bcj}\kappa_{dkl}n/\{(n-1)(n-2)\}$$
$$+ \overset{24}{\Sigma}\kappa_{ai}\kappa_{bj}\kappa_{ck}\kappa_{dl}n(n+1)/\{(n-1)(n-2)(n-3)\}. \qquad (13.58)$$

These formulae embrace 41 formulae given by Cook (1951) and a number of others besides, but as Kaplan remarks, some care is required in deriving particular cases.

Example 13.3

For the variance of k_{11} from an infinite population we have, as in Example 13.5,

$$\operatorname{var} k_{11} = \frac{1}{n}\kappa_{22} + \frac{1}{n-1}\kappa_{20}\kappa_{02} + \frac{1}{n-1}\kappa_{11}^2.$$

Let us specialize this to the case of a normal parent in which the variances are unity. We then find

$$\operatorname{var} k_{11} = \frac{1}{n-1}(1+\rho^2). \qquad (13.29)$$

Let us find the third and fourth cumulants of k_{11}.

The third cumulant is given by (13.54) with a, i, p the same and b, j, q the same. We find

$$\kappa\begin{pmatrix}1 & 1 & 1\\ 1 & 1 & 1\end{pmatrix} = \frac{1}{n^2}\kappa_{33} + \frac{1}{n(n-1)}\{6\kappa_{22}\kappa_{11} + 3\kappa_{20}\kappa_{13} + 3\kappa_{31}\kappa_{02}\}$$
$$+ \frac{n-2}{n(n-1)^2}\{3\kappa_{21}\kappa_{12} + \kappa_{30}\kappa_{03}\} + \frac{1}{(n-1)^2}\{2\kappa_{11}^3 + 6\kappa_{11}\kappa_{20}\kappa_{02}\}. \qquad (13.60)$$

For the normal case the only surviving terms yield

$$\kappa_3(k_{11}) = \frac{1}{(n-1)^2}(2\rho^3 + 6\rho). \qquad (13.61)$$

Likewise we find from (13.57), taking a, i, p, u the same and b, j, q, v the same, the surviving terms, which can only arise from the concluding term,

$$\kappa_4(k_{11}) = \frac{1}{(n-1)^3}\{36\rho^2 + 6 + 6\rho^4\}. \qquad (13.62)$$

Thus, for instance, the kurtosis of the distribution of k_{11} is given by (13.59) and (13.62) as

$$\beta_2 - 3 = \frac{6}{n-1}\left(1 + \frac{4\rho^2}{(1+\rho^2)^2}\right). \qquad (13.63)$$

EXERCISES

13.1 Show that the correlation between k_{30} and k_{03} in samples of n from a bivariate normal population is ρ^3, where ρ is the correlation parameter of the population.

13.2 Show generally that, in the circumstances of Exercise 13.1, the correlation between k_{ro} and k_{or} is ρ^r.

13.3 A four-variate normal population has variates x_1, x_2, x_3, x_4 with unit variances and the correlation of x_i and x_j is ρ_{ij}. Show that the correlation between the sample covariance of x_1, x_2 and that of x_3, x_4 is given by

$$\frac{\rho_{13}\rho_{24}+\rho_{14}\rho_{23}}{\sqrt{(1+\rho_{12}^2)(1+\rho_{34}^2)}}.$$

13.4 Derive the following formula

$$\kappa\begin{pmatrix}1&1&1\\1&1&1\end{pmatrix}=\frac{1}{n^2}\kappa_{33}+\frac{6}{n(n-1)}\kappa_{22}\kappa_{11}+\frac{3}{n(n-1)}\kappa_{20}\kappa_{13}+\frac{3}{n(n-1)}\kappa_{31}\kappa_{02}+\frac{3(n-2)}{n(n-1)^2}\kappa_{21}\kappa_{12}$$
$$+\frac{n-2}{n(n-1)^2}\kappa_{30}\kappa_{03}+\frac{2}{(n-1)^2}\kappa_{11}^3+\frac{6}{(n-1)^2}\kappa_{11}\kappa_{20}\kappa_{02}$$

(Cook, 1951)

13.5 Use the methods of this chapter to derive the large-sample result for the variance of the correlation coefficient in bivariate normal samples

$$\text{var } r = (1-\rho^2)^2/n.$$

13.6 If x is a random variable with finite cumulants of all orders satisfying $\kappa_r = O(n^{-r+1})$ show that the transformed variate $y = f(x)$ has cumulants which also satisfy this condition. (This does not preclude the cumulants from being of lower order, i.e. $o(n^{-r+1})$.)

(James, 1955, *Biometrika*, **42**, 529)

13.7 Show that for a bivariate normal population k_{tu} and k_{vw} have zero covariance unless $t+u = v+w$.

(Wishart, 1929)

13.8 Show that for a bivariate normal population

$$dF = \frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{\frac{1}{2}}}\exp\left\{\frac{-1}{2(1-\rho^2)}\left(\frac{x_1^2}{\sigma_1^2}-\frac{2\rho x_1x_2}{\sigma_1\sigma_2}+\frac{x_2^2}{\sigma_2^2}\right)\right\}dx_1\,dx_2,$$

$$\text{var}\,(k_{tu}) = \kappa\begin{pmatrix}t&t\\u&u\end{pmatrix} = t!\,u!\sum_{j=1}^{t+u}\frac{(j-1)!}{j}\frac{\Delta^j 0^{t+u}}{n^{[j]}}\sigma_1^{2t}\sigma_2^{2u}F(-t,-u,1,\rho^2),$$

where $\Delta^j 0^k$ is the jth difference of the kth power of zero and F refers to the hypergeometric function.

(Wishart, 1929)

13.9 In the notation of **13.15**, show that

$$k_{ijkl} = \{n^2(n+1)s_{ijkl}-n(n+1)\sum^4 s_i s_{jkl}-n(n-1)\sum^3 s_{ij}s_{kl}+2n\sum^6 s_i s_j s_{kl}-6s_i s_j s_k s_l\}/n^{[4]},$$

and hence verify the formulae for k_{31} and k_{22} of (13.2).

(Kaplan, 1952)

13.10 Referring to the result of **13.9**, show that for a normal parent population the effect of adding a new part 2 to $\kappa(a_1^{\alpha_1} \ldots a_s^{\alpha_s})$ is to give pattern functions $1/(n-1)$ times those of the original. Show also that the effect on the numerical coefficient in an array is to multiply by twice the number of rows in the array. Deduce that the effect of adding a new part 2 is equivalent to operating by

$$\frac{2\kappa_2^2}{n-1}\frac{d}{d\kappa_2}.$$

(Fisher and Wishart, 1931)

13.11 Use the previous exercise to establish equations (12.71) and (12.72).

13.12 In generalization of Exercise 13.10 show that for a multivariate normal parent the effect of adding a covariance k_{pq} (p, q referring to the pth and qth variates) is equivalent to operating by

$$\sum_{rs}\frac{1}{(n-1)}(\kappa_{pr}\kappa_{qs}+\kappa_{ps}\kappa_{qr})\frac{d}{d\kappa_{rs}},$$

where κ_{pq} is the covariance of the variates p and q.

(Fisher and Wishart, 1931)

13.13 Show that in normal samples

$$\text{var}(m_r/m_2^{\frac{1}{2}r}) \sim r!/n.$$

CHAPTER 14

ORDER-STATISTICS

14.1 In Chapters 10 and 11 we have briefly discussed some of the sampling properties of medians, quantiles and, more generally, the order-statistics of a sample of observations. It will be remembered that the rth order-statistic of a sample of n is simply the rth smallest variate-value in the sample. We shall denote it, where necessary, by a suffix in parentheses : $x_{(r)}$. Unlike statistics which are based on symmetric functions of the observations, the order-statistics do not lend themselves to investigation by the methods we have been using in Chapters 12 and 13. They have the distinctive property that if we take an ordinary random sample of values $x_1, x_2, \ldots, x_n$ and rearrange them in order, as $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$, the values so derived are no longer independent even though the original x's were so. We may therefore expect to encounter special features in their sampling distributions.

14.2 Notwithstanding lack of symmetry and independence, some remarkably simple results can be obtained. We have seen in **10.10** that, in samples of size n from any distribution function $F(x)$ with a continuous frequency function $f(x)$, the sampling distribution of $x_{(r)}$ is given by G_r, say, where

$$dG_r = \frac{\{F(x_{(r)})\}^{r-1}\{1-F(x_{(r)})\}^{n-r}dF(x_{(r)})}{B(r, n-r+1)}. \tag{14.1}$$

Likewise, the joint distribution of $x_{(r)}$ and $x_{(s)}$, $r<s$, is given by

$$dG_{r,s} = \frac{\{F(x_{(r)})\}^{r-1}\{F(x_{(s)})-F(x_{(r)})\}^{s-r-1}\{1-F(x_{(s)})\}^{n-s}dF(x_{(r)})\,dF(x_{(s)})}{B(r, s-r)\,B(s, n-s+1)}. \tag{14.2}$$

The generalization to the joint distribution of several order-statistics is immediate. For instance, for $x_{(r)}, x_{(s)}, x_{(t)}, x_{(u)}$, $r<s<t<u$, we have, writing F_r for $F(x_{(r)})$, etc.,

$$dG_{r,s,t,u} = \frac{F_r^{r-1}(F_s-F_r)^{s-r-1}(F_t-F_s)^{t-s-1}(F_u-F_t)^{u-t-1}(1-F_u)^{n-u}dF_r\,dF_s\,dF_t\,dF_u}{B(r, s-r)\,B(s, t-s)\,B(t, u-t)\,B(u, n-u+1)}. \tag{14.3}$$

14.3 Distributions of the type (14.1) have been studied for various given forms of the parental distribution $F(x)$. When the parent is rectangularly distributed in the range 0 to 1, $F(x) = x$ and the distribution of $x_{(r)}$ is particularly simple, reducing to the Beta-distribution itself. A second type of parent leading to relatively simple results is the exponential $dF = \exp(-x/\sigma)\,dx/\sigma$. For others, and in particular the normal parent, the evaluation of exact results requires, in general, the quadrature of certain integrals.

Median and quartiles in the normal case

14.4 Suppose that the sample size is odd, say $n = 2r+1$, so that the median is the $(r+1)$th order-statistic. Let us write

$$A_x = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2}\, dt. \tag{14.4}$$

Then the distribution of the median, from (14.1), is seen to be (writing A for A_x),

$$dG = n!\, A^r (1-A)^r\, dA/(r!)^2. \tag{14.5}$$

From symmetry, the expected value of the median is clearly zero, and hence its variance is given by

$$\operatorname{var} m_e = \int x^2\, dG = \frac{n!}{(r!)^2} \int_{-\infty}^{\infty} x^2 A^r (1-A)^r . \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2}\, dx. \tag{14.6}$$

We expand the term $(1-A)^r$ binomially to obtain

$$\operatorname{var} m_e = \frac{n!}{(r!)^2} \int_{-\infty}^{\infty} \sum_{j=0}^{r} (-1)^j \binom{r}{j} A^{r+j} \frac{1}{\sqrt{(2\pi)}} x^2 e^{-\frac{1}{2}x^2}\, dx. \tag{14.7}$$

Each term may be integrated by parts, taking $x e^{-\frac{1}{2}x^2}$ for one part. We have

$$\frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} x A^{r+j} . x e^{-\frac{1}{2}x^2}\, dx = \left[\frac{-x A^{r+j} e^{-\frac{1}{2}x^2}}{\sqrt{(2\pi)}} \right]_{-\infty}^{\infty} + \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} \left\{ A^{r+j} + (r+j) A^{r+j-1} \frac{1}{\sqrt{(2\pi)}} x e^{-\frac{1}{2}x^2} \right\} dx. \tag{14.8}$$

The first term on the right vanishes. With a further integration by parts on the third term we find, for the integral on the left of (14.8),

$$\frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} A^{r+j}\, dx + \frac{(r+j)(r+j-1)}{4\pi} \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} A^{r+j-2} \exp\left(-\tfrac{3}{2}x^2\right) dx. \tag{14.9}$$

In (14.9), the first integral is simply $1/(r+j+1)$ and on substitution in (14.7) we find

$$\operatorname{var} m_e = \frac{n!}{(r!)^2} \sum_{j=0}^{r} (-1)^j \binom{r}{j} \frac{1}{r+j+1} + \frac{n!}{(r!)^2 4\pi} \sum_{j=0}^{r} (-1)^j \binom{r}{j} (r+j)(r+j-1) \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} A^{r+j-2} \exp\left(-\tfrac{3}{2}x^2\right) dx. \tag{14.10}$$

The first summation on the right of (14.10) yields unity, for

$$\sum_{j=0}^{r} (-1)^j \binom{r}{j} \frac{1}{r+j+1} = \int_0^1 \Sigma (-1)^j \binom{r}{j} t^{r+j}\, dt = \int_0^1 t^r (1-t)^r\, dt = B(r+1, r+1). \tag{14.11}$$

Thus we have, from (14.10) and (14.11),

$$\operatorname{var} m_e = 1 + \frac{n!}{4\pi (r!)^2} \Sigma (-1)^j \binom{r}{j} (r+j)(r+j-1)\, T_{r+j-2}, \tag{14.12}$$

where

$$T_{r+j-2} = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} A^{r+j-2} \exp\left(-\tfrac{3}{2}x^2\right) dx. \tag{14.13}$$

14.5 In this kind of way we can ascertain as many moments as we wish of any order-statistic. The results depend on integrals of type

$$\int_{-\infty}^{\infty} A^p \exp(-\tfrac{1}{2}qx^2)\,dx, \tag{14.14}$$

of which (14.13) is a particular case. These are sometimes known as Hojo's integrals, from the name of the author who first (1931) studied them in detail and tabulated them for a range of values of p and q. By an easy extension of the method, we may ascertain product-moments of two order-statistics, and hence the moments of such quantities as the interquartile range (the variate-distance between the upper and lower quartile) and the mid-range (half the sum of the smallest and greatest values). The quadratures involved are extensive and for ordinary purposes the methods of approximation given below are adequate. We may, however, quote a few of Hojo's exact results by way of illustration.

14.6 In normal samples the standard error of the mean is $\sigma/\sqrt{n}$. The standard error of the median is always larger than this, say $c_n\sigma/\sqrt{n}$ where $c_n \geqslant 1$ (and is only equal to unity for the trivial case $n = 2$). With a definition of the median for even $n = 2r$ as $\frac{1}{2}(x_{(r)}+x_{(r+1)})$ we have the following values for c_n :—

n	c_n	n	c_n
2	1·000	10	1·177
4	1·092	12	1·189
6	1·135	20	1·214
8	1·160	∞	1·253

Likewise, for the mid-range $\frac{1}{2}(x_{(1)} + x_{(n)})$ we have a standard error, say $d_n\sigma/\sqrt{n}$, but d_n tends to infinity with n, as we shall see below (Example 14.4). For increasing sample size the mid-range is an increasingly inferior estimator of the mean of the normal population. The following are some values for lower n :—

n	d_n	n	d_n
2	1·000	10	1·362
4	1·092	20	1·691
6	1·190	∞	∞

14.7 The interquartile distance, divided by twice 0·67449 (the distance between the quartiles in a normal population with unit variance), is sometimes used as an estimator of the parent standard deviation. It is interesting to compare the standard error of this quantity with the standard error of the sample s.d., namely $\sigma/\sqrt{(2n)}$. If the standard error is $e_n\sigma/\sqrt{(2n)}$, the following are some values of e_n :—

n	e_n	n	e_n
2	1·000	10	1·497
4	1·047	12	1·419
6	1·421	∞	1·649
8	1·313		

The values proceed somewhat irregularly owing to the arbitrary element in the definition of quartiles for sample-numbers not of the form $4r+1$.

14.8 Hojo (1931, 1933) and K. Pearson (1931) give a number of further results of this character. For some more recent numerical work reference may be made to Hastings and others (1947) and to Godwin (1949). Reference is made later in this chapter to work on extreme values. See also **14.10** below.

Pearson's expansion

14.9 In Chapter 6 we have considered the expansion of a distribution function $F(x)$ or a frequency function $f(x)$ as a power series in the variate x. We now consider an inverse expansion of x in terms of F, due to Karl Pearson. Let X_r be the parent value such that

$$F(X_r) = \frac{r}{n+1}. \tag{14.15}$$

We expand $x_{(r)}$ about X_r in a Taylor series

$$x_{(r)} = X_r + h_r X_r' + \frac{1}{2!} h_r^2 X_r'' + \dots \tag{14.16}$$

where

$$\left.\begin{aligned} h_r &= F(x_{(r)}) - F(X_r) = F_r - r/(n+1), \\ X_r' &= \frac{dX_r}{dF} = \left.\frac{dx}{dF}\right|_{x=X_r}, \\ X_r'' &= \left.\frac{d^2x}{dF^2}\right|_{x=X_r}. \end{aligned}\right\} \tag{14.17}$$

and so on.

From (14.16) we can express powers of x in series of powers of h; and the expectation of any power of h is easily derivable from (14.1). Provided, then, that our series converge in a suitable manner (or, more generally, give good approximations of an asymptotic kind) we may derive approximations to as many moments as we wish of any order-statistic.

Example 14.1

Consider again the distribution of the median—the $(r+1)$th value—in odd samples $(n = 2r+1)$ from the normal distribution.

We find, since $(r+1)/(n+1) = \frac{1}{2}, X_{r+1} = 0$,

$$\begin{aligned} X_r' &= \left.\frac{dx}{dF}\right|_{x=0} = 1\Big/\left.\frac{dF}{dx}\right|_{x=0} = \sqrt{(2\pi)} \\ X_r'' &= \frac{d}{dF}\left(1\Big/\left.\frac{dF}{dx}\right)\right|_{x=0} = \frac{dx}{dF}\cdot\frac{d}{dx}\left(1\Big/\left.\frac{dF}{dx}\right)\right|_{x=0} \\ &= \sqrt{(2\pi)}.0 \\ &= 0. \end{aligned}$$

Similarly $X_r''' = (2\pi)^{3/2}$. Thus, from (14.16) and (14.4) with $h_{r+1} = A - \frac{1}{2}$, we have

$$X_{me} = x_{(r+1)} = (2\pi)^{\frac{1}{2}}(A-\tfrac{1}{2}) + \tfrac{1}{6}(2\pi)^{3/2}(A-\tfrac{1}{2})^3 + \dots$$

and

$$x_{me}^2 = 2\pi(A-\tfrac{1}{2})^2 + \tfrac{1}{3}(2\pi)^2(A-\tfrac{1}{2})^4 + \dots$$

Substituting the expansion to this order for x_{me}^2 in (14.6), we find

$$\text{var } m_e = \frac{n!}{(r!)^2}\int_{-\infty}^{\infty} A^r(1-A)^r\{2\pi(A-\tfrac{1}{2})^2+\tfrac{1}{3}(2\pi)^2(A-\tfrac{1}{2})^4\}\,dA,$$

which reduces to

$$\text{var } m_e = \frac{\pi}{2(n+2)}+\frac{\pi^2}{4(n+2)(n+4)}. \tag{14.18}$$

This formula is accurate enough for ordinary purposes. For $n = 11$ it gives a variance of 0·137 agreeing to 3 d.p. with the known true value of 0·137,227.

Example 14.2

For the rectangular population

$$dF = dx, \qquad 0 \leqslant x \leqslant 1,$$

we find, for the median in odd samples,

$$X' = 1, \qquad X^{(r)} = 0, \qquad r > 1.$$

The variance of the median is then the expectation of $(x-\frac{1}{2})^2$ in

$$dG = \frac{n!}{(r!)^2}x^r(1-x)^r\,dx, \qquad 0 \leqslant x \leqslant 1,$$

and this is obviously an exact result. We find

$$\text{var } m_e = \frac{1}{4(n+2)}. \tag{14.19}$$

14.10 David and Johnson (1954) have pursued this subject systematically and give expansions for cumulants and product-cumulants of order-statistics up to and including the fourth-order cumulants and degree three in $1/(n+2)$. They choose expansions in terms of $(n+2)^{-1}$ rather than n^{-1} because of the natural appearance of the former quantity in elementary cases—cf. equations (14.18) and (14.19). It is not clear how far these expansions have the required asymptotic properties, but they appear to work well in practice.

The expected values of $x_{(r)}$ in samples of n from a normal population are often required in ranking studies and have been tabulated fairly extensively. See *Biometrika Tables for Statisticians*, Part 1, which gives the means of all normal order-statistics for $n = 2(1)26(2)50$ to at least two decimal places. Harter (1961, *Biom.*, **48**, 151) gives them to 5 d.p. for $n = 2(1)100(25)250(50)400$. Teichroew (1956) gives the means, and Sarhan and Greenberg (1956) give the variances and covariances, of all normal order-statistics to 10 decimal places for $n = 2(1)20$.

14.11 In considering the asymptotic distributions of the order-statistics we may have two types of limiting process. In the first type we let r and n tend to infinity with the ratio r/n held constant: we are then considering the asymptotic distributions of *quantiles* of the sample. In the second type, r remains fixed as n tends to infinity: this leads to results of a radically different nature from those of the quantiles, and this branch of the subject is often designated as the theory of *extreme values*, the name deriving from the fact that the cases of particular interest are those in which r is close to 1 or to n.

14.12 We have already (**10.10** and **10.11**) derived the limiting results for quantiles. We saw there that the sample quantile corresponding to $x_{(r)}$ is asymptotically

normal. Its mean is X_r (given by $F(X_r) = r/n$) and its variance is

$$p_r q_r/(nf_{r,n}^2), \tag{14.20}$$

where $p_r = r/n$, $q_r = 1-p_r$ and $f_{r,n}$ is the ordinate of the frequency function at X_r. Similarly for two quantiles based on $x_{(r)}, x_{(s)}$ we have bivariate normality, variances given by (14.20) and covariance

$$p_r q_s/(nf_{r,n} f_{s,n}). \tag{14.21}$$

The result generalizes readily to more than two quantiles, the variances and covariances being of type (14.20) and (14.21), and the limiting distribution multivariate normal.

Asymptotic distributions of extreme values

14.13 We now turn to the asymptotic theory of extreme values, which we have not so far encountered.(*) In (14.1), we make the transformation $y = nF(x)$, with $dF(x) = dy/n$, obtaining as the distribution of y

$$dH_r(y) = \left(\frac{y}{n}\right)^{r-1}\left(1-\frac{y}{n}\right)^{n-r} d\left(\frac{y}{n}\right)\Big/ B(r, n-r+1), \tag{14.22}$$

so that y/n is a Beta variate of the first kind. As n tends to infinity, for any fixed r,

$$\lim_{n\to\infty} dH_r(y) = y^{r-1}e^{-y}dy/\Gamma(r), \tag{14.23}$$

a Gamma variate with parameter r. When $r = 1$, (14.23) becomes, for the transformed smallest value in the sample, an exponential distribution.

If we transform (14.1) by $y = n(1-F(x))$, we get a parallel result to (14.23), with $n-r$ written for $r-1$, and this also reduces to an exponential distribution for the case $r = n$.

These results apply to the transformed variable y. We now inquire directly into the distribution of the original variable $x_{(r)}$. If $F(x)$ is known explicitly, the transformation leading to (14.23) can, of course, be reversed to give the limiting distribution of $x_{(r)}$ itself.

14.14 In the case of the extreme values ($r = 1$ or n), direct progress can be made by the use of an argument due to Fisher and Tippett (1928). Consider the largest value, and suppose that $n = km$. The largest value $x_{(n)}$ may then be regarded as the largest of k largest values of samples of size m. If $x_{(n)}$ has a limiting distribution, it will be identical with that of the largest values in samples of size m, as m tends to infinity with k fixed, except possibly for location and scale factors due to the factor k. Thus, if $G(x)$ is the limiting distribution function, it satisfies the functional equation

$$G^k(x) = G(a_k x + b_k). \tag{14.24}$$

This therefore characterizes the possible forms of limiting distribution. To solve (14.24), we use a method due to Jenkinson (1955). Put

$$q(x) = -\log\{-\log G(x)\}, \tag{14.25}$$

(*) A very full account of the subject has now been given by E. J. Gumbel in his book *Statistics of Extremes*, Columbia U.P., 1958.

so that on taking logarithms twice, (14.24) becomes

$$\log k = q(x) - q(a_k x + b_k). \tag{14.26}$$

We now expand $q(x)$ and $q(a_k x + b_k)$ in a Taylor series about the point x_{on} for which

$$q(x_{on}) = 0. \tag{14.27}$$

(14.26) becomes

$$\log k = -q(a_k x_{on} + b_k) + \sum_{r=1}^{\infty} \{q^{(r)}(x_{on}) - q^{(r)}(a_k x_{on} + b_k)\} \frac{(x - x_{on})^r}{r!}. \tag{14.28}$$

This is an identity in x. Equating coefficients, we have

$$\log k = -q(a_k x_{on} + b_k) \tag{14.29}$$

and

$$q^{(r)}(x_{on}) = q^{(r)}(a_k x_{on} + b_k), \qquad r \geqslant 1. \tag{14.30}$$

(14.30) gives

$$\frac{q^{(r)}(a_k x_{on} + b_k)}{\{q'(a_k x_{on} + b_k)\}^r} = \frac{q^{(r)}(x_{on})}{\{q'(x_{on})\}^r} = c_r, \tag{14.31}$$

where c_r is independent of k. Since a_k and b_k are constants depending on the value of k, which is arbitrary, we may rewrite (14.31) generally as

$$q^{(r)}(x) = c_r \{q'(x)\}^r. \tag{14.32}$$

Differentiating (14.32) with respect to x, we obtain

$$\begin{aligned} q^{(r+1)}(x) &= r c_r \{q'(x)\}^{r-1} . q''(x) \\ &= r c_r c_2 \{q'(x)\}^{r+1}, \end{aligned} \tag{14.33}$$

on using (14.32) with $r = 2$. Identifying (14.32) and (14.33), we have

$$c_{r+1} = r c_r c_2 = r! c_2^r, \qquad r \geqslant 1. \tag{14.34}$$

Putting (14.34) back into (14.32) gives

$$q^{(r)}(x) = (r-1)! c_2^{r-1} \{q'(x)\}^r, \qquad r \geqslant 1. \tag{14.35}$$

If (14.35) is used in the Taylor expansion of $q(x)$, we obtain, using (14.27),

$$\begin{aligned} q(x) &= \sum_{r=1} c_2^{r-1} \{q'(x)\}^r (x - x_{on})^r / r \\ &= -\frac{1}{c_2} \log \{1 - c_2 q'(x_{on})(x - x_{on})\}. \end{aligned} \tag{14.36}$$

(14.25) and (14.36) give for the characteristic limiting form

$$G(x) = \exp[-\{1 - c_2 q'(x_{on})(x - x_{on})\}^{1/c_2}]. \tag{14.37}$$

14.15 From **14.13** it follows that the distribution function $G(x)$ may alternatively be written

$$G(x) = \exp[-n\{1 - F(x)\}], \tag{14.38}$$

so that, from (14.25),

$$q(x) = -\log[n\{1 - F(x)\}]. \tag{14.39}$$

At x_{on}, in virtue of (14.27) and (14.39), $q(x_{on}) = -\log[n\{1 - F(x_{on})\}] = 0$, so that

$$n\{1 - F(x_{on})\} = 1. \tag{14.40}$$

The left side of (14.40) is the expected number of values exceeding x_{on} in a sample of size n, so that x_{on} is that value of the variate which has expectation of being exceeded just once in such a sample.(*)

From (14.39) and (14.40)

$$q'(x) = F'(x)/\{1-F(x)\}, \tag{14.41}$$

$$q'(x_{on}) = nF'(x_{on}), \tag{14.42}$$

while, from (14.35),

$$\begin{aligned} c_2 &= q''(x)/\{q'(x)\}^2 \\ &= \frac{d}{dx}\{-1/q'(x)\} \\ &= -\frac{d}{dx}\left\{\frac{1-F(x)}{F'(x)}\right\}, \end{aligned} \tag{14.43}$$

by (14.41). If

$$\lim_{x\to\infty} c_2 = 0, \tag{14.44}$$

the limiting form (14.37) becomes, on inserting (14.42) and (14.44),

$$G(x) = \exp[-\exp\{-nF'(x_{on})(x-x_{on})\}], \tag{14.45}$$

a limiting form put forward originally by Fisher and Tippett (1928), the explicit condition (14.44) being due to Von Mises (1936). For the exponential distribution

$$dF(x) = e^{-x}dx, \qquad 0 \leqslant x \leqslant \infty,$$

the expression in curly brackets in (14.43) is zero identically in x, and by analogy distributions for which (14.44), and consequently (14.45), holds have been termed exponential-type.

14.16 If $F(x)$ is not exponential-type, (14.45) is replaced by a different limiting form. From (14.43), we see that

$$c_2 = 1+F''(x)\{1-F(x)\}/\{F'(x)\}^2. \tag{14.46}$$

If now

$$\lim_{x\to\infty} xF'(x)/\{1-F(x)\} = k > 0, \tag{14.47}$$

(14.46) and (14.47) give

$$\lim_{x\to\infty} c_2 = 1+xF''(x)/\{kF'(x)\}. \tag{14.48}$$

On using l'Hôpital's rule, (14.47) becomes

$$1+xF''(x)/F'(x) = -k, \tag{14.49}$$

and (14.48) and (14.49) imply that

$$\lim_{x\to\infty} c_2 = -1/k. \tag{14.50}$$

(*) Gumbel (e.g., 1954) calls x_{on} " the expected largest value " though, as he points out, this is confusing since it is not the expected value of the largest order-statistic.

The use of (14.42), (14.47) and (14.50) in (14.37) gives us the limiting form

$$G(x) = \exp[-\{1+n(1-F(x_{on}))(x-x_{on})/x_{on}\}^{-k}],$$

which finally becomes, on using (14.40),

$$G(x) = \exp\left[-\left(\frac{x}{x_0}\right)^{-k}\right], \tag{14.51}$$

the second type of limiting form, due to Fréchet (1927) and to Fisher and Tippett (1928). The condition (14.47) is implied by

$$\lim_{x\to\infty} x^k\{1-F(x)\} = A>0, \tag{14.52}$$

as may be verified by applying l'Hôpital's rule to (14.52). Distributions satisfying (14.52) (and therefore (14.47)) have no moments of order k or more, and hence are termed Cauchy-type.

14.17 Finally, we consider an original distribution $F(x)$ with a finite terminal x_u where

$$F(x_u) = 1.$$

If, in addition

$$\left.\begin{aligned} F^{(r)}(x_u) &= 0 \qquad (r = 1,2,\ldots,k-1),\\ F^{(k)}(x_u) &\neq 0, \end{aligned}\right\} \tag{14.53}$$

we may expand the exponent of (14.38) in a Taylor series about x_u whose first non-vanishing term is that in $F^{(k)}(x_u)$. If $F^{(k+1)}(x)$ is bounded near x_u, the remainder term is negligible since $x-x_u$ is small, and we obtain

$$G(x) = \exp[-\{(-nF^{(k)}(x_u)/k!)^{1/k}(x-x_u)\}^k], \tag{14.54}$$

the third limiting form for extreme-value distributions given by Fisher and Tippett (1928).

14.18 In **14.14–14.17** we have been discussing the possible limiting distributions of $x_{(n)}$. If we reverse the sign of the variable, $-x_{(1)}$ becomes $x_{(n)}$ and $G(x)$ becomes $1-G(x)$. Thus the distribution of the smallest value $x_{(1)}$ is simply

$$H(x) = 1-G(-x) \tag{14.55}$$

and the results for $x_{(n)}$ apply with only this modification.

Example 14.3

For the normal distribution

$$dF(x) = e^{-\frac{1}{2}x^2}\,dx/\sqrt{(2\pi)},$$

we have, on using (5.68) in (14.43), for large x,

$$c_2 = -\frac{d}{dx}\{(1-F(x))/F'(x)\} = -\frac{d}{dx}\left(\frac{1}{x}\right) = \frac{1}{x^2}$$

so that

$$\lim_{x\to\infty} c_2 = 0.$$

(14.44) is satisfied in this case, giving (14.45) as the limiting form of the distribution. To evaluate its exponent, we use the fact that at x_{on}, (14.40) holds, so that

$$n\{1-F(x_{on})\} = nF'(x_{on})/x_{on} = 1.$$

Thus

$$nF'(x_{on}) = x_{on},$$

and the exponent of (14.45) is simply $-x_{on}(x-x_{on})$. It also follows immediately that

$$\log n + \log F'(x_{on}) - \log x_{on} = 0,$$

i.e.

$$\log n - \tfrac{1}{2}\log(2\pi) - \tfrac{1}{2}x_{on}^2 - \log x_{on} = 0,$$

whence, taking leading terms only,

$$x_{on} \sim (2\log n)^{\frac{1}{2}}.$$

Thus, (14.45) is, for the normal distribution,

$$\exp[-\exp\{-(2\log n)^{\frac{1}{2}}(x-(2\log n)^{\frac{1}{2}})\}].$$

Asymptotic distribution of *m*th values

14.19 Reverting to (14.1), we now turn from the extreme values to a more general investigation of limiting distributions for other order-statistics. We confine ourselves to the case of exponential-type distributions, which are the most important, and consider the mth largest order-statistic, $x_{(n-m+1)}$.

If we put $r = n-m+1$ in (14.1) and differentiate its logarithm, as in **10.10**, we obtain

$$(n-m)F'/F - (m-1)F'/(1-F) + F''/F' = 0 \tag{14.56}$$

at the mode of $x_{(n-m+1)}$. From (14.41) and (14.43) it follows that if (14.44) holds

$$q''(x) = F''/(1-F) + (F')^2/(1-F)^2 \to 0$$

so that since $(1-F) \to 0$, F'' must also do so. Applying l'Hôpital's rule to the last term on the left of (14.56), we obtain

$$\lim F''/F' = \lim F'/\{-(1-F)\}, \tag{14.57}$$

so that (14.56) becomes

$$(n-m)F'/F - mF'(1-F) = 0, \tag{14.58}$$

whence, at the mode $\tilde{x}$ of the distribution of $x_{(n-m+1)}$,

$$F(\tilde{x}) = 1-m/n. \tag{14.59}$$

Comparison of (14.59) with (14.40) shows that, in the exponential case, when $m = 1$ we have

$$x_{on} = \tilde{x},$$

i.e. the mode of the distribution of $x_{(n)}$ coincides with the value which is on the average exceeded by one member.

We now expand $F(x)$ in a Taylor series about $\tilde{x}$,

$$F(x) = F(\tilde{x}) + (x-\tilde{x})F'(\tilde{x}) + (x-\tilde{x})^2F''(\tilde{x})/2 + \ldots \tag{14.60}$$

Since, by (14.57) and (14.59)

$$F'' \sim -(F')^2/(1-F) = -n(F')^2/m, \tag{14.61}$$

we rewrite (14.60), using (14.59) and (14.61), as

$$F(x) = (1-m/n)+(x-\tilde{x})F'(\tilde{x})-n\{F'(\tilde{x})\}^2(x-\tilde{x})^2/(2m)+\ldots \tag{14.62}$$

$$= 1-(m/n)\left[1-\frac{(x-\tilde{x})F'(\tilde{x})}{m/n}+\left\{\frac{(x-\tilde{x})F'(\tilde{x})}{m/n}\right\}^2\cdot\frac{1}{2!}+\ldots\right]$$

$$\sim 1-(m/n)\exp\{-(x-\tilde{x})F'(\tilde{x})/(m/n)\} \tag{14.63}$$

approximately. Now (14.1) may be rewritten

$$dG_{n-m+1}(x) \propto \left(\frac{1-F}{F}\right)^{m-1} d(F^n). \tag{14.64}$$

If we write $-y_m$ for the exponent in (14.63) and substitute it into (14.64), we obtain

$$\left(\frac{1}{F}-1\right)^{m-1} = \left(\frac{1}{1-\frac{m}{n}e^{-y_m}}-1\right)^{m-1}$$

$$\sim \left(\frac{m}{n}\right)^{m-1}\exp\{-(m-1)y_m\},$$

and

$$d(F^n) = n\left(1-\frac{m}{n}e^{-y_m}\right)^{n-1}\cdot\frac{m}{n}e^{-y_m}dy_m.$$

Thus, from (14.64), we have for large n,

$$dG_{n-m+1}(x) \propto \exp[-my_m - m\,e^{-y_m}]\,dy_m$$

and, on evaluating the constants by integration, we get

$$dG_{n-m+1}(x) = \frac{m^m}{(m-1)!}\exp[-my_m - m\,e^{-y_m}]\,dy_m. \tag{14.65}$$

In particular, when $m = 1$, we have

$$dG_n(x) = \exp[-y-e^{-y}]\,dy, \tag{14.66}$$

which is the frequency function corresponding to the distribution function (14.45).

14.20 The unusual limiting form (14.65), which is due to Gumbel (1934), is very far from normal for moderate or low values of m. For its moments, we have (omitting the suffix of y for convenience)

$$\mu_1' = \frac{m^m}{(m-1)!}\int_{-\infty}^{\infty}\exp[-my-m\,e^{-y}]\,y\,dy.$$

Put $e^{-y} = t/m$. We get

$$\mu_1' = -\frac{1}{(m-1)!}\int_0^{\infty}(\log t-\log m)\,t^{m-1}e^{-t}\,dt$$

$$= \log m-\frac{d}{dm}\{\log\Gamma(m)\}$$

$$= \log m+\gamma-\sum_{r=1}^{m-1}\frac{1}{r} \tag{14.67}$$

where γ is Euler's constant. For the rth moment about the mean we have

$$\mu_r = \frac{(-1)^r}{(m-1)!}\int_0^\infty (\log t - \log m + \mu_1')^r\, t^{m-1}\, e^{-t}\, dt$$

$$= \frac{(-1)^r}{\Gamma(m)}\left[\frac{d^r}{dq^r}\exp\{q(\mu_1' - \log m)\}\,\Gamma(m+q)\right]_{q=0}. \qquad (14.68)$$

These formulae have been worked out further by Gumbel, from whose numerical results the following are chosen :—

m	Mean	$\sqrt{\beta_1}$	$\beta_2 - 3$
1	0·577	1·139	2·400
3	0·176	0·621	0·763
5	0·103	0·468	0·437
10	0·051	0·324	0·212

These figures, which relate to the distribution from the top, show clearly that the limiting distribution is far from normal. The distribution from the bottom is similar, odd moments including the mean having the same magnitude but opposite sign, even moments being the same.

J. J. Dronkers (1958, *Biom.*, **45**, 447) has studied the asymptotic distribution of mth values in detail, deriving limiting forms for all three types of initial distribution. Transitional approximations are also given. In the exponential case, it is found that (14.65) is very slowly approached for the normal distribution, although the speed of approach may be improved by adjusting the parameters of (14.65) as in **14.29** below for a related problem.

Exact distributions of mth values in the normal case

14.21 Many authors have considered the exact distributions of the order-statistics in samples from a normal population. The most comprehensive theoretical results have been obtained by Ruben (1954) who gives general recurrence relations for the moments of $x_{(r)}$ for sample size n in terms of those for lower sample sizes. Ruben tabulates the first ten moments for the largest order-statistic for $n = 1(1)50$, and from these calculates the first four central moments, and the measures of skewness and kurtosis to a minimum of 7 significant figures. For the largest order-statistic, Tippett (1925) had previously evaluated the distribution for certain values of n up to 1000. The following values are quoted from his results :—

n	Mean	Standard Deviation	β_1	β_2
2	0·564	0·826	0·019	3·062
5	1·163	0·669	0·092	3·202
10	1·539	0·587	0·168	3·331
100	2·508	0·429	0·429	3·765
500	3·037	0·370	0·570	4·003
1000	3·241	0·351	0·618	4·088

The values of β_1 and β_2 illustrate the point that, as n increases, the distribution of the extreme value diverges more and more from the normal form.

14.22 Hartley and David (1954) have established an upper bound for the mean

value of the largest order-statistic, as follows. By the Cauchy-Schwarz inequality

$$\int_0^1 n F^{n-1}\left(x+\frac{1}{a}\right)dF \leqslant \left[\left\{\int_0^1 (nF^{n-1})^2\,dF\right\}\left\{\int_0^1\left(x+\frac{1}{a}\right)^2 dF\right\}\right]^{\frac{1}{2}},$$

where F is any distribution function whatever, and the constant $a = (n-1)/(2n-1)^{\frac{1}{2}}$. If these integrals are all finite, this inequality becomes (on putting the mean of the distribution equal to zero and its variance to 1 without loss of generality)

$$\int_0^1 n F^{n-1} x\,dF + \frac{1}{a} \leqslant \left[\frac{n^2}{2n-1}\cdot\left(1+\frac{1}{a^2}\right)\right]^{\frac{1}{2}}.$$

The integral on the left is simply the expected value of the largest order-statistic. Thus, on substituting for a, we have

$$E(x_{(n)}) \leqslant (n-1)/(2n-1)^{\frac{1}{2}} = a. \tag{14.69}$$

If the population is symmetrical, we may use the fact that $F(x) = 1-F(-x)$ in the original inequality to obtain the sharper bound

$$E(x_{(n)}) \leqslant n\left[\left\{1-1\Big/\binom{2n-2}{n-1}\right\}\Big/\{2(2n-1)\}\right]^{\frac{1}{2}}. \tag{14.70}$$

This result is due to Moriguti (1951), who gives similar but more complicated results for upper and lower bounds of the variance and the coefficient of variation of $x_{(n)}$ in the symmetrical case. The obvious lower bound of zero for $E(x_{(n)})$ cannot in general be improved. It may be verified that it attains its upper bound for the distribution function

$$F(x) = \left(\frac{ax+1}{n}\right)^{1/(n-1)}, \qquad -1/a \leqslant x \leqslant (n-1)/a. \tag{14.71}$$

For the smallest order-statistic, analogously, the same results hold with $E(x_{(n)})$ replaced by $|E(x_{(1)})|$. Hartley and David (1954) give the general result for any order-statistic

$$|E(x_{(m)})| \leqslant \left[\frac{B(2m-1,2n-2m+1)}{\{B(m,n-m+1)\}^2}-1\right]^{\frac{1}{2}}, \tag{14.72}$$

but this can only be attained by a distribution function if $m = 1$ or n, the special cases considered above.

Joint distribution of two order-statistics

14.23 If in (14.2) we make the transformation

$$\begin{cases} y = nF(x_{(r)}) \\ z = n\{1-F(x_{(s)})\}, \end{cases}$$

we obtain

$$dH_{r,s}(y,z) = \frac{\left(\frac{y}{n}\right)^{r-1}\left(1-\frac{y}{n}-\frac{z}{n}\right)^{s-r-1}\left(\frac{z}{n}\right)^{n-s} dy\,dz}{B(r,s-r)\,B(s,n-s+1).n^2}. \tag{14.73}$$

We keep r and $n-s$ fixed and small and let n tend to infinity. Putting temporarily $t = n-s+1$ we find for the limit

$$dH = \frac{y^{r-1}e^{-y}\,dy}{\Gamma(r)}\cdot\frac{z^{t-1}e^{-z}\,dz}{\Gamma(t)}. \tag{14.74}$$

Thus the rth member from one end and the tth from the other are asymptotically independent, which is what we might expect; and each is distributed in the Gamma form. In the particular case when $t = r$ we may also write this as

$$dH = \frac{y^{r-1}e^{-y}dy}{\Gamma(r)} \cdot \frac{z^{n-r+1}e^{-z}dz}{\Gamma(n-r+1)}. \tag{14.75}$$

Sarhan and Greenberg (1956) have tabulated the covariances of all normal order-statistics for $n = 2(1)20$ to 10 d.p.

Ranges and mid-ranges

14.24 The range of a sample has been briefly mentioned in **2.17.** We now generalize the idea by defining the mth range as the difference R_m given by

$$R_m = x_{(n-m+1)} - x_{(m)}, \tag{14.76}$$

namely the distance between the mth values from either end of the sample. The range itself is the special case $m = 1$. We also define the mth mid-range analogously by

$$M_m = \tfrac{1}{2}(x_{(n-m+1)} + x_{(m)}). \tag{14.77},$$

The particular value M_1 is sometimes called the centre of the sample, but more usually simply the mid-range.

From (14.75) it is clear that the asymptotic distributions of ranges and mid-ranges are distributions of the differences and sums of transformed independent variates. We may, however, derive the exact distributions directly from (14.2) by putting $r = m = n-s+1$ and then transforming from (14.76) and (14.77). We obtain for the joint distribution of R_m and M_m,

$$dH_m \propto \{F(M-\tfrac{1}{2}R)\}^{m-1}\{F(M+\tfrac{1}{2}R)-F(M-\tfrac{1}{2}R)\}^{n-2m} \times \{1-F(M+\tfrac{1}{2}R)\}^{m-1} f(M-\tfrac{1}{2}R) f(M+\tfrac{1}{2}R)\, dR\, dM, \tag{14.78}$$

where the suffix m has been omitted on the right for convenience.

The marginal distributions of R_m and of M_m are obtained by integrating out the other variate from (14.78). A little care is necessary in determining the limits of integration, since these depend in each case on the range of the original distribution F. Equivalently, we may proceed direct from (14.2).

The exact distribution of the range

14.25 Putting $m = 1$ in (14.78), and integrating out M, we obtain for the frequency function of the range

$$dG(R) = dR.n(n-1)\int_{-\infty}^{\infty} \{F(x+R)-F(x)\}^{n-2} f(x+R)\, dF(x). \tag{14.79}$$

For some particular cases, the integral can be evaluated explicitly. In others, quadrature is necessary. From (14.79), we have for the mean value of the range,

$$\begin{aligned} E(R) &= n(n-1)\int_{-\infty}^{\infty} dF_n \int_{-\infty}^{x_{(n)}} (F_n - F_1)^{n-2}(x_{(n)} - x_{(1)})\, dF_1 \\ &= -\int_{-\infty}^{\infty} dx_{(n)} \int_{-\infty}^{x_{(n)}} \frac{d^2}{dx_{(1)}\, dx_{(n)}} (F_n - F_1)^n . (x_{(n)} - x_{(1)})\, dx_{(1)}. \end{aligned} \tag{14.80}$$

On integration by parts under the second integral sign, this becomes

$$E(R) = -\int_{-\infty}^{\infty} dx_{(n)}\left\{\left[-(x_{(n)}-x_{(1)})\frac{d}{dx_{(n)}}F_n^n\right]_{x_{(1)}=-\infty} + \int_{-\infty}^{x_{(n)}} \frac{d}{dx_{(n)}}(F_n-F_1)^n\, dx_{(1)}\right\}. \quad (14.81)$$

Owing to the fact that $(F_n-F_1)^n = 0$ at the upper limit of integration, the order of integration and differentiation may be interchanged in the last term in (14.81), which then becomes

$$\begin{aligned} E(R) &= \int_{-\infty}^{\infty}\left[(x_{(n)}-x_{(1)})\frac{d}{dx_{(n)}}F_n^n\right]_{x_{(1)}=-\infty} dx_{(n)} - \int_{-\infty}^{\infty}(1-F_1)^n\, dx_{(1)} \\ &= \int_{-\infty}^{\infty}\left\{\frac{d}{dx_{(n)}}[(x_{(n)}-x_{(1)})F_n^n]-F_n^n\right\}_{x_{(1)}=-\infty} dx_{(n)} - \int_{-\infty}^{\infty}(1-F_1)^n\, dx_{(1)} \\ &= \left[(x_{(n)}-x_{(1)})F_n^n\right]_{\substack{x_{(n)}=\infty \\ x_{(1)}=-\infty \\ x_{(n)}=-\infty \\ x_{(1)}=-\infty}} - \int_{-\infty}^{\infty} F_n^n\, dx_{(n)} - \int_{-\infty}^{\infty}(1-F_1)^n\, dx_{(1)} \\ &= \int_{-\infty}^{\infty}\{1-F^n-(1-F)^n\}\, dx. \end{aligned} \quad (14.82)$$

Tippett, to whom (14.82) is due, also showed that

$$\operatorname{var} R = 2\int_{-\infty}^{\infty}\int_{-\infty}^{x_{(n)}}\{1-F_n^n-(1-F_1)^n-(F_n-F_1)^n\}\, dx_{(1)}\, dx_{(n)} - \{E(R)\}^2.$$

The normal case has been fairly completely studied by Tippett (1925) and E. S. Pearson ((1926) and (1932)). Tippett found the first four moments of the distribution of the range, tabulated the mean values for values of n up to 1000 and gave a diagram for determining standard errors. His table of mean values to 5 d.p. is reproduced in the *Biometrika Tables for Statisticians*, Vol 1. The following values illustrate the general behaviour of the distribution :—

n	Standard deviation	β_1 (approximate)	β_2 (approximate)
2	0·853	0·99	3·87
10	0·797	0·16	3·20
100	0·605	0·21	3·39
500	0·524	0·29	3·50
1000	0·497	0·31	3·54

It appears that as n increases, the distribution of the range diverges more and more from the normal form.

Pearson and Hartley (1942) have tabulated to 4 d.p. the distribution function of R/σ in samples of size 2 to 20 from a normal population. Their results, too, are reproduced in the *Biometrika Tables for Statisticians*, Vol. 1. H. L. Harter (1960, *Ann. Math. Statist.*, **31**, 1122) gives to 6 d.p. for $n = 2(1)20(2)40(10)100$ the values of R/σ for which the d.f. is equal to 0·0001, 0·0005, 0·001, 0·005, 0·01, 0·025, 0·05, 0·1(0·1)0·9, 0·95, 0·975, 0·99, 0·995, 0·999, 0·9995, 0·9999. He also gives the first four moments to at least 8 significant figures for $n = 2(1)100$.

Cox (1954) has considered the effect of departures from parent normality on the distribution of the range in small samples, concluding that kurtosis, rather than skewness, is an appropriate guide to the behaviour of the distribution.

14.26 From (14.70) it follows that the range has its mean value bounded for symmetrical populations (with zero mean and unit variance) by

$$0 \leqslant E(R) \leqslant n\left[2\left\{1-1\Big/\binom{2n-2}{n-1}\right\}\Big/(2n-1)\right]^{\frac{1}{2}} \sim (n+\tfrac{1}{2})^{\frac{1}{2}},$$

a result due originally to Plackett (1947), who showed that it also holds for non-symmetrical populations. Hartley and David (1954) confirmed this and also derived upper and lower bounds for $E(R)$ in samples from populations with finite range. The following table (extracted from Plackett (1947)) gives values for the upper bound, its asymptotic value and the exact values for the normal and rectangular populations.

n	Upper bound for $E(R)$	$(n+\frac{1}{2})^{\frac{1}{2}}$	Exact values	
			Normal	Rectangular
2	1·155	1·581	1·128	1·155
3	1·732	1·871	1·693	1·732
4	2·084	2·121	2·059	2·078
8	2·921	2·915	2·847	2·694
12	3·539	3·536	3·258	2·931

The asymptotic distribution of the range

14.27 We now turn to the limiting distribution of R as n becomes large. From (14.2) with $r = n-s+1 = 1$, we have, for the joint distribution of the smallest and largest observations,

$$dG_{1,n}(x_1, x_n) = n(n-1)(F_n - F_1)^{n-2}\, dF_1\, dF_n.$$

We now transform by the relations

$$\left.\begin{aligned} u &= 2n\{F_1(1-F_n)\}^{\frac{1}{2}}, \\ v &= \tfrac{1}{2}\log\{F_1/(1-F_n)\}, \end{aligned}\right\} \tag{14.83}$$

the Jacobian of the transformation being $u/\{2n^2 f_1 f_n\}$. The joint distribution of u and v is then

$$dF(u,v) = \left(\frac{n-1}{2n}\right) u\left(1-\frac{u\cosh v}{n}\right)^{n-2} du\, dv. \tag{14.84}$$

As $n \to \infty$, (14.84) tends to

$$\lim_{n\to\infty} dF(u,v) = \tfrac{1}{2} u \exp(-u\cosh v)\, du\, dv. \tag{14.85}$$

The domain of variation is bounded by

$$u \geqslant 0, \qquad \cosh v \leqslant n/u,$$

in (14.84), so that in (14.85) $\cosh v$, and therefore v, can range to infinity. Hence we obtain from (14.85) the limiting marginal distribution of u:

$$\left.\begin{aligned} dF(u) &= u\, du \int_0^\infty \exp(-u\cosh v)\, dv \\ &= u\, du \int_1^\infty \exp(-ut)(t^2-1)^{-\frac{1}{2}}\, dt. \end{aligned}\right\} \tag{14.86}$$

Now u may be expressed as

$$u = 2n\{F(M-\tfrac{1}{2}R)[1-F(M+\tfrac{1}{2}R)]\}^{\frac{1}{2}}. \tag{14.87}$$

If now we assume that the mid-range M converges in probability to the known population mid-range μ, u will similarly behave like

$$u' = 2n\{F(\mu-\tfrac{1}{2}R)[1-F(\mu+\tfrac{1}{2}R)]\}^{\frac{1}{2}}. \tag{14.88}$$

If we may take $\mu = 0$ and if the population is symmetrical, (14.88) reduces to

$$u' = 2n\{F(-\tfrac{1}{2}R)\}, \tag{14.89}$$

so that u will tend to be simply the probability integral transform of the range, apart from constant multipliers.

14.28 The value of these results, which are due to Elfving (1947), is restricted by the condition, imposed in order to reach (14.88), that the mid-range M converges to the population mid-range μ in probability. Even when the population is symmetrical and exponential-type, this condition is not necessarily fulfilled. To show this, we put $m = 1$ and $r = 2$ in (14.68), and find that the variance of the exponential-type extreme distribution (14.66) is $\pi^2/6$. By (14.45), the variate here is

$$y = nF'(x_{on})(x-x_{on}), \tag{14.90}$$

where x is written for $x_{(n)}$. Thus $\operatorname{var} y = \pi^2/6$ and hence, for large samples,

$$\operatorname{var} x = \pi^2/\{6[nF'(x_{on})]^2\}. \tag{14.91}$$

Since $x_{(n)}$ and $x_{(1)}$ are asymptotically independent, the mid-range, which is their mean, has variance

$$\operatorname{var} M = \pi^2/\{12[nF'(x_{on})]^2\}. \tag{14.92}$$

Example 14.4

In the normal case, we have seen (Example 14.3) that

$$nF'(x_{on}) \sim (2\log n)^{\frac{1}{2}}$$

so that, from (14.92),

$$\operatorname{var} M = \pi^2/(24\log n).$$

Thus, in the normal case, M converges to the population mid-range (which, by symmetry, is the mean).

Example 14.5

The double exponential distribution

$$dF(x) = \tfrac{1}{2}\exp(-|x|)\,dx, \qquad -\infty \leqslant x \leqslant \infty,$$

has finite moments of all orders. Its distribution function is

$$F(x) = 1-\tfrac{1}{2}e^{-x}, \qquad x \geqslant 0. \tag{14.93}$$

It is easily verified that (14.44) is satisfied. By (14.40) and (14.93), $x_{on} = \log(\frac{1}{2}n)$ and therefore

$$nF'(x_{on}) = \tfrac{1}{2}n\exp(-\log\tfrac{1}{2}n) = 1.$$

Thus, for this case, from (14.92)

$$\operatorname{var} M = \pi^2/12$$

and M does not converge in probability to zero, the population mid-range (and mean).

The reduced range and mid-range

14.29 In view of the result of the last section, we return to the problem of **14.27**, and seek a more general solution of the asymptotic distribution problem for the range and mid-range, or some simple functions of them. We confine our attention to symmetrical exponential-type distributions.

If we equate (14.38) and (14.45), we find

$$1-F(x_{(n)}) = \frac{1}{n}\exp\{-nF'(x_{on})(x_{(n)}-x_{on})\},$$

and similarly, using (14.55) also,

$$F(x_{(1)}) = \frac{1}{n}\exp\{nF'(x_{on})(x_{(1)}+x_{on})\}.$$

Thus (14.83) becomes

$$\left.\begin{aligned} u &= 2\exp\{-\tfrac{1}{2}nF'(x_{on})(x_{(n)}-x_{(1)}-2x_{on})\},\\ v &= \tfrac{1}{2}nF'(x_{on})(x_{(n)}+x_{(1)}). \end{aligned}\right\} \tag{14.94}$$

We now define the *reduced range* R^* and the *reduced mid-range* M^* by

$$\left.\begin{aligned} R^* &= nF'(x_{on})(R-2x_{on}),\\ M^* &= nF'(x_{on})M. \end{aligned}\right\} \tag{14.95}$$

(14.94) then becomes

$$u = 2\exp(-\tfrac{1}{2}R^*), \tag{14.96}$$

$$v = M^*. \tag{14.97}$$

If we apply (14.96) as a transformation of the distribution of u at (14.86), we obtain, for the distribution of R^*,

$$\begin{aligned} dF(R^*) &= 2\exp(-R^*)\,dR^*.\int_0^\infty \exp\{-2\exp(-\tfrac{1}{2}R^*)\cosh v\}\,dv\\ &= 2\exp(-R^*)K_0\{2\exp(-\tfrac{1}{2}R^*)\}\,dR^*, \end{aligned} \tag{14.98}$$

where K_0 is a Bessel function of zero order and imaginary argument. Gumbel (1949), who had earlier obtained (14.98) by a different method, has given tables of the frequency function and of the distribution function of R^*, and also tabulated its percentage points. Cox (1948), who also obtained this result, used the method of steepest descents to obtain a second approximation, which he compared with (14.98) and Elfving's result at (14.86). The latter is the most accurate in the normal case, and the steepest descents approximation, as might be expected, is also better than (14.98). But as will be seen in the next paragraph, even (14.98) gives a good approximation for small n.

The range proper, R, is a linear function of R^*, by (14.95). As in **14.28**, we know that, asymptotically,

$$\left.\begin{aligned} E(R^*) &= E(y_{(n)})-E(y_{(1)}) = 2\gamma,\\ \operatorname{var}(R^*) &= \operatorname{var}(y_{(n)})+\operatorname{var}(y_{(1)}) = \pi^2/3. \end{aligned}\right\} \tag{14.99}$$

Using (14.95), we have from (14.99), asymptotically,

$$\left.\begin{aligned} E(R) &= 2\gamma/\{nF'(x_{on})\}+2x_{on},\\ \operatorname{var}(R) &= \pi^2/[3\{nF'(x_{on})\}^2]. \end{aligned}\right\} \tag{14.100}$$

By using this and the exact normal parent values of $E(R)$ and $\text{var}(R)$ (tabulated by Tippett (1925) and E. S. Pearson (1926)) for small n, Gumbel (1947) estimated x_{on} and $F'(x_{on})$ and investigated the closeness of the approximation to the distribution of R obtained from (14.98) by comparing the results it gave with those obtained from the exact tables given by Pearson and Hartley (1942). He found that the distribution function of R was closely approximated by (14.98) for n as low as 6. Thus in the normal case, at least, (14.98) solves the problem of the distribution of the range.

14.30 Turning now to the reduced mid-range M^*, we see from (14.97) that its asymptotic distribution, obtained by integrating u out of (14.85), is

$$dG(M^*) = \tfrac{1}{2}dM^* \int_0^\infty u \exp(-u \cosh M^*)\, du. \tag{14.101}$$

The substitution

$$z = u \cosh M^*$$

reduces this to

$$dG(M^*) = dM^*/(2 \cosh^2 M^*). \tag{14.102}$$

(14.95) immediately gives the distribution of the mid-range proper, M, from (14.102). It is symmetrical, with zero mean, so that M is, for symmetrical exponential-type populations, an unbiassed estimator of the population mid-range (mean), which is zero. But Example 14.5 shows that the even moments of M are heavily dependent on the form of the population.

Tables for extreme-value distributions

14.31 *Probability Tables for the Analysis of Extreme-Value Data* (National Bureau of Standards, Applied Mathematics Series, 22, Washington, 1953) gives:—

(1) The distribution function and frequency function of the exponential-type extreme-value distribution (14.45) and (14.66) to 7 decimal places over the interval $-3{\cdot}0$ to $17{\cdot}0$, including $-2{\cdot}40\,(0{\cdot}05)\,0\,(0{\cdot}1)\,4{\cdot}0$, with auxiliary inverse tables giving the extreme value to 5 d.p. for values of its distribution function $0\,(0{\cdot}0001)\,0{\cdot}005\,(0{\cdot}001)\,0{\cdot}988\,(0{\cdot}0001)\,0{\cdot}9994\,(0{\cdot}000{,}01)\,1$, and giving the ordinate to 5 d.p. for distribution function values $0\,(0{\cdot}0001)\,0{\cdot}010\,(0{\cdot}001)\,0{\cdot}999$.

(2) Percentage points for the distributions of mth values (14.65) for $m = 1\,(1)\,15\,(5)\,50$. The 12 percentage points range between 0·005 and 0·995.

(3) The distribution and frequency functions of the reduced range for exponential-type distributions (14.98), for values of the reduced range $1{\cdot}0\,(0{\cdot}05)\,11{\cdot}0\,(0{\cdot}5)\,20{\cdot}0$, to 7 d.p., with inverse tables giving the reduced range to 3 or 4 d.p. for selected values of its distribution function.

14.32 When we come in Chapters 19 and 32 (Volume 2) to consider problems of estimation and testing hypotheses, we shall have to turn to some additional sampling properties of the order-statistics.

EXERCISES

14.1 Show that the median in samples of (odd) n from the Cauchy distribution

$$dF = \frac{dx}{\pi(1+x^2)}, \quad -\infty \leqslant x \leqslant \infty,$$

has a variance given approximately by

$$\tfrac{1}{4}\pi^2(n+2)^{-1} + \tfrac{1}{8}\pi^4(n+2)^{-2}.$$

(David and Johnson, 1954)

14.2 For the "logistic" d.f.

$$F(x) = (1+e^{-x})^{-1},$$

show directly that the limiting d.f. of $z = x_{(n)} - \log n$ is

$$G(z) = \exp(-e^{-z})$$

and verify that this is equivalent to (14.45) in this case.

14.3 In the notation of **14.9** show that for the median in (odd) samples, to order $(n+2)^{-2}$,

$$\mathrm{E}(m_e) = X_{r+1} + \frac{1}{8(n+2)} X''_{r+1} + \frac{1}{128(n+2)^2} X^{(iv)}_{r+1},$$

$$\text{var } m_e = \frac{1}{4(n+2)}(X'_{r+1})^2 + \frac{1}{32(n+2)^2}\{2X'_{r+1}X'''_{r+1} + (X''_{r+1})^2\},$$

$$\gamma_1(m_e) = \frac{3}{2(n+2)^{\frac{1}{2}}} \cdot \frac{X''_{r+1}}{X'_{r+1}}, \qquad \gamma_2(m_e) = \frac{1}{n+2}\left\{\frac{X'''_{r+1}}{X'_{r+1}} + \frac{3(X''_{r+1})^2}{(X'_{r+1})^2} - 6\right\}.$$

(David and Johnson, 1954)

14.4 Use the c.f.'s of Exercises 4.21 and 14.21 to establish the result (14.102) for the limiting distribution of the reduced mid-range.

14.5 Show that if $G_{r,n}$ refers to the distribution (14.1) for the rth order-statistic in a sample of n,

$$(n+1)\,dG_{r,n} = r\,dG_{r+1,n+1} + (n-r+1)\,dG_{r,n+1}$$

and hence that the c.f. and moments of $x_{(r)}$ obey a similar recurrence relation. Similarly, if $G_{r,s,n}$ refers to (14.2), show that

$$(n+1)\,dG_{r,s,n} = r\,dG_{r+1,s+1,n+1} + (s-r)\,dG_{r,s+1,n+1} + (n-s+1)\,dG_{r,s,n+1}.$$

14.6 Rewriting (14.20) as $V_p = p(1-p)/(nf_p^2)$, $f_p > 0$, show that

$$\frac{\partial \log V_p}{\partial p} = \frac{1-2p}{p(1-p)} - 2\frac{f'_p}{f_p},$$

and

$$\frac{\partial^2 \log V_p}{dp^2} = -\left(\frac{1}{p^2} + \frac{1}{(1-p)^2}\right) + 2\left(\frac{f'_p}{f_p^2}\right)^2 - 2\frac{f''_p}{f_p^2}.$$

Hence show that V_p has a turning-point at the sample median ($p = 0{\cdot}5$) if and only if f_p has a turning-point there; that if $f_{0.5}$ is a minimum, $V_{0.5}$ is a maximum; that if $f_{0.5}$ is a maximum, $V_{0.5}$ is a minimum if $|f''_{0.5}| > 4f^2_{0.5}$, and a maximum otherwise; and that if $f_{0.5}$ is a point of inflexion, $V_{0.5}$ is a maximum. Show also that if f_p has a turning-point at $p_0 \neq 0{\cdot}5$, V_p has no turning-point either at p_0 or at $0{\cdot}5$.

(Sen, 1961, *J. Roy. Statist. Soc.*, **B, 23,** 453)

14.7 For a continuous symmetrical distribution with frequency function $f(x)$ which possesses an absolute maximum at its median ξ, show that the variance of the sample median in samples of odd n satisfies

$$\text{var } m_e \geqslant [4\{f(\xi)\}^2.(n+2)]^{-1},$$

and verify that the equality holds for the rectangular distribution. Hence show that the asymptotic value for the variance effectively provides a lower bound for the small-sample variance.
(Chu, 1955)

14.8 Show that in samples of n from

$$dF = e^{-x}dx, \quad 0 \leqslant x \leqslant \infty,$$

the variance of the smallest sample value is $1/n^2$.

14.9 Verify that the three limiting forms of equations (14.45), (14.51) and (14.54) satisfy the functional relation (14.24).

14.10 A sample of n is drawn from a normal population with unit variance, and that sample value is chosen which is closest to the true mean. Show that the variance of this value is

$$v_n = 1 - \frac{n(n-1)}{\pi} + \tfrac{1}{2}n(n-1)(n-2)\left(\frac{2}{\pi}\right)^{\frac{1}{2}n}\int_0^\infty \exp\left(-\frac{3}{2}x^2\right)\left\{\int_x^\infty \exp(-\tfrac{1}{2}t^2)\,dt\right\}^{n-3} dx.$$

(Kendall, 1954)

14.11 In the example of the previous exercise show that

$$v_5 = 1 - \frac{20}{\pi} + \frac{240}{\pi^2\sqrt{3}} \arctan\left(\sqrt{\frac{5}{3}} - \frac{\pi}{6}\right)$$

exactly and that

$$v_n = \pi/n^2$$

approximately.
(Kendall, 1954)

14.12 Show that the joint distribution of the sample range and mid-range in samples from a distribution rectangular on the interval $(-\frac{1}{2}, \frac{1}{2})$ is

$$dF(R, M) = n(n-1)R^{n-2}\,dR\,dM, \quad 0 \leqslant R \leqslant 1 - 2|M| \leqslant 1.$$

14.13 From the previous exercise show that the marginal distribution of M is

$$dG(M) = n(1-2|M|)^{n-1}\,dM, \quad |M| \leqslant \tfrac{1}{2},$$

with moments

$$\left.\begin{aligned} \mu_{2k-1} &= 0 \\ \mu_{2k} &= \left\{2^{2k}\binom{2k+n}{2k}\right\}^{-1} \end{aligned}\right\} k = 1, 2, \ldots$$

In particular show that $\mu_2 = \{2(n+1)(n+2)\}^{-1}$.

14.14 Show that in a normal sample of size $n \geqslant 2$, the order-statistics measured from the sample mean are distributed independently of that mean. Hence show that

$$\begin{aligned} \kappa_r(x_{(j)} - \bar{x}) &= \kappa_r(x_{(j)}), \quad r > 2 \\ &= \kappa_2(x_{(j)}) - n^{-1}, \quad r = 2. \end{aligned}$$

(McKay, 1935)

14.15 Show that the expectation of the range in samples of n is given by

$$\mathrm{E}(R) = \int_{-\infty}^{\infty} x\,d\{F^n + (1-F)^n - 1\}$$

and hence that it exists if and only if the population mean is finite, when its value is

$$\int_{-\infty}^{\infty} \{1-F^n-(1-F)^n\}\,dx.$$

(Cox, 1954)

14.16 Continuing Exercise 14.12, show that the marginal distribution of R is

$$dH(R) = n(n-1)R^{n-2}(1-R)\,dR, \quad 0 \leqslant R \leqslant 1,$$

with moments

$$\mu_k' = n(n-1)/\{(n+k)(n+k-1)\}.$$

In particular

$$\mu_1' = (n-1)/(n+1)$$
$$\mu_2 = 2(n-1)/\{(n+1)^2(n+2)\}.$$

14.17 Continuing Exercise 14.12, show that the distribution of the ratio M/R is given by

$$dF\left(\frac{M}{R}\right) = (n-1)\left\{1+2\left|\frac{M}{R}\right|\right\}^{-n} d\left(\frac{M}{R}\right),$$

with moments

$$\left.\begin{aligned} \mu_{2k-1} &= 0, \\ \mu_{2k} &= 1\Big/\left\{2^{2k}\binom{n-2}{2k}\right\} \end{aligned}\right\} \quad k = 1,2,\ldots$$

(Carlton, 1946)

14.18 In Exercise 14.13, show that if $t = nM$

$$\lim_{n\to\infty} dF(t) = \exp(-2\,|\,t\,|)\,dt,$$

and hence that the difference between two such reduced mid-ranges, $t_1 - t_2 = z$, has the distribution

$$dG(z) = (\tfrac{1}{2}+|\,z\,|)\exp(-2\,|\,z\,|)\,dz,$$

with moments

$$\left.\begin{aligned} \mu_{2k-1} &= 0, \\ \mu_{2k} &= (2k)!\,(k+1)/2^{2k}. \end{aligned}\right\} \quad k = 1,2,\ldots$$

(Carlton, 1946)

14.19 Show that the distribution function of the range is given exactly by

$$G(R) = n\int_{-\infty}^{\infty} \{F(x+R)-F(x)\}^{n-1}\,dF(x).$$

(Gumbel, 1947)

14.20 Show that u in (14.86) has mean $\frac{1}{2}\pi$ and variance $(4-\frac{1}{4}\pi^2)$.

(Elfving, 1947)

14.21 Show that the distribution (14.66) has characteristic function $\Gamma(1-it)$ and hence that

$\kappa_1 = \gamma$ (Euler's constant, $= 0{\cdot}577$); $\kappa_2 = \Sigma n^{-2} = \pi^2/6 = 1{\cdot}645$;
$\kappa_3 = 2\Sigma n^{-3} = 2{\cdot}404$; $\kappa_4 = 6\Sigma n^{-4} = \pi^4/15 = 6{\cdot}494$.

Hence derive $\beta_1 = 1{\cdot}299$, $\beta_2 = 5{\cdot}4$.

CHAPTER 15

THE MULTIVARIATE NORMAL DISTRIBUTION AND QUADRATIC FORMS

15.1 The univariate normal distribution

$$dF = (2\pi)^{-\frac{1}{2}} \exp\left\{-\tfrac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}\frac{dx}{\sigma}, \qquad -\infty \leqslant x \leqslant \infty,$$

has been discussed in many contexts in earlier chapters. Apart from location and scale parameters, μ and σ, the exponent of the frequency function is simply $-\frac{1}{2}x^2$. We now seek to generalize the normal distribution to the joint distribution of two or more variables, and it is natural to look for this generalization in a frequency function which has as its exponent a quadratic form in the variables, p in number, i.e.

$$dF \propto \exp\left\{-\tfrac{1}{2}\sum_{j=1}^{p}\sum_{k=1}^{p} a_{jk}\left(\frac{x_j-\mu_j}{\sigma_j}\right)\left(\frac{x_k-\mu_k}{\sigma_k}\right)\right\}\prod_{j=1}^{p}\frac{dx_j}{\sigma_j}, \qquad -\infty \leqslant x_j \leqslant \infty. \tag{15.1}$$

Here, as in the univariate case, the μ_j and σ_j are simply location and scale parameters which may be removed by a standardizing transformation.

In this chapter, and generally when dealing with multivariate problems, we shall find it convenient to use vector and matrix notation for economy and clarity in representation. A bold-face capital letter will represent a matrix, a bold-face small letter a column vector, and a prime will denote transposition, so that a row vector will be represented by a bold-face primed small letter. With this notation, (15.1) becomes

$$dF \propto \exp\{-\tfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\mathbf{A}(\mathbf{x}-\boldsymbol{\mu})\}\prod_j dx_j, \tag{15.2}$$

where $\mathbf{A}$ is a symmetric $(p \times p)$ matrix.

Characteristic function and moments

15.2 The condition that (15.2) be a properly defined frequency function is simply that its exponent is non-negative definite. For in such a case (and only in such a case) we can find a real linear transformation which transforms the exponent into a negative sum of squares, and the integral of dF will converge. In particular there is an *orthogonal* transformation

$$\mathbf{x} = \mathbf{B}\mathbf{y} \tag{15.3}$$

which transforms the exponent of (15.2) into a sum of squares, the coefficients in this sum being the latent roots of $\mathbf{A}$. Confining ourselves to the non-singular case, where the rank of $\mathbf{A}$ is p,(*) and measuring from the vector $\boldsymbol{\mu}$, we now apply this transformation to obtain the characteristic function of the distribution, and incidentally to evaluate the proportionality constant which ensures that its integral has the value unity.

(*) If $\mathbf{A}$ has rank $r < p$, the distribution degenerates into r dimensions, i.e. one or more variates are redundant.

In our notation, the characteristic function is (cf. (4.27))

$$\phi(\mathbf{t}) \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp(i\mathbf{t}'\mathbf{x} - \tfrac{1}{2}\mathbf{x}'\mathbf{A}\mathbf{x}) \prod_j dx \tag{15.4}$$

which, on applying (15.3), becomes

$$\phi(\mathbf{t}) \propto \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp(i\mathbf{t}'\mathbf{B}\mathbf{y} - \tfrac{1}{2}\mathbf{y}'\mathbf{C}\mathbf{y}) \prod_j dy_j, \tag{15.5}$$

the Jacobian being unity since $\mathbf{B}$ is orthogonal. We have written

$$\mathbf{C} = \mathbf{B}'\mathbf{A}\mathbf{B} \tag{15.6}$$

in (15.5). Since $\mathbf{C}$ is a diagonal matrix, the right side of (15.5) factorizes into p single integrals of type

$$I_j = \int_{-\infty}^{\infty} \exp(iu_j b_j y_j - \tfrac{1}{2}c_{jj}y_j^2)\, dy_j,$$

where u_j is linear in the dummy variables t_j. Then I_j is obtained directly from the characteristic function of the univariate normal distribution (Example 3.11) as

$$I_j = \left(\frac{2\pi}{c_{jj}}\right)^{\frac{1}{2}} \exp\left(-\tfrac{1}{2}\frac{u_j^2 b_j^2}{c_{jj}}\right). \tag{15.7}$$

Applying (15.7) to each term in (15.5) gives

$$\phi(\mathbf{t}) \propto (2\pi)^{\frac{1}{2}p} \left(\prod_j c_{jj}\right)^{-\frac{1}{2}} \exp\left(-\tfrac{1}{2}\sum_j \frac{u_j^2 b_j^2}{c_{jj}}\right). \tag{15.8}$$

Since $\mathbf{C}$ is diagonal, its determinant is

$$|C| = \prod_j c_{jj}$$

and that of its reciprocal matrix $\mathbf{C}^{-1}$ is

$$|C^{-1}| = |C|^{-1} = \left(\prod_j c_{jj}\right)^{-1}. \tag{15.9}$$

Using (15.9) in (15.8) and rewriting its exponent in matrix form, we have

$$\phi(\mathbf{t}) \propto (2\pi)^{\frac{1}{2}p} |C^{-1}|^{\frac{1}{2}} \exp(-\tfrac{1}{2}\mathbf{t}'\mathbf{B}\mathbf{C}^{-1}\mathbf{B}'\mathbf{t}). \tag{15.10}$$

Since, by (15.6),

$$\mathbf{B}\mathbf{C}^{-1}\mathbf{B}' = \mathbf{B}(\mathbf{B}'\mathbf{A}\mathbf{B})^{-1}\mathbf{B}' = \mathbf{B}\mathbf{B}^{-1}\mathbf{A}^{-1}(\mathbf{B}')^{-1}\mathbf{B}' = \mathbf{A}^{-1}, \tag{15.11}$$

we have, since $\mathbf{B}$ is orthogonal,

$$|A^{-1}| = |BC^{-1}B'| = |B||C^{-1}||B'| = |C^{-1}|. \tag{15.12}$$

We may substitute (15.11) and (15.12) into (15.10) to give

$$\phi(\mathbf{t}) \propto (2\pi)^{\frac{1}{2}p} |A^{-1}|^{\frac{1}{2}} \exp(-\tfrac{1}{2}\mathbf{t}'\mathbf{A}^{-1}\mathbf{t}). \tag{15.13}$$

If we put $\mathbf{t} = \mathbf{0}$ in (15.13), we obtain immediately for the proportionality constant of the multivariate normal distribution

$$(2\pi)^{-\frac{1}{2}p} |A^{-1}|^{-\frac{1}{2}} = (2\pi)^{-\frac{1}{2}p} |A|^{\frac{1}{2}}.$$

Thus (15.2) becomes

$$dF = (2\pi)^{-\frac{1}{2}p} |A|^{\frac{1}{2}} \exp\{-\tfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\mathbf{A}(\mathbf{x}-\boldsymbol{\mu})\} \prod dx_j \tag{15.14}$$

with characteristic function

$$\phi(\mathbf{t}) = \exp(-\tfrac{1}{2}\mathbf{t}'\mathbf{A}^{-1}\mathbf{t})\exp(i\mathbf{t}'\boldsymbol{\mu}), \tag{15.15}$$

the second factor on the right of (15.15) arising if the location vector $\boldsymbol{\mu}$ is non-null, as may be seen from the effect on (15.5) of putting $(\mathbf{x}-\boldsymbol{\mu})$ for $\mathbf{x}$ in the second term of the exponent in (15.14) and in (15.3).

15.3 From the characteristic function (15.15), the moments of the distribution, or, more conveniently, its cumulants, are easily obtainable. The cumulant-generating function is

$$\psi(\mathbf{t}) = -\tfrac{1}{2}\mathbf{t}'\mathbf{A}^{-1}\mathbf{t} + i\mathbf{t}'\boldsymbol{\mu}. \tag{15.16}$$

It is clear from the form of (15.16), which has no terms of degree above 2, that all cumulants of order greater than 2 are equal to zero, thus generalizing the univariate property of the normal distribution. The mean of the jth variate is

$$\left[\frac{\partial\psi(\mathbf{t})}{\partial(it_j)}\right]_{\mathbf{t}=\mathbf{0}} = \mu_j,$$

so that the location vector $\boldsymbol{\mu}$ is the vector of means of the variables. The variance of the jth variate is

$$\left[\frac{\partial^2\psi(\mathbf{t})}{\partial(it_j)^2}\right]_{\mathbf{t}=\mathbf{0}} = a_{jj}^{-1}, \tag{15.17}$$

the element in the jth row and jth column of $\mathbf{A}^{-1}$. The covariance of the jth and kth variates is

$$\left[\frac{\partial^2\psi(\mathbf{t})}{\partial(it_j)\,\partial(it_k)}\right]_{\mathbf{t}=\mathbf{0}} = a_{jk}^{-1}. \tag{15.18}$$

(15.17) and (15.18) show that $\mathbf{A}^{-1}$ is the matrix of variances and covariances of the variates, which we shall call their *dispersion matrix*. We may now make our notation more suggestive by writing $\mathbf{V}$ for $\mathbf{A}^{-1}$, and obtain, as our final expression for the p-variate normal distribution in terms of its first- and second-order moments,

$$dF = (2\pi)^{-\frac{1}{2}p}|V^{-1}|^{\frac{1}{2}}\exp\{-\tfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\mathbf{V}^{-1}(\mathbf{x}-\boldsymbol{\mu})\}\Pi\,dx \tag{15.19}$$

with characteristic function

$$\phi(\mathbf{t}) = \exp(-\tfrac{1}{2}\mathbf{t}'\,\mathbf{V}\mathbf{t})\exp(i\,\mathbf{t}'\boldsymbol{\mu}). \tag{15.20}$$

Example 15.1 The bivariate normal distribution

We have already encountered the bivariate specializations of (15.19) and (15.20) in Example 3.17 and elsewhere. In this case the matrix of variances and covariances is

$$\mathbf{V} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

where ρ, the coefficient of correlation, is defined by $\mu_{11}/\sigma_1\sigma_2$. Hence

$$\mathbf{V}^{-1} = \begin{pmatrix} \dfrac{1}{\sigma_1^2(1-\rho^2)} & \dfrac{-\rho}{\sigma_1\sigma_2(1-\rho^2)} \\ \dfrac{-\rho}{\sigma_1\sigma_2(1-\rho^2)} & \dfrac{1}{\sigma_2^2(1-\rho^2)} \end{pmatrix},$$

and the frequency function with zero means is, from (15.19),

$$dF = \frac{dx_1\,dx_2}{2\pi\sigma_1\sigma_2(1-\rho^2)^{\frac{1}{2}}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{x_1^2}{\sigma_1^2}-\frac{2\rho x_1 x_2}{\sigma_1\sigma_2}+\frac{x_2^2}{\sigma_2^2}\right)\right\},$$

with characteristic function, from (15.20),

$$\phi(t_1, t_2) = \exp\{-\tfrac{1}{2}(\sigma_1^2 t_1^2 + 2\rho\sigma_1\sigma_2 t_1 t_2 + \sigma_2^2 t_2^2)\}.$$

Linear functions of normal variates

15.4 We have seen (Example 11.2) that a linear function of independent univariate normal variates is itself normally distributed. We may now obtain a much more general result, to the effect that a set of variates, each of which is a linear function of a set of multinormal variates (i.e. variates which have a multivariate normal distribution), has itself a multinormal distribution. (15.20) is the characteristic function of a p-variate non-singular normal distribution. It may be written

$$\mathrm{E}\{\exp(i\mathbf{t}'\mathbf{x})\} = \exp(-\tfrac{1}{2}\mathbf{t}'\mathbf{V}\mathbf{t}). \tag{15.21}$$

Suppose that we write

$$\mathbf{t}' = \mathbf{s}'\mathbf{A}, \tag{15.22}$$

where $\mathbf{s}'$ is a vector of q dummy variables, just as $\mathbf{t}'$ is a vector of p such variables, and $\mathbf{A}$ is any $(q\times p)$ matrix. (15.22) substituted into (15.21) gives

$$\mathrm{E}\{\exp(i\mathbf{s}'\mathbf{A}\mathbf{x})\} = \exp(-\tfrac{1}{2}\mathbf{s}'\mathbf{A}\mathbf{V}\mathbf{A}'\mathbf{s}). \tag{15.23}$$

(15.23) shows that the set of q linear functions of the p x-variates defined by $\mathbf{y} = \mathbf{A}\mathbf{x}$ has a multinormal characteristic function with dispersion matrix $\mathbf{A}\mathbf{V}\mathbf{A}'$. If $\mathbf{A}$ is a non-singular matrix, the multivariate form of the Inversion Theorem (**4.17**) gives us a multinormal frequency function (15.19) for $\mathbf{y}$, with $\mathbf{A}\mathbf{V}\mathbf{A}'$ replacing $\mathbf{V}$.

If $\mathbf{A}$ is singular, the inversion cannot be explicitly carried out, since $(\mathbf{A}\mathbf{V}\mathbf{A}')^{-1}$ is not defined, but (15.23) is nevertheless the characteristic function of a degenerate multinormal distribution. In particular, this is so if $q>p$. Looking at the matter from this standpoint, it is evident that any degenerate multinormal distribution may be regarded, through its characteristic function, as having arisen by a singular transformation from a non-degenerate one. Such singularity causes no difficulty in practice, since a distribution can be made non-singular by discarding redundant variates.

An important corollary of the result of this section, obtained by making $\mathbf{A}$ an identity matrix with $q\leqslant p-1$, is that the (marginal) distribution of any subset of the original variates is multinormal, with variances and covariances unchanged, although the inverse of their dispersion matrix, and hence the coefficients in the exponent of their (marginal) frequency function, will of course be changed.

The multivariate normal integral

15.5 Apart from the location and scale parameters, the frequency function (15.19) contains $\frac{1}{2}p(p-1)$ parameters, namely the covariances between the p variates. It is thus clear that tabulation of its distribution function is a formidable undertaking for $p > 2$. Bivariate tables are given in *Tables of the Bivariate Normal Distribution and Related Functions* (National Bureau of Standards, Applied Mathematics Series, 50, Washington, 1959) for $\pm\rho = 0(0{\cdot}05)0{\cdot}95(0{\cdot}01)1$ and variates in the range $0(0{\cdot}1)4$, to 6 or 7 d.p. Zelen and Severo (1960, *Ann. Math. Statist.*, **31**, 619) give charts for reading the bivariate normal integral value with an error of 1 per cent. or less.

Kendall (1941) gave an expression for the distribution function in terms of an infinite series, but this, though convergent, converges too slowly to be of general use. In the general case, Plackett (1954) has given a reduction formula which enables the three- and four-variable cases to be tabulated as required by using tabulated integrals.

D. B. Owen (1957), *The Bivariate Normal Probability Integral*, Sandia Corporation, Washington, D.C., and (1956) *Ann. Math. Statist.*, **27**, 1075, gives tables for computing the bivariate integral over regions bounded by straight lines and certain other regions.

G. P. Steck (1958, *Ann. Math. Statist.*, **29**, 780) gives tables for computing the trivariate integral.

15.6 A special problem which frequently arises in statistical theory is the evaluation of the multivariate normal integral over that part of the range of variation for which all p variates are positive. Although this is a considerable simplification of the general problem, it still presents analytical difficulties for $p>3$, and a number of special devices have been used to overcome them in particular cases.

We deal first with the two- and three-variate cases, the only ones for which analytical solutions of the problem are known.

In the two-variate case, we write the frequency function as the Fourier transform of its characteristic function and evaluate its integral over the positive quadrant as

$$\mathrm{P}_2^0 = \int_0^\infty \int_0^\infty \left\{ \frac{1}{(2\pi)^2} \int_{-\infty}^\infty \int_{-\infty}^\infty \exp(-i\mathbf{t}'\mathbf{x} - \tfrac{1}{2}\mathbf{t}'\mathbf{V}\mathbf{t})\, dt_1\, dt_2 \right\} dx_1\, dx_2, \qquad (15.24)$$

where the vectors and matrix are now of order two. Reversing the order of integration in (15.24), we integrate out x_1 and x_2 and obtain

$$4\pi^2 \mathrm{P}_2^0 = \int_{-\infty}^\infty \int_{-\infty}^\infty \exp(-\tfrac{1}{2}\mathbf{t}'\mathbf{V}\mathbf{t}) \frac{dt_1}{it_1} \frac{dt_2}{it_2}. \qquad (15.25)$$

The coefficient of $(-t_1 t_2)$ in the exponent of the integrand is the covariance of the two variates, which is (see Example 15.1) $\rho\sigma_1\sigma_2$. Differentiating (15.25) with respect to ρ, we obtain

$$4\pi^2 \frac{\partial \mathrm{P}_2^0}{\sigma_1 \sigma_2 \partial \rho} = \int_{-\infty}^\infty \int_{-\infty}^\infty \exp(-\tfrac{1}{2}\mathbf{t}'\mathbf{V}\mathbf{t})\, dt_1\, dt_2. \qquad (15.26)$$

The integrand on the right of (15.26) is a bivariate normal form in t_1 and t_2 apart from the constant term. The integral in (15.26) therefore equals the reciprocal of that constant, and (15.26) becomes

$$\frac{\partial \mathrm{P}_2^0}{\sigma_1 \sigma_2 \partial \rho} = \frac{1}{4\pi^2} \cdot \frac{2\pi}{|V|^{\frac{1}{2}}} = \frac{1}{2\pi |V|^{\frac{1}{2}}} = \{2\pi\sigma_1\sigma_2(1-\rho^2)^{\frac{1}{2}}\}^{-1}, \qquad (15.27)$$

from Example 15.1. On integration, (15.27) gives

$$\mathrm{P}_2^0 = \text{arc } \sin\rho/(2\pi) + \text{a constant},$$

and the constant is seen to be $\frac{1}{4}$ by putting $\rho = 0$. Thus

$$\mathrm{P}_2^0 = \tfrac{1}{4} + \text{arc } \sin\rho/(2\pi). \qquad (15.28)$$

(15.28) is usually called Sheppard's (1898) theorem on median dichotomy. By symmetry, the same result holds for the negative quadrant, while the two " mixed " quadrants have content $(\frac{1}{2} - \mathrm{P}_2^0)$ each. P_2^0 is tabulated by National Bureau of Standards (1959)—cf. reference in **15.5**—for $\rho = 0(0{\cdot}01)1$.

15.7 The corresponding three-variate result is derived from this fairly easily. We have seen (Exercise 7.1) that the probability of occurrence of at least one of a set of p events is

$$\Sigma P_1 - \Sigma P_2 + \Sigma P_3 - \ldots + (-1)^{p-1} P_p, \tag{15.29}$$

the summations being of the probabilities of occurrence of the events singly, in pairs, and so on to the last term, which contains the probability that all p events occur jointly.

If these p events are the positiveness of p multinormal variates, the probability that at least one is positive is the complement of the probability that none is positive, and, by symmetry, also the complement of the probability that none is negative, which is what we seek. Using (15.29) and writing the zero superscript as before, we therefore have

$$1 - P_p^0 = \Sigma P_1^0 - \Sigma P_2^0 + \Sigma P_3^0 \ldots + (-1)^{p-1} P_p^0. \tag{15.30}$$

If p is even, the last term on the right has a negative sign, and (15.30) is an identity; but if p is odd, it becomes

$$P_p^0 = \tfrac{1}{2}\{1 - \Sigma P_1^0 + \Sigma P_2^0 \ldots + (-1)^{p-1} \Sigma P_{p-1}^0\}. \tag{15.31}$$

(15.31) enables us to obtain P_p^0 for odd p from lower-order probabilities. With $p = 3$, it gives

$$\begin{aligned} P_3^0 &= \tfrac{1}{2}(1 - \Sigma P_1^0 + \Sigma P_2^0) = \tfrac{1}{2}\{1 - 3 . \tfrac{1}{2} + \Sigma(\tfrac{1}{4} + \text{arc } \sin\rho)/(2\pi)\} \\ &= \tfrac{1}{8} + \frac{1}{4\pi}(\text{arc } \sin\rho_{12} + \text{arc } \sin\rho_{13} + \text{arc } \sin\rho_{23}), \end{aligned} \tag{15.32}$$

a very simple generalization of Sheppard's formula.

15.8 No such simple result is available for $p > 3$. The best that has been achieved up to the present in the general case is an expansion in a power series in the covariance parameters by Kendall (1941). We may assume all variances to be unity, without losing generality. The Fourier inversion of the characteristic function (cf. (15.24)) can then be expanded as a power series in the correlations ρ_{ij}. Thus

$$\begin{aligned} P_p^0 &= \int_0^\infty \cdots \int_0^\infty \left\{ \frac{1}{(2\pi)^p} \int_{-\infty}^\infty \cdots \int_{-\infty}^\infty \exp(-i\mathbf{t}'\mathbf{x} - \tfrac{1}{2}\mathbf{t}'\mathbf{V}\mathbf{t})\, dt_1 \ldots dt_p \right\} dx_1 \ldots dx_p \\ &= \int_0^\infty \cdots \int_0^\infty \left\{ \frac{1}{(2\pi)^p} \int_{-\infty}^\infty \cdots \int_{-\infty}^\infty \exp(-i\mathbf{t}'\mathbf{x} - \tfrac{1}{2}\mathbf{t}'\mathbf{t}) . \Sigma \left[(-1)^{n..} \frac{\rho_{12}^{n_{12}} \rho_{13}^{n_{13}} \ldots}{n_{12}!\, n_{13}! \ldots} \right] \right. \\ &\qquad \left. \times t_1^{n_1.} t_2^{n_2.} \ldots t_p^{n_p.}\, dt_1 \ldots dt_p \right\} dx_1 \ldots dx_p, \end{aligned} \tag{15.33}$$

where the summation is over all possible sets of the ρ_{ij} taken over all non-negative values of the n_{ij}; $n_{i.} = \sum_j (n_{ij} + n_{ji})$ and $n_{..} = \sum_i n_{i.}$.

Interchanging summation and integration, we may re-write (15.33) as

$$P_p^0 = \Sigma \left[(-1)^{n..} \frac{\rho_{12}^{n_{12}} \rho_{13}^{n_{13}} \ldots}{n_{12}!\, n_{13}! \ldots} \right] \prod_{j=1}^{p} G_{n_j.}, \tag{15.34}$$

where

$$G_{n_j.} = \frac{1}{2\pi} \int_0^\infty \int_{-\infty}^\infty t_j^{n_j.} \exp(-it_j x_j - \tfrac{1}{2}t_j^2)\, dt_j\, dx_j. \tag{15.35}$$

If $n_{j.} = 0$, $G_{n_{j.}} = \frac{1}{2}$ from (15.35). If $n_{j.}$ is an integer, we integrate for x_j as at (15.24) to obtain

$$G_{n_{j.}} = \frac{1}{2\pi i}\int_{-\infty}^{\infty} t_j^{n_{j.}-1} e^{-\frac{1}{2}t_j^2} dt_j,$$

and by the formulae for the moments of the univariate normal distribution (5.60), we have

$$G_{n_{j.}} \begin{cases} = 0 & \text{if } n_{j.} \text{ is even} \\ = \dfrac{(2m)!}{\sqrt{(2\pi)}\, i\, 2^m m!} & \text{if } n_{j.} \text{ is odd} = 2m+1. \end{cases}$$

Thus finally

$$G_{n_{j.}} = \begin{cases} \frac{1}{2} & \text{if } n_{j.} = 0 \\ 0 & \text{if } n_{j.} = 2m \quad (m = 1,2\ldots) \\ \dfrac{(2m)!}{\sqrt{(2\pi)}\, i\, 2^m m!} & \text{if } n_{j.} = 2m+1 \quad (m = 0,1,2\ldots) \end{cases} \tag{15.36}$$

Thus partitions vanish from (15.34) which contain any t_i an even number of times.

(15.34) and (15.36) formally solve the problem. But although the series (15.34) for P_p^0 always converges, it does so only slowly if the ρ_{ij} are not small. Moreover, the terms in (15.34) are not all positive. In the four-variate case, e.g. Moran (1948), (15.34) becomes

$$P_4^0 = \frac{1}{16} + \frac{1}{8\pi}\sum_{i,j}\rho_{ij} + \frac{1}{4\pi^2}\sum_{ijkl}\rho_{ij}\rho_{kl} + \ldots$$

but the coefficient of $\rho_{12}\,\rho_{13}\,\rho_{14}$ is negative, being $-1/(4\pi^2)$.

15.9 In the special case when all the correlations are equal, a remarkable simplification is possible.

Let $x_0, x_1, \ldots, x_p$ be a set of $p+1$ independent normal variates with zero means and unit variances. Consider the derived set of variables

$$y_i = x_i - bx_0, \qquad i = 1, \ldots p, \qquad b > 0.$$

These will be multinormally distributed with zero means, variances $(1+b^2)$ and covariances b^2. The correlation between any pair of y's is thus

$$\rho = b^2/(1+b^2). \tag{15.37}$$

This ranges from zero when $b = 0$ to 1 when $b \to \infty$. The y_i can thus be taken to represent the general multinormal distribution with equal (positive) correlations. The value of P_p^0 is immediately deducible from the construction of the distribution; it is simply

$$P_p^0 = \int_{-\infty}^{\infty} (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}x^2} \left\{ \int_{bx}^{\infty} (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}t^2}\, dt \right\}^p dx, \tag{15.38}$$

the probability that p independent normal variates exceed bx, integrated over a standard normal distribution. This is an integral of the Hojo type (14.5) and can be simply calculated with little error by replacing the integration by a summation. Evidently,

the argument leading to the result is basically unaffected if some y_i are defined by

$$y_i = x_i + bx_0$$

so that some of the y_i have negative correlations with the others, of magnitude $-\rho$. In this case, of course, (15.38) must be suitably modified.

Ruben (1954) has tabulated P_p^0 in the equal-correlations case for $\rho = 1/j$, $p = 1, 2, \ldots, 51-j$, where $j = 2, 3, \ldots, 12$.

Example 15.2

To illustrate the approximation of the outer integral in (15.38) by a sum, we consider the case $p = 2$ when $b = 1$ in (15.37), i.e. $\rho = \frac{1}{2}$. Replacing the integral by the sum of the values at $x = \pm 3, \pm 2, \pm 1, 0$, we find from Appendix Tables 1 and 2 the following results :—

x	(1) $= \int_x^\infty (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}t^2}\,dt$	(2) $= \{(1)\}^2$	(3) = Sums of (2)	(4) $= (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}x^2}$	(3) × (4)
−3	0·99865	0·99730	0·99730	0·00443	0·00442
+3	0·00135	0·00000			
−2	0·97725	0·95502	0·95554	0·05399	0·05159
+2	0·02275	0·00052			
−1	0·84134	0·70785	0·73302	0·24197	0·17737
+1	0·15866	0·02517			
0	0·50000	0·25000	0·25000	0·39894	0·09973
					0·33311

The value obtained is 0·3331 to 4 d.p. The exact value, given by (15.28), is $\frac{1}{3}$. Greater accuracy could have been obtained by using a sum of more than seven values. For example, if we add in a further pair of terms for $x = \pm 4$, they contribute 0·00013 to the sum, giving a total of 0·33324. To 5 d.p., no further gain in accuracy results from adding terms for $x = \pm 5$, but accuracy as great as is desired could be obtained by spacing the values at narrower intervals, say 0·5 instead of 1.

Moran (1956) gives a bound for the relative error in the sum compared to the integral. It is, in our notation,

$$2\exp\left\{-\frac{\pi^2}{h^2(1+pb^2)}\right\},$$

where h is the interval of summation. In our example, this is

$$2\exp\left(-\frac{\pi^2}{3h^2}\right).$$

With $h = 1$, this gives a bound considerably in excess of the error of about 1 in 10,000 obtained above with nine values. With smaller h, the bound decreases sharply and also seems to be closer to the actual error.

Quadratic forms in normal variates

15.10 Suppose that $\mathbf{x}$ is distributed with the multinormal density

$$dF \propto \exp(-\tfrac{1}{2}\mathbf{x}'\mathbf{A}\mathbf{x}) \prod_{j=1}^{p} dx_j,$$

where $\mathbf{A}$ is not necessarily non-singular. Consider the distribution of the quadratic form $\mathbf{x}'\mathbf{A}\mathbf{x}$ itself. Its characteristic function is, writing $\theta = it$,

$$\phi(t) \propto \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\{-\tfrac{1}{2}\mathbf{x}'\mathbf{A}\mathbf{x}(1-2\theta)\} \prod_j dx_j. \tag{15.39}$$

The transformation (15.3) orthogonalizes the quadratic form as in **15.2** and (15.39) becomes

$$\phi(t) \propto \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\{-\tfrac{1}{2}\mathbf{y}'\mathbf{C}\mathbf{y}(1-2\theta)\} \prod_j dy_j. \tag{15.40}$$

The number of non-zero components of $\mathbf{y}$ is precisely the number of non-zero latent roots of $\mathbf{A}$, say r. If we integrate out the $(p-r)$ variables corresponding to the zero latent roots, the exponent of (15.40) is unaffected. To the remaining r-fold integral, we apply the transformation

$$z_j = y_j c_{jj}(1-2\theta)^{\frac{1}{2}}$$

to obtain

$$\phi(t) \propto (1-2\theta)^{-\frac{1}{2}r} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp(-\tfrac{1}{2}\mathbf{z}'\mathbf{z}) \prod_j dz_j. \tag{15.41}$$

The integral in (15.41) is independent of θ. Thus

$$\phi(t) = (1-2\theta)^{-\frac{1}{2}r}, \tag{15.42}$$

the elimination of constants being verified by the fact that $\phi(0) = 1$. By the Inversion Theorem and Example 3.6, (15.42) shows that the distribution of the quadratic form in the exponent of a multinormal distribution is that of a distribution known as the χ^2 distribution with r degrees of freedom defined by (11.36) with $n-p = r$ (see also Chapter 16), where r is the *rank* (of the quadratic form) of the distribution. This generalizes the result of (11.8), which deals effectively with the case where $\mathbf{A}$ is diagonal and non-singular.

Example 15.3 The χ^2 goodness-of-fit distribution

Given a sample of n observations from a multinomial distribution (**5.30**) with probabilities p_i $(i = 1, \ldots, k)$, it follows from the multivariate form of the Central Limit theorem that the observed numbers x_i in the k groups will tend, with increasing n, to have a multivariate normal distribution, with means np_i and dispersion matrix given (as at (5.80)) by

$$\mathbf{V} = n\begin{pmatrix} p_1(1-p_1), & -p_1p_2, & -p_1p_3, & \ldots, & -p_1p_k \\ -p_2p_1, & p_2(1-p_2), & -p_2p_3, & \ldots, & -p_2p_k \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ -p_kp_1, & -p_kp_2, & \cdots & \ldots, & p_k(1-p_k) \end{pmatrix}$$

Because of the linear restriction $\sum_{j=1}^{k} x_j = n$, the multinormal distribution is singular, of rank $(k-1)$, so $|\mathbf{V}| = 0$ and we cannot invert $\mathbf{V}$. However, we may omit one of the x_i (say x_k), as it is redundant, the remaining $(k-1)$ x_i being multinormally distributed with means np_i and dispersion matrix $\mathbf{V}^*$ which is simply $\mathbf{V}$ with its last row and column eliminated. It is easily verified by a direct matrix multiplication that

$$(\mathbf{V}^*)^{-1} = \frac{1}{n}\begin{pmatrix} \frac{1}{p_1}+\frac{1}{p_k}, & \frac{1}{p_k}, & \frac{1}{p_k}, & \cdots & \frac{1}{p_k} \\ \frac{1}{p_k}, & \frac{1}{p_2}+\frac{1}{p_k}, & \frac{1}{p_k}, & \cdots & \frac{1}{p_k} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \frac{1}{p_k}, & \frac{1}{p_k}, & \frac{1}{p_k}, & \cdots & \frac{1}{p_{k-1}}+\frac{1}{p_k} \end{pmatrix}$$

The exponent of the multinormal distribution thus has the quadratic form

$$\begin{aligned}(\mathbf{x}-\boldsymbol{\mu})'(\mathbf{V}^*)^{-1}(\mathbf{x}-\boldsymbol{\mu}) &= \sum_{i=1}^{k-1}\frac{(x_i-np_i)^2}{np_i} + \frac{1}{np_k}\sum_{i=1}^{k-1}\sum_{j=1}^{k-1}(x_i-np_i)(x_j-np_j) \\ &= \sum_{i=1}^{k-1}\frac{(x_i-np_i)^2}{np_i} + \frac{1}{np_k}\left\{\sum_{i=1}^{k-1}(x_i-np_i)\right\}^2 \\ &= \sum_{i=1}^{k}\frac{(x_i-np_i)^2}{np_i}.\end{aligned}$$

By the result of **15.10**, this variate is distributed like χ^2 with $(k-1)$ degrees of freedom. This result, due to Karl Pearson (1900), is the basis of the χ^2 goodness-of-fit test which we shall discuss in Volume 2.

15.11 We now suppose that $\mathbf{x}$ is a vector of p independent standardized normal variates, and consider the distribution of the general quadratic form

$$Q = \mathbf{x}'\mathbf{A}\mathbf{x}.$$

By the orthogonal transformation (15.3), $\mathbf{x}$ is transformed to $\mathbf{y}$, a vector of r $(\leqslant p)$ independent standardized normal variates, and

$$Q = \sum_{i=1}^{r} a_i y_i^2, \tag{15.43}$$

where the a_i are the non-zero latent roots of $\mathbf{A}$. The cumulant-generating function of $a_i y_i^2$, which is a multiple of a χ^2 variable with one degree of freedom, is (cf. **15.10** and Chapter 16), with $\theta = it$,

$$\psi_i(t) = -\tfrac{1}{2}\log(1-2a_i\theta), \tag{15.44}$$

and that of Q is, by the additive property of cumulant-generating functions, obtained from (15.43) and (15.44) as

$$\psi(t) = -\tfrac{1}{2}\sum_{i=1}^{r}\log(1-2a_i\theta). \tag{15.45}$$

From (15.45), it follows that the cumulants of Q are

$$\kappa_s = 2^{s-1}(s-1)! \sum_{i=1}^{r} a_i^s. \tag{15.46}$$

Now $\sum_{i=1}^{r} a_i^s$ is the sum of the elements of the leading diagonal of the sth power of $\mathbf{A}$, and we write this sum as $\operatorname{tr}\mathbf{A}^s$. Thus (15.46) is

$$\kappa_s = 2^{s-1}(s-1)!\operatorname{tr}\mathbf{A}^s. \tag{15.47}$$

(15.46) makes it clear that the distribution of Q depends only on the values of the non-zero latent roots of $\mathbf{A}$.

Comparing (15.46) with the cumulants of a χ^2 variate (obtainable from (15.42) or from Chapter 16), we see that if and only if every $a_i = 1$, Q is distributed like χ^2 with r degrees of freedom. (In this case, (15.47) shows that $\mathbf{A}$ is an idempotent matrix, since $\mathbf{A} = \mathbf{A}^s$ for all s.) This is a special case of the result of **15.10.**

15.12 The characteristic function of Q, corresponding to the cumulant-generating function (15.45), is

$$\phi(t) = E\{\exp(\theta\,\mathbf{x}'\mathbf{A}\,\mathbf{x})\} = \frac{1}{(2\pi)^{\frac{1}{2}p}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\{-\tfrac{1}{2}\mathbf{x}'[\mathbf{I}-2\theta\,\mathbf{A}]\mathbf{x}\}\,dx_1 \ldots dx_n. \tag{15.48}$$

This is the integral of a multinormal distribution with $\mathbf{V}^{-1} = \mathbf{I}-2\theta\,\mathbf{A}$, apart from a factor $|V|^{-\frac{1}{2}}$. It therefore equals $|V|^{\frac{1}{2}}$, so that

$$\phi(t) = |\mathbf{I}-2\theta\,\mathbf{A}|^{-\frac{1}{2}}. \tag{15.49}$$

If we now turn to the joint distribution of two quadratic forms,

$$Q_1 = \mathbf{x}'\mathbf{A}\,\mathbf{x},$$
$$Q_2 = \mathbf{x}'\mathbf{B}\,\mathbf{x},$$

we see similarly from evaluating the expectation of $\exp(\theta_1 Q_1 + \theta_2 Q_2)$ that their joint c.f. is

$$\phi(t_1, t_2) = |\mathbf{I}-2\theta_1\mathbf{A}-2\theta_2\mathbf{B}|^{-\frac{1}{2}}, \tag{15.50}$$

and their joint cumulant-generating function is

$$\psi(t_1, t_2) = \sum_{s=1}^{\infty} 2^{s-1}\operatorname{tr}(\theta_1\mathbf{A}+\theta_2\mathbf{B})^s/s. \tag{15.51}$$

As we have seen at **4.16**, the necessary and sufficient condition for the independence of two variates is that their joint characteristic function factorize into their individual characteristic functions. (15.49) and (15.50) therefore yield, as a necessary and sufficient condition for the independence of Q_1 and Q_2,

$$|\mathbf{I}-2\theta_1\mathbf{A}|\,|\mathbf{I}-2\theta_2\mathbf{B}| = |\mathbf{I}-2\theta_1\mathbf{A}-2\theta_2\mathbf{B}| \tag{15.52}$$

or equivalently, since θ_1 and θ_2 are dummy variables,

$$|\lambda\mathbf{I}-2\theta_1\mathbf{A}|\,|\lambda\mathbf{I}-2\theta_2\mathbf{B}| = \lambda^p|\lambda\mathbf{I}-2\theta_1\mathbf{A}-2\theta_2\mathbf{B}|, \tag{15.53}$$

a result due to Cochran (1934).

15.13 An alternative way of stating the condition for the independence of Q_1 and Q_2 is that the product-cumulants of their joint distribution must all be equal to zero (Example 12.7). In the joint cumulant-generating function (15.51), the coefficient of $\theta_1^k \theta_2^l$ must be zero whenever k and $l \geqslant 1$. Thus we require

$$\operatorname{tr}(\mathbf{A}^k \mathbf{B}^l \mathbf{A}^m \mathbf{B}^n \ldots) = 0 \qquad (\text{all } k, l, m, n, \ldots \geqslant 1). \tag{15.54}$$

We now prove Craig's (1943) theorem: Q_1 and Q_2 are independent if and only if

$$\mathbf{AB} = \mathbf{0}. \tag{15.55}$$

The sufficiency of (15.55) for (15.54) is evident, for since we may commute matrices under the trace operation the matrix product in (15.54) may be written to include at least one factor $\mathbf{AB}$.

Its necessity follows on putting $k = l = 2$ and $m = n = \ldots = 0$ in (15.54), for

$$0 = \operatorname{tr}(\mathbf{A}^2 \mathbf{B}^2) = \operatorname{tr}(\mathbf{B}.\mathbf{A}^2 \mathbf{B}) = \operatorname{tr}\{(\mathbf{AB})'(\mathbf{AB})\} \tag{15.56}$$

since $\mathbf{A}$ and $\mathbf{B}$ are symmetric. Now the leading diagonal of $(\mathbf{AB})'(\mathbf{AB})$ contains only the squares of all the elements of $\mathbf{AB}$ which are, of course, real. Thus (15.56) implies that every element of $\mathbf{AB}$ is zero, i.e. that (15.55) holds. It is remarkable that (15.56) alone implies (15.55), no use having been made of the other relations in (15.54). This proof is due to Lancaster (1954).

15.14 Further, since $\mathbf{A}$ and $\mathbf{B}$ are symmetric, we may write

$$\mathbf{A} = \mathbf{TT}'$$
$$\mathbf{B} = \mathbf{UU}',$$

and

$$\mathbf{AB} = \mathbf{TT}'\mathbf{UU}'.$$

Hence

$$\begin{aligned} \operatorname{tr}(\mathbf{AB}) &= \operatorname{tr}(\mathbf{T}\,\mathbf{T}'\mathbf{U}.\,\mathbf{U}') = \operatorname{tr}(\mathbf{U}'\mathbf{T}.\mathbf{T}'\mathbf{U}) \\ &= \operatorname{tr}\{(\mathbf{T}'\mathbf{U})'(\mathbf{T}'\mathbf{U})\}. \end{aligned}$$

The leading diagonal of $(\mathbf{T}'\mathbf{U})'(\mathbf{T}'\mathbf{U})$ contains only the squares of all the elements of $(\mathbf{T}'\mathbf{U})$, but these are not necessarily all real. But if $\mathbf{A}$ and $\mathbf{B}$ are non-negative forms, the elements are all real, and hence

$$\operatorname{tr}(\mathbf{AB}) = \operatorname{tr}\{(\mathbf{T}'\mathbf{U})'(\mathbf{T}'\mathbf{U})\} = 0 \tag{15.57}$$

implies

$$\mathbf{T}'\mathbf{U} = \mathbf{0}$$

and hence

$$\mathbf{AB} = \mathbf{T}.\mathbf{T}'\mathbf{U}.\mathbf{U}' = \mathbf{0}.$$

Thus for non-negative quadratic forms, (15.57) implies (15.55) and independence, a result due to Matérn (1949).

In particular, $\mathbf{A}$ and $\mathbf{B}$ are non-negative if Q_1 and Q_2 are χ^2 variates. In this case, the equivalence of (15.57) and (15.55) may be seen directly since $\mathbf{A}$, $\mathbf{B}$ are idempotent and

$$\mathbf{AB} = \mathbf{A}^2\mathbf{B}^2,$$

so that (15.57) implies (15.56) and hence (15.55) and independence.

Example 15.4 The independence of sample mean and variance in normal samples

We have seen (Examples 11.3, 11.7) that in samples of size n from a univariate normal distribution, which without loss of generality we may take to have zero mean and unit variance, the sample mean $\bar{x}$ is itself normally distributed with mean 0 and variance $1/n$, and the sample sum of squares $\sum_{i=1}^{n}(x_i-\bar{x})^2$ is distributed like χ^2 with $(n-1)$ degrees of freedom. It follows that $n\bar{x}^2$ is also distributed like χ^2 with 1 degree of freedom. The quadratic forms $n\bar{x}^2$ and $\sum_{i=1}^{n}(x_i-\bar{x})^2$ have matrices

$$\mathbf{A} = \frac{1}{n}\mathbf{U}, \qquad \mathbf{B} = \mathbf{I}-\frac{1}{n}\mathbf{U},$$

where $\mathbf{U}$ is the matrix with unit for each of its elements. Thus

$$\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{A}-\mathbf{A}^2) = 0$$

by the idempotency of χ^2 variates. Hence, from (15.57), $n\bar{x}^2$ and $\sum_{i=1}^{n}(x_i-\bar{x})^2$ are independently distributed, as we have already seen in Example 11.3, where we obtained the result in a slightly different form.

15.15 As a consequence of Craig's condition (15.55), we may obtain a necessary and sufficient condition for the independence of a quadratic form $Q = \mathbf{x}'\mathbf{Ax}$ and a linear form $L = \mathbf{b}'\mathbf{x}$. For Q and L are independent if and only if Q and L^2 are independent, i.e. if $\mathbf{x}'\mathbf{Ax}$ and $(\mathbf{b}'\mathbf{x})'.\mathbf{b}'\mathbf{x}$ are. From (15.55), this will be so if and only if

$$\mathbf{A}.\mathbf{bb}' = \mathbf{0}. \tag{15.58}$$

From (15.58) it follows that

$$\mathbf{Abb}'\mathbf{b} = \mathbf{0}, \tag{15.59}$$

and as $\mathbf{b}'\mathbf{b}$ is a non-zero scalar, we may drop it from (15.59) to give

$$\mathbf{Ab} = \mathbf{0}. \tag{15.60}$$

This only establishes the necessity of (15.60) for (15.58). Its sufficiency follows immediately.

Example 15.5

We may apply (15.60) to obtain the result of Example 15.4. Here the quadratic form is, using $\mathbf{1}$ to denote a vector of units,

$$\sum_{i=1}^{n}(x_i-\bar{x})^2 = \mathbf{x}'(\mathbf{I}-\frac{1}{n}\mathbf{U})\mathbf{x},$$

and the linear form is

$$\bar{x} = \frac{1}{n}\mathbf{1}'\mathbf{x}.$$

The product of these matrices is

$$\left(\mathbf{I}-\frac{1}{n}\mathbf{U}\right)\cdot\frac{1}{n}\mathbf{1} = \frac{1}{n}\mathbf{1}-\frac{1}{n^2}.n\mathbf{1} = \mathbf{0},$$

so (15.60) is obeyed and $\sum_{i=1}^{n}(x_i-\bar{x})^2$ and $\bar{x}$ are independent.

The decomposition of a sum of squares of independent normal variates

15.16 We now prove a remarkable result due to Cochran (1934), whose fundamental theorem has been amplified by James (1952). With $\mathbf{x}$ a vector of p independent standardized normal variates as before, suppose that the sum of squares $\mathbf{x}'\mathbf{x}$ is decomposed into k quadratic forms $Q_i = \mathbf{x}'\mathbf{A}_i\mathbf{x}$ with ranks r_i, i.e. that

$$\sum_{i=1}^{k} \mathbf{x}'\mathbf{A}_i\mathbf{x} = \mathbf{x}'\mathbf{I}\mathbf{x}. \tag{15.61}$$

Then any one of the following three conditions implies the other two :—

(a) The ranks of the Q_i add to p.
(b) Each of the Q_i is distributed like χ^2.
(c) Each is Q_i independent of every other.

We show that (b) is implied by (a) or (c), and implies them both, using a method of proof due to Lancaster (1954).

15.17 Select an arbitrary Q_i, say Q_1. If we make an orthogonal transformation (15.3) which diagonalizes $\mathbf{A}_1$, we obtain

$$\mathbf{y}'\mathbf{B}'\mathbf{A}_1\mathbf{B}\mathbf{y} + \mathbf{y}'\mathbf{B}'(\mathbf{I}-\mathbf{A}_1)\mathbf{B}\mathbf{y} = \mathbf{y}'\mathbf{I}\mathbf{y}, \tag{15.62}$$

the identity matrix, of course, remaining invariant under orthogonal transformation. Since the first and last terms in (15.62) are diagonal, so must the second be. Moreover, since $(p-r_1)$ of the leading diagonal elements of $\mathbf{B}'\mathbf{A}_1\mathbf{B}$ arezero, the corresponding elements of $\mathbf{B}'(\mathbf{I}-\mathbf{A}_1)\mathbf{B}$ are unity, and since by condition (a) the rank of the latter matrix is $(p-r_1)$, the other elements of its leading diagonal are zero, and the corresponding elements of $\mathbf{B}'\mathbf{A}_1\mathbf{B}$ are unity. Hence, from **15.11**, Q_1 is a χ^2 variable, $\mathbf{A}_1$ being idempotent. The same result holds for the other $\mathbf{A}_i$. Thus we have established (b) from (a).

15.18 Further, from (15.61)

$$\mathbf{I} = \sum_i \mathbf{A}_i, \tag{15.63}$$

and since the $\mathbf{A}_i$ are idempotent, (15.63) squared gives

$$\mathbf{I} = \sum_i \mathbf{A}_i + \sum_{i \neq j} \mathbf{A}_i\mathbf{A}_j$$

or

$$\sum_{i \neq j} \mathbf{A}_i\mathbf{A}_j = \mathbf{0}. \tag{15.64}$$

Taking traces on both sides of (15.64),

$$\sum_{i \neq j} \mathrm{tr}(\mathbf{A}_i\mathbf{A}_j) = 0. \tag{15.65}$$

But since the $\mathbf{A}_i$ are idempotent, and latent roots and traces are unaffected by orthogonal transformation, we have

$$\mathrm{tr}(\mathbf{A}_i\mathbf{A}_j) \geqslant 0 \qquad (i \neq j) \tag{15.66}$$

and (15.65) and (15.66) together imply

$$\text{tr}(\mathbf{A}_i\mathbf{A}_j) = 0 \qquad (\text{all } i \neq j) \tag{15.67}$$

which by **15.14** implies $\mathbf{A}_i\mathbf{A}_j = \mathbf{0}$ (all $i \neq j$), and the independence of Q_i from Q_j. Thus (c) is established from (b).

15.19 If (b) holds, $\mathbf{A}_i$ has r_i latent roots of unity and $(p-r_i)$ of zero, and from (15.63) it follows, on taking traces of both sides, that

$$p = \sum_{i=1}^{k} r_i,$$

which establishes (a) from (b).

If (c) holds, taking powers of (15.63) gives

$$\sum_i \mathbf{A}_i^s = \mathbf{I} \qquad (\text{all } s), \tag{15.68}$$

and taking traces of both sides of (15.68) gives

$$\text{tr}\sum \mathbf{A}_i^s = p \qquad (\text{all } s). \tag{15.69}$$

(15.69) can only hold if every latent root of each $\mathbf{A}_i$ is unity, i.e. if each Q_i is distributed like χ^2. This establishes (b) from (c).

We have thus seen that (a) implies (b) and conversely, and also that (b) implies (c) and conversely. This establishes the sufficiency of any one of the conditions for the other two. See also Exercise 15.16.

It is not true generally that if two variates are distributed like χ^2 and their sum is also distributed like χ^2 with degrees of freedom equal to the sum of their individual degrees of freedom, then the variates are independent. Example 7.6 is a counterexample due to James (1952): the restriction to quadratic forms in standard normal variates is crucial here.

Example 15.6

Reverting to the distribution of mean and variance in normal samples, discussed in Example 15.4, we see that the decomposition of the sum of squares

$$\sum_{i=1}^{n}(x_i-\bar{x})^2 + n\bar{x}^2 = \sum_{i=1}^{n} x_i^2$$

obeys the conditions for the application of Cochran's theorem. Condition (b) of **15.16** was established in Examples 11.3 and 11.7, and immediately implies the independence of the two sums on the left. It is easily verified that their ranks are $(n-1)$ and 1 respectively (agreeing with the degrees of freedom in their χ^2 distributions), and this satisfaction of condition (a) is enough to imply both their χ^2 distributions and their independence.

15.20 From **15.11** onwards, we have supposed $\mathbf{x}$ to be a vector of independent standardized normal variates, but no essential difference is made to our results if we take $\mathbf{x}$ to be multinormally and non-singularly distributed with mean $\mathbf{0}$ and dispersion matrix $\mathbf{V}$. For since $\mathbf{V}$ is positive definite, we may write

$$\mathbf{V} = \mathbf{T}\mathbf{T}' \tag{15.70}$$

where $\mathbf{T}$ has real elements, and the transformation $\mathbf{x} = \mathbf{T}\mathbf{y}$ transforms the exponent

of the distribution of $\mathbf{x}$ from $\mathbf{x}'\mathbf{V}^{-1}\mathbf{x}$ to

$$\mathbf{y}'\mathbf{T}'\mathbf{V}^{-1}\mathbf{T}\mathbf{y} = \mathbf{y}'\mathbf{T}'(\mathbf{T}')^{-1}\mathbf{T}^{-1}\mathbf{T}\mathbf{y} = \mathbf{y}'\mathbf{y},$$

using (15.70), and we may treat the vector of independent variates $\mathbf{y}$ as before. Thus the quadratic forms $\mathbf{x}'\mathbf{A}\mathbf{x}$, $\mathbf{x}'\mathbf{B}\mathbf{x}$ become $\mathbf{y}'.\mathbf{T}'\mathbf{A}\mathbf{T}.\mathbf{y}$ and $\mathbf{y}'.\mathbf{T}'\mathbf{B}\mathbf{T}.\mathbf{y}$ respectively, and Craig's condition (15.55) becomes, using (15.70),

$$\mathbf{0} = \mathbf{T}'\mathbf{A}\mathbf{T}.\mathbf{T}'\mathbf{B}\mathbf{T} = \mathbf{T}'.\mathbf{A}\mathbf{V}\mathbf{B}.\mathbf{T}$$

or

$$\mathbf{A}\mathbf{V}\mathbf{B} = \mathbf{0}, \tag{15.71}$$

a generalization due to Aitken (1950). Similarly, Lancaster's condition (15.56) becomes

$$0 = \text{tr}\,\{(\mathbf{T}'\mathbf{A}\mathbf{T})^2(\mathbf{T}'\mathbf{B}\mathbf{T})^2\} = \text{tr}\,\{(\mathbf{A}\mathbf{V})^2(\mathbf{B}\mathbf{V})^2\}, \tag{15.72}$$

and Matérn's condition (15.57) for non-negative forms becomes

$$0 = \text{tr}\,\{(\mathbf{T}'\mathbf{A}\mathbf{T})(\mathbf{T}'\mathbf{B}\mathbf{T})\} = \text{tr}\,(\mathbf{A}\mathbf{V}\mathbf{B}\mathbf{V}). \tag{15.73}$$

15.21 Cochran's theorem of **15.16** also holds for the more general situation where, in (15.61), the right-hand side is replaced by any quadratic form distributed like χ^2. For we may transform this to a sum of squares as above and the ranks of the singular matrices on the left of (15.61) are unaffected by the non-singular transformation. Conditions (a), (b) and (c) of **15.16** on the untransformed matrices are then equivalent to the same conditions on the transformed matrices, and we may therefore apply Cochran's theorem directly to the former.

If the x_i are not independent, the theorem still holds with obvious modifications (Exercise 15.17).

Characterizations of the normal distribution

15.22 We have seen in Example 12.7 that the independence of sample mean and variance characterizes the univariate normal distribution—no other distribution with finite cumulants of all orders possesses this property. This result may be generalized to the multivariate case in the form : the independence of the sample mean vector and the elements of the sample dispersion matrix characterizes the multinormal distribution among all multivariate distributions with finite dispersion matrices. It will be noticed that the existence of higher-order cumulants is not necessary to the result, which is due to Lukacs (1942). We first give his proof for the univariate case, and sketch the generalization.

15.23 We suppose a population to have finite mean μ and variance σ^2, and characteristic function (c.f.) $\phi(t)$. The joint c.f. of the sample mean $\bar{x}$ and variance s^2 in samples of size n is

$$\phi_{12}(t_1, t_2) = \int \cdots \int \exp(it_1\bar{x} + it_2 s^2)\, dF_n. \tag{15.74}$$

A necessary and sufficient condition for the independence of $\bar{x}$ and s^2 is that (15.74) factorizes, i.e. that

$$\phi_{12}(t_1, t_2) = \phi_1(t_1)\phi_2(t_2). \tag{15.75}$$

Differentiating (15.75) with respect to t_2 we obtain

$$\left[\frac{\partial \phi_{12}}{\partial t_2}\right]_{t_2=0} = \phi_1(t_1)\left[\frac{\partial \phi_2}{\partial t_2}\right]_{t_2=0}. \tag{15.76}$$

Now in (15.76), we may use the relation connecting the c.f. of the sample mean with that of the parent distribution

$$\phi_1(t_1) = \left\{\phi\left(\frac{t_1}{n}\right)\right\}^n \tag{15.77}$$

Further
$$\left[\frac{\partial \phi_2}{\partial t_2}\right]_{t_2=0} = i\mathrm{E}(s^2) = i.\frac{(n-1)}{n}\sigma^2. \tag{15.78}$$

Finally, from (15.74),

$$\frac{\partial \phi_{12}}{\partial t_2} = i\int \cdots \int s^2 \exp(it_1\bar{x} + it_2 s^2)\, dF_n, \tag{15.79}$$

and since

$$s^2 = \frac{1}{n}\sum_i x_i^2 - \bar{x}^2 = \frac{1}{n^2}\left\{(n-1)\sum_i x_i^2 - \sum\sum_{i\neq j} x_i x_j\right\}, \tag{15.80}$$

(15.79) and (15.80) give

$$\begin{aligned}\left[\frac{\partial \phi_{12}}{\partial t_2}\right]_{t_2=0} &= \frac{i}{n^2}\int \cdots \int \left\{(n-1)\sum_i x_i^2 - \sum\sum_{i\neq j} x_i x_j\right\} \exp\left(it_1 \sum_i x_i/n\right) dF_n \\ &= \frac{i(n-1)}{n}\left\{\int x^2 \exp(it_1 x/n)\, dF_1 . \left[\phi\left(\frac{t_1}{n}\right)\right]^{n-1} \right. \\ &\qquad \left. - \left[\int x \exp(it_1 x/n)\, dF_1\right]^2 \left[\phi\left(\frac{t_1}{n}\right)\right]^{n-2}\right\} \\ &= \frac{i(n-1)}{n}\left[\phi\left(\frac{t_1}{n}\right)\right]^{n-2} . \left\{-\phi\left(\frac{t_1}{n}\right) . \frac{\partial^2 \phi\left(\frac{t_1}{n}\right)}{\partial\left(\frac{t_1}{n}\right)^2} + \left[\frac{\partial \phi\left(\frac{t_1}{n}\right)}{\partial\left(\frac{t_1}{n}\right)}\right]^2\right\}. \end{aligned} \tag{15.81}$$

Writing t for $\frac{t_1}{n}$, we put (15.81), (15.77) and (15.78) into (15.76) and obtain

$$-\phi(t)\phi''(t) + \{\phi'(t)\}^2 = \{\phi(t)\}^2\sigma^2. \tag{15.82}$$

(15.82) may be rewritten

$$\frac{d}{dt}\left\{\frac{d}{dt}\log\phi(t)\right\} = -\sigma^2, \tag{15.83}$$

so that integration of (15.83) gives

$$\phi'(t)/\phi(t) = -\sigma^2 t + c. \tag{15.84}$$

From the initial conditions

$$\phi(0) = 1,$$
$$\phi'(0) = i\mu,$$

we find in (15.84)

$$c = i\mu,$$

and (15.84) becomes

$$\frac{d}{dt}\log\phi(t) = i\mu - \sigma^2 t. \tag{15.85}$$

Integrating (15.85), we have

$$\log\phi(t) = i\mu t - \tfrac{1}{2}\sigma^2 t^2,$$

the constant of integration being zero since $\log \phi(0) = 0$. Thus finally

$$\phi(t) = \exp(i\mu t - \tfrac{1}{2}\sigma^2 t^2), \tag{15.86}$$

which is the c.f. unique to the normal distribution. (15.86) thus establishes the independence of sample mean and variance as a characteristic of the normal distribution alone among distributions with finite variance.(*)

15.24 The generalization of the result of **15.23** to the multivariate case is straightforward. If the distribution of the sample mean vector is independent of $s_{l,m}$, the sample covariance of x_l and x_m, we obtain, by the argument leading to (15.82),

$$-\phi''_{l,m}/\phi + \phi'_l \phi'_m/\phi^2 = \sigma_{lm}, \tag{15.87}$$

where ϕ is the c.f. of the parent distribution and suffixes denote the t-variable with respect to which differentiation has been carried out. If (15.87) holds for $l, m = 1, 2, \ldots, p$, we have a system of partial differential equations which may be written in matrix form, corresponding to (15.83),

$$\left(\frac{\partial}{\partial t_l}\frac{\partial}{\partial t_m}\log\phi\right) = -\mathbf{V}, \tag{15.88}$$

so that on applying the initial conditions, (15.88) yields after two integrations

$$\phi(\mathbf{t}) = \exp(i\mathbf{t}'\boldsymbol{\mu} - \tfrac{1}{2}\mathbf{t}'\mathbf{V}\mathbf{t}),$$

the c.f. of the multinormal distribution.

15.25 Other interesting and important characterizations of the normal distribution are known. For example, in our discussion of the geometrical method of deriving sampling distributions in Chapter 11, we saw that the probability density of a sample of independent observations from a standardized univariate normal distribution is constant on the sphere $\Sigma x^2 =$ constant, since the density is a function of Σx^2 only. The question arises whether this property of radial symmetry is possessed by any other distribution. The following argument, due to Bartlett (1934), shows that it is not.

If the density may be written

$$\mathrm{L} = \prod_{i=1}^{n} f(x_i) \propto g(\Sigma x^2), \tag{15.89}$$

we must have

$$\frac{\partial \mathrm{L}}{\partial x_i} = \frac{\partial \log \mathrm{L}}{\partial x_i} = 0$$

for

$$\sum_i x_i^2 = \text{constant}.$$

Using Lagrange's method with a multiplier λ, we therefore have

$$\frac{\partial \log f(x_i)}{\partial x_i} + \lambda x_i = 0 \qquad \text{(all } i\text{)}. \tag{15.90}$$

Integration yields

$$\log f(x_i) = -\tfrac{1}{2}\lambda x_i^2 + k_i$$

(*) Kawata and Sakamoto, 1949, *Jour. Math. Soc. Japan*, **1**, 111, relax even the restriction of requiring a finite variance.

or

$$f(x_i) \propto \exp(-\tfrac{1}{2}\lambda x_i^2). \tag{15.91}$$

(15.89) thus characterizes the normal distribution.

15.26 The argument generalizes at once to the multivariate case. Instead of radial symmetry with respect to a single vector of length $(\Sigma x^2)^{\frac{1}{2}}$, we now consider a group of p vectors. If the distribution has a density which is a function only of the squared lengths of these vectors (the sums of squares) and the angles between them (the correlation coefficients—cf. **16.24**), it must be multinormal. For if we define $\mathbf{x}_i$ as the $(p \times 1)$ vector containing the ith set of observations on the p variates,(*) and $\mathbf{X}'$ as the $(p \times n)$ matrix formed by these vectors, the sums of squares and products are the elements of $\mathbf{X}'\mathbf{X}$, and we have

$$\mathrm{L} = \prod_i f(\mathbf{x}_i) \propto g(\mathbf{X}'\mathbf{X}). \tag{15.92}$$

For $\mathbf{X}'\mathbf{X}$ constant, we now introduce a $(p \times p)$ matrix $\mathbf{\Lambda}$ of Lagrange multipliers, and differentiate logarithmically to obtain

$$\frac{\partial \log f(\mathbf{x}_i)}{\partial \mathbf{x}_i} + \mathbf{\Lambda}\mathbf{x}_i = \mathbf{0} \qquad (\text{all } i),$$

whence

$$f(\mathbf{x}_i) \propto \exp(-\tfrac{1}{2}\mathbf{x}_i'\mathbf{\Lambda}\mathbf{x}_i),$$

so that (15.92) characterizes the multinormal distribution.

15.27 Finally, we prove a characterization of the normal distribution which illuminates the position of central importance which it holds in statistical theory. We have seen (Examples 11.2 and 11.3) that an orthogonal transformation from a set of independent standardized normal variates yields a set of linear functions of the original variates which are also independent standardized normal variates. We now prove that if $\mathbf{x}$ is a vector of (not necessarily identical) independent standardized variates with finite cumulants, and the non-trivial transformation

$$\mathbf{x} = \mathbf{C}\mathbf{y} \tag{15.93}$$

gives a vector $\mathbf{y}$ of independent standardized variates, then each of the x_i is normal (and hence so is each y_i) and the transformation is orthogonal. The result is due to Lancaster (1954).

The orthogonality of $\mathbf{C}$ follows at once, for since the variates are standardized

$$\sum_j c_{ij}^2 = 1, \tag{15.94}$$

while since they are independent

$$\sum_j c_{ij}\, c_{kj} = 0 \qquad (i \neq k). \tag{15.95}$$

(15.94) and (15.95) are the conditions for orthogonality. Thus $\mathbf{C}' = \mathbf{C}^{-1}$ and

$$\mathbf{y} = \mathbf{C}'\mathbf{x}. \tag{15.96}$$

(*) These new vectors are, of course, distinct from the $(n \times 1)$ vectors of observations on single variates just discussed.

Since the y_i are independent, (15.93) gives for the sth cumulant of x_i, κ_{si},

$$\kappa_{si} = \sum_k c_{ik}^s \lambda_{sk} \qquad \text{(all } s\text{)}, \tag{15.97}$$

where λ_{sk} is the sth cumulant of y_k. Conversely, from (15.96),

$$\lambda_{si} = \sum_j (c'_{ij})^s \kappa_{sj} \qquad \text{(all } s\text{)}. \tag{15.98}$$

Combination of (15.97) and (15.98) gives

$$\kappa_{si} = \sum \kappa_{sj} \sum_k c_{ik}^s c_{jk}^s \qquad \text{(all } s\text{)}. \tag{15.99}$$

For $s = 1, 2$, (15.99) reduces to $\kappa_{1i} = 0$, $\kappa_{2i} = 1$, as it must for standardized variates. Now

$$\sum_j |\sum_k c_{ik}^s c_{jk}^s| \leqslant \sum_j \sum_k |c_{ik}^s||c_{jk}^s|, \tag{15.100}$$

and using (15.94), (15.100) becomes, for $s \geqslant 3$,

$$\sum_j |\sum_k c_{ik}^s c_{jk}^s| < 1. \tag{15.101}$$

Taking absolute values in (15.99), we have

$$|\kappa_{si}| \leqslant \sum_j |\kappa_{sj}||\sum_k c_{ik}^s c_{jk}^s|. \tag{15.102}$$

(15.102) has on the right-hand side a weighted sum of the κ_{sj}, the weights adding to less than 1 by (15.101). If for fixed s we take that value of i for which κ_{si} is largest, (15.102) cannot hold unless that $\kappa_{si} = 0$, i.e. all $\kappa_{si} = 0$. This holds for all s. Thus

$$\kappa_{si} = 0 \qquad \text{(all } s \geqslant 3\text{)} \tag{15.103}$$

and each x_i is normal. The normality of the y_i follows at once.

Lukacs (1956, *Proc. Third Berkeley Sympos.*, **2**, 195) reviews characterization methods and results generally. Lancaster (1960, *J. Austral. Math. Soc.*, **1**, 368) reviews the characterizations of normality by independence properties, and shows in particular that the latter imply the existence of all moments, which therefore need not be assumed.

EXERCISES

15.1 From a multivariate normal distribution, a conditional distribution is found by fixing the values of certain of the variates. Show that this distribution is itself multivariate normal, with a mean vector which is a function of the eliminated variates and with the same elements in the inverse variance matrix as the variates possessed in the original distribution.

15.2 A trivariate distribution, not necessarily normal, has correlations ρ_{12}, ρ_{13}, ρ_{23}. Show that $1 + 2\rho_{12}\rho_{13}\rho_{23} \geqslant \rho_{12}^2 + \rho_{13}^2 + \rho_{23}^2$.

15.3 In a p-variate distribution all the correlations ρ are equal. Show that this is possible if, and only if, $\rho \geqslant -1/(p-1)$.

15.4 Show that the frequency function of the multivariate normal distribution satisfies the system of equations

$$\frac{\partial f}{\partial \rho_{ij}} = \frac{\partial^2 f}{\partial x_i \, \partial x_j}$$

where ρ_{ij} is the correlation coefficient of x_i and x_j.

15.5 By considering the bivariate normal distribution with independent variates, show that

$$\int_{-x}^{x} (2\pi)^{-\frac{1}{2}} \exp(-\tfrac{1}{2}t^2)\, dt \leqslant \{1-\exp(-2x^2/\pi)\}^{\frac{1}{2}}.$$

(The error in taking the equality to hold is always less than $\frac{3}{4}$ per cent.)

(Williams, 1946)

15.6 By considering the area of a suitable spherical triangle, derive the probability (15.32) that in a trivariate normal population with zero means and unit variances the variate values are all positive.

15.7 Shots are fired at a vertical circular target of unit radius. The distribution of horizontal and vertical deviations from the centre of the target is bivariate normal, with zero means, equal variances v and correlation ρ. Show that the probability of hitting the target is

$$\frac{(1-\rho^2)^{\frac{1}{2}}}{\pi}\int_{1-\rho}^{1+\rho}\left[1-\exp\left\{-\frac{u}{2v(1-\rho^2)}\right\}\right]\frac{du}{u\,\{\rho^2-(1-u)^2\}^{\frac{1}{2}}}.$$

15.8 The variates x and y are non-singularly distributed in a bivariate normal form. Show that $x^2+2axy+y^2$ cannot be independent of $x^2+2bxy+y^2$ unless a and b are both unity in absolute value and of opposite sign.

15.9 By considering a sequence of $(p+1)$ identically distributed variates, show that for *any* continuous distribution of p variates with identical marginal distributions and correlations all equal to $\frac{1}{2}$, the probability that all the variates exceed their expectation is $1/(p+1)$.

(Foster and Stuart, 1954)

15.10 In the previous exercise, if, instead of the correlations being $\frac{1}{2}$, the dispersion matrix is given by

$$\mathbf{V} = \begin{pmatrix} 1 & -\frac{1}{2} & 0 & 0 & \cdots & 0 \\ -\frac{1}{2} & 1 & -\frac{1}{2} & 0 & \cdots & 0 \\ 0 & -\frac{1}{2} & 1 & -\frac{1}{2} & \cdots & 0 \\ 0 & 0 & -\frac{1}{2} & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdots & 1 \\ 0 & \cdot & \cdot & \cdot & \cdots & -\frac{1}{2} \end{pmatrix} \quad \begin{matrix} 0 \\ 0 \\ 0 \\ 0 \\ \cdot \\ \cdot \\ -\frac{1}{2} \\ 1 \end{matrix}$$

show that the probability is $1/(p+1)!$.

15.11 For a multivariate normal distribution with $\mathbf{V}^{-1}$ equal to $\mathbf{V}$ of the previous exercise, show that the probability that the variates are all non-negative is $1/(p+1)$.

(Anis and Lloyd, 1953)

15.12 By considering the variates $y_i = x_i - b_i x_0$, where the x_i are independent standardized normal variates with frequency function f, show that for a multinormal distribution with correlations $\rho_{ij} = a_i a_j$ $(i \neq j)$ the probability that all the variables are positive is

$$\int_{-\infty}^{\infty}\left\{\prod_i \int_{b_i x}^{\infty} f(t)\, dt\right\} f(x)\, dx.$$

15.13 If $\mathbf{x}$ has a multinormal distribution with mean $\mathbf{0}$ and dispersion matrix $\mathbf{V}$, show that a necessary and sufficient condition for the independence of a quadratic form $\mathbf{x}'\mathbf{Ax}$ and a linear form $\mathbf{b}'\mathbf{x}$ is that $\mathbf{AVb} = \mathbf{0}$.

(Aitken, 1950)

15.14 If $\mathbf{V}$ in the previous exercise is non-singular, with rank p, show that a necessary and sufficient condition that $\mathbf{x}'\mathbf{A}\mathbf{x}$ be distributed like χ^2 with s degrees of freedom is that $(\mathbf{V}^{-1}-\mathbf{A})$ has rank $p-s$.

15.15 If $\mathbf{x}'\mathbf{A}\mathbf{x}$, $\mathbf{x}'\mathbf{B}\mathbf{x}$ are distributed like χ^2 with m and n ($\geqslant m$) degrees of freedom, show that a necessary and sufficient condition that one be a component of the other is that their correlation coefficient takes the value $(m/n)^{\frac{1}{2}}$.

(Lancaster, 1954)

15.16 Show by modifying the proof in **15.18** that if Q_1 is a χ^2 variable and $(\mathbf{x}'\mathbf{I}\mathbf{x}-Q_1)$ is non-negative definite, the latter is distributed as χ^2 independently of Q_1 with degrees of freedom $(p-r_1)$.

(Hogg and Craig (1958), *Ann. Math. Statist*, **29**, 608)

15.17 If $\mathbf{x}$ is multinormal with mean $\mathbf{0}$ and dispersion matrix $\mathbf{V}$, and $\sum_{i=1}^{k}\mathbf{x}'\mathbf{A}_i\mathbf{x} = \mathbf{x}'\mathbf{A}\mathbf{x}$ where $\mathbf{A}\mathbf{V}$ is idempotent, show that any of the conditions (a), (b), (c) of **15.16** implies the other two if (b) is amended to require the idempotency of each $\mathbf{A}_i\mathbf{V}$ and (c) to require that $\mathbf{A}_i\mathbf{V}\mathbf{A}_j = \mathbf{0}$, all $i \neq j$.

15.18 x and y are distributed in a bivariate normal form with zero means, unit variance and correlation ρ. Show that the distribution of x and y, conditional upon $ax+by = 0$, may be regarded as the univariate normal distribution of $(bx-ay)$ with variance

$$(1-\rho^2)(a^2+b^2)^2/(a^2+2\rho ab+b^2).$$

15.19 $x_1, x_2, \ldots, x_n$ are distributed in a multivariate normal form, the covariance of x_i and x_j being c_{ij}. Show that their covariance, conditional upon $l_1x_1+l_2x_2+\ldots+l_nx_n =$ constant, is

$$c_{ij} - \frac{(\sum_k l_k c_{ik})(\sum_k l_k c_{jk})}{\sum\sum_{k\,m} l_k l_m c_{km}}$$

15.20 In a bivariate normal distribution with zero means, all pairs of values which are not both positive or both negative have the sign of x changed. Show that when this is done, the two variates are marginally normally distributed but that their joint distribution is not bivariate normal. Generalize to show that, for a set of n variates, it is possible for any subset of $p<n$ variates to be jointly normally distributed without the n variates being so distributed.

15.21 Use the result of Exercise 15.19 to establish, as in Example 15.3, that the sum $\sum_i \frac{(x_i-np_i)^2}{np_i}$ has a χ^2 distribution with $k-1$ degrees of freedom. (Consider k independent Poisson variables conditional upon $\sum_{i=1}^{k} x_i = n$.)

15.22 If $x_1, \ldots, x_k$ are independent variables having (not necessarily identical) Gamma-distributions, show by using characteristic functions that any scale-free function of them, $h(x_1, \ldots, x_k)$, is distributed independently of $S = \sum_{i=1}^{k} x_i$. In particular this holds for

$$h(x_1, \ldots, x_k) = \sum_{i=1}^{n} a_i x_i \Big/ \sum_{i=1}^{n} x_i = R.$$

(Pitman (1937), *Proc. Camb. Phil. Soc.*, **33**, 212. Laha (1954), *Ann. Math. Statist.*, **25**, 784, gives a converse: if the x_i are independently *and identically* distributed with finite variance, the independence of R and S implies that each x_i has a Gamma-distribution.)

CHAPTER 16

DISTRIBUTIONS ASSOCIATED WITH THE NORMAL

16.1 In the course of our investigations in earlier chapters, we have encountered several distributions, related to the normal distribution, which play important parts in the theory of statistics precisely because they are the forms taken by the sampling distributions of various statistics in samples from normal populations. The special position which the normal distribution holds, mainly by virtue of the Central Limit theorem in one or other of its forms, is reflected in the positions of central importance occupied by these related distributions. Whenever some statistic (under regularity conditions) has a sampling distribution which is asymptotically normal, other statistics exist which have the corresponding related distributions. We have already encountered an instance of this in discussing the χ^2 goodness-of-fit test in Example 15.3.

In this chapter, the three important distributions related to the univariate normal, namely the χ^2 distribution, " Student's " t-distribution, and Fisher's variance ratio (F) distribution, will in turn be examined, and the salient properties of each described. Later in the chapter, sampling distributions arising from the bivariate normal distribution will be discussed.

The χ^2 distribution

16.2 We have already seen (Example 11.6) that the sum z of the squares of n independent standardized normal variates is distributed with frequency function

$$dF = \frac{1}{2^{\frac{1}{2}n}\Gamma(\frac{1}{2}n)} \exp(-\tfrac{1}{2}z) z^{\frac{1}{2}n-1}\, dz, \qquad 0 \leqslant z \leqslant \infty, \tag{16.1}$$

and that if we impose p independent linear constraints upon these variates, i.e. consider the distribution of z conditional upon $p\,(<n)$ linear relations among the normal variates being satisfied, the effect of the constraints is simply to replace n in (16.1) by $(n-p)$. Further, we have also seen (Example 11.7) that if the sum of squares is taken about the sample mean instead of the population mean, the effect on the distribution of z is to impose one linear constraint, replacing n in (16.1) by $(n-1)$.

The distribution (16.1) is called the χ^2 distribution with n degrees of freedom. If rewritten, with $z=\chi^2$ and $n = \nu$, as

$$dF = \frac{1}{\Gamma(\frac{1}{2}\nu)} \exp(-\tfrac{1}{2}\chi^2)(\tfrac{1}{2}\chi^2)^{\frac{1}{2}\nu-1}\, d(\tfrac{1}{2}\chi^2), \qquad \nu>0\,;\ \ 0 \leqslant \chi^2 \leqslant \infty, \tag{16.2}$$

it is at once evident that $\frac{1}{2}\chi^2$ is distributed in the Gamma-distribution with parameter $\frac{1}{2}\nu$. (16.2) is a Pearson Type III distribution. Although we have so far considered distributions of the type (16.2) with ν a positive integer, we have taken the opportunity here to widen the definition of the distribution to all cases for which ν is positive. For such cases, the distribution function always converges, so that the distribution is always properly defined.

By differentiation of (16.2), we find that for $\nu > 2$, the χ^2 distribution has a mode at $\nu - 2$. For $\nu = 2$, the distribution is J-shaped, with maximum ordinate at zero, while for $0 < \nu < 2$ the distribution is J-shaped and has infinite ordinate at the origin.

Properties of the χ^2 distribution

16.3 The characteristic function of χ^2 is obtained at once from (16.1) (see Example 3.6) as

$$\phi(t) = \frac{1}{(1-2it)^{\frac{1}{2}\nu}} \tag{16.3}$$

whence, for the cumulants, we have

$$\kappa_r = \nu 2^{r-1}(r-1)! \tag{16.4}$$

and for moments about the mean

$$\left.\begin{aligned} \mu_2 &= 2\nu \\ \mu_3 &= 8\nu \\ \mu_4 &= 48\nu + 12\nu^2 \\ \mu_5 &= 32\nu(5\nu + 12) \\ \mu_6 &= 40\nu(3\nu^2 + 52\nu + 96). \end{aligned}\right\} \tag{16.5}$$

Since κ_r is linear in ν, μ_r, which can contain only $[\frac{1}{2}r]$ powers of μ_2, must be of degree $[\frac{1}{2}r]$ in ν, i.e. $\frac{1}{2}r$ if r is even and $\frac{1}{2}(r-1)$ if r is odd.

As ν tends to infinity the χ^2 distribution tends to normality, for in standard measure we have

$$\phi(t) = e^{-\frac{\nu it}{\sqrt{(2\nu)}}}\left(1 - \frac{2it}{\sqrt{(2\nu)}}\right)^{-\frac{1}{2}\nu}$$

and

$$\log\phi(t) = \frac{-\nu it}{\sqrt{(2\nu)}} - \frac{\nu}{2}\left\{\frac{-2it}{\sqrt{(2\nu)}} - \frac{1}{2}\left(\frac{2it}{\sqrt{(2\nu)}}\right)^2 + o(\nu^{-1})\right\}$$
$$\to -\tfrac{1}{2}t^2.$$

The tendency is, however, rather slow, and there are better approximations, as we shall see in a moment.

16.4 The distribution function of (16.2) is an incomplete Gamma-function. We have

$$F(\zeta) = \int_0^\zeta \frac{1}{2^{\frac{1}{2}\nu}\Gamma(\frac{1}{2}\nu)} \exp(-\tfrac{1}{2}z)\, z^{\frac{1}{2}\nu - 1}\, dz$$
$$= \Gamma_{\frac{1}{2}\zeta}(\tfrac{1}{2}\nu)/\Gamma(\tfrac{1}{2}\nu),$$

or, in the notation of Pearson's tables,

$$= I\left(\frac{\frac{1}{2}\zeta}{\sqrt{(\frac{1}{2}\nu)}},\ \tfrac{1}{2}\nu - 1\right).$$

Many special tables of the χ^2 distribution have been computed. Of these, the most useful and easily accessible are listed below:

(a) Pearson and Hartley (1950) give the distribution function of χ^2 for $\nu = 1(1)20(2)70$ and $\chi^2 = 0{\cdot}001(0{\cdot}001)0{\cdot}01(0{\cdot}01)0{\cdot}1(0{\cdot}1)2{\cdot}0(0{\cdot}2)10(0{\cdot}5)20(1)40(2)70$. This table is fully reproduced in the *Biometrika Tables*.

(b) Thompson (1941) gives an inverted table of the percentage points of the distribution, i.e. the values of χ^2 corresponding to the values $(1-P)$ of the distribution function, for $P = 0{\cdot}995$, $0{\cdot}990$, $0{\cdot}975$, $0{\cdot}950$, $0{\cdot}900$, $0{\cdot}750$, $0{\cdot}500$, $0{\cdot}250$, $0{\cdot}100$, $0{\cdot}050$, $0{\cdot}025$, $0{\cdot}010$, $0{\cdot}005$ and $\nu = 1(1)30(10)100$. This table, too, is reproduced in the *Biometrika Tables*, with the addition of entries for $P = 0{\cdot}001$.

(c) Fisher and Yates (1953) give an inverted table for $P = 0{\cdot}99, 0{\cdot}98, 0{\cdot}95, 0{\cdot}90(0{\cdot}10)0{\cdot}10$, $0{\cdot}05$, $0{\cdot}02$, $0{\cdot}01$, $0{\cdot}001$ and $\nu = 1(1)30$. A selection from this table is given as Appendix Table 3.

(d) A table by Yule, reproduced in Appendix Table 4A, gives P for $\nu = 1$, $\chi^2 = 0(0{\cdot}01)1(0{\cdot}1)10$.

16.5 Two approximations to the distribution of χ^2 are in common use, each of which selects an appropriate function of the variate which is normally distributed to a closer approximation than is χ^2 itself. They are:

(a) Fisher's result that $\sqrt{(2\chi^2)}$ is approximately normally distributed with mean $\sqrt{(2\nu-1)}$ and unit variance;

(b) Wilson and Hilferty's (1931) result that $(\chi^2/\nu)^{1/3}$ is approximately normally distributed with mean $1-2/(9\nu)$ and variance $2/(9\nu)$.

The second of these is the more accurate approximation, but involves more computation in applications.

16.6 The relative speed of approach to normality of χ^2 and $\sqrt{(2\chi^2)}$ may be compared as follows:—

For χ^2 we have, from (16.5),

$$\left.\begin{aligned}\gamma_1 &= \sqrt{\beta_1} = (8/\nu)^{\frac{1}{2}}\\ \gamma_2 &= \beta_2 - 3 = \frac{12}{\nu}.\end{aligned}\right\} \tag{16.6}$$

For the moments of χ we have

$$\begin{aligned}\mu'_r &= \frac{1}{2^{\frac{1}{2}(\nu-2)}\Gamma(\frac{1}{2}\nu)}\int_0^\infty e^{-\frac{1}{2}\chi^2}\chi^{\nu+r-1}\,d\chi\\ &= \frac{2^{\frac{1}{2}r}\Gamma\{\frac{1}{2}(\nu+r)\}}{\Gamma(\frac{1}{2}\nu)}.\end{aligned} \tag{16.7}$$

Thus

$$\mu'_1 = \sqrt{2}\,\frac{\Gamma\{\frac{1}{2}(\nu+1)\}}{\Gamma(\frac{1}{2}\nu)}.$$

Using Stirling's expansion

$$\log\Gamma(x+1) \sim \tfrac{1}{2}\log(2\pi) + (x+\tfrac{1}{2})\log x - x + \frac{1}{12x} - \frac{1}{360x^3} + o\left(\frac{1}{x^3}\right),$$

we find after substitution and reduction

$$\mu'_1 = \sqrt{\nu}\left(1 - \frac{1}{4\nu} + \frac{1}{32\nu^2} + \frac{5}{128\nu^3} + \dots\right),$$

whence $$\mu_1'^2 = \nu\left(1-\frac{1}{2\nu}+\frac{1}{8\nu^2}+\frac{1}{16\nu^3}+\ldots\right).$$

Also
$$\begin{aligned}\mu_2' &= \nu,\\ \mu_3' &= (\nu+1)\mu_1',\\ \mu_4' &= (\nu+2)\nu,\end{aligned}$$

whence we find for moments about the mean

$$\begin{aligned}\mu_2 &= \frac{1}{2}-\frac{1}{8\nu}+o\left(\frac{1}{\nu}\right),\\ \mu_3 &= \frac{1}{4\sqrt{\nu}}+o\left(\frac{1}{\sqrt{\nu}}\right),\\ \mu_4 &= \frac{3}{4}-\frac{3}{8\nu}+o\left(\frac{1}{\nu}\right).\end{aligned}$$

Hence for the moments of $\sqrt{(2\chi^2)}$ we have

$$\left.\begin{aligned}\mu_2 &= 1-\frac{1}{4\nu}+o\left(\frac{1}{\nu}\right),\\ \gamma_1 &= \frac{1}{\sqrt{(2\nu)}}+o\left(\frac{1}{\sqrt{\nu}}\right)\\ \gamma_2 &= \frac{3}{4\nu^2}+o\left(\frac{1}{\nu^2}\right).\end{aligned}\right\} \tag{16.8}$$

A comparison of (16.8) with (16.6) shows that $\sqrt{(2\chi^2)}$ tends to normality with considerably greater rapidity than χ^2. Moreover, the expression for μ_1' of χ is equal to $\sqrt{(\nu-\frac{1}{2})}$ to order $\nu^{-3/2}$ and hence $\sqrt{(2\chi^2)}$ is distributed about mean $\sqrt{(2\nu-1)}$ to that order, with variance which is unity to order ν^{-1}.

16.7 For the Wilson–Hilferty approximation, consider the distribution of χ^2 about its mean value ν. Let us find the distribution of $\left(\frac{\chi^2}{\nu}\right)^h = y$, say, h as yet being undetermined. Write $\xi = \chi^2-\nu$. Then

$$\begin{aligned}y &= \left(1+\frac{\xi}{\nu}\right)^h\\ &= 1+\frac{h\xi}{\nu}+\frac{h(h-1)}{2!}\frac{\xi^2}{\nu^2}+\ldots\end{aligned} \tag{16.9}$$

Taking mean values and using the results of (16.5), we find, after some reduction,

$$\begin{aligned}\mu_1'(y) &= 1+\frac{h(h-1)}{2!}\frac{\mu_2(\chi^2)}{\nu^2}+\ldots\\ &= 1+\frac{h(h-1)}{1}\cdot\frac{1}{\nu}+\frac{h(h-1)(h-2)(3h-1)}{6\nu^2}\\ &\quad+\frac{h^2(h-1)^2(h-2)(h-3)}{6\nu^3}+O(\nu^{-4}).\end{aligned} \tag{16.10}$$

If in (16.9) we put rh for h and find the mean value of y^r, we obtain

$$\mu'_r(y) = 1 + \frac{rh(rh-1)}{1} \cdot \frac{1}{\nu} + \dots \tag{16.11}$$

We now choose h so that the third term in (16.10) vanishes, i.e. we take $h = \frac{1}{3}$. We then find, from (16.10) and (16.11),

$$\begin{aligned}
\mu'_1(y) &= 1 - \frac{2}{9\nu} + \frac{80}{3^7\nu^3} + O(\nu^{-4}),\\
\mu'_2(y) &= 1 - \frac{2}{9\nu} + \frac{4}{3^4\nu^2} + \frac{56}{3^7\nu^3} + O(\nu^{-4}),\\
\mu'_3(y) &= 1,\\
\mu'_4(y) &= 1 + \frac{4}{9\nu} - \frac{4}{3^3\nu^2} + \frac{80}{3^7\nu^3} + O(\nu^{-4}),
\end{aligned}$$

or, transferring to the mean,

$$\left.\begin{aligned}
\mu_2(y) &= \frac{2}{9\nu} - \frac{104}{3^7\nu^3} + O(\nu^{-4}),\\
\mu_3(y) &= \frac{32}{3^6\nu^3} + O(\nu^{-4}),\\
\mu_4(y) &= \frac{4}{3^3\nu^2} - \frac{16}{3^6\nu^3} + O(\nu^{-4}).
\end{aligned}\right\} \tag{16.12}$$

We now find

$$\left.\begin{aligned}
\gamma_1 &= \frac{2^{7/2}}{3^3\nu^{3/2}} + o(\nu^{-3/2})\\
\gamma_2 &= -\frac{4}{9\nu} + o(\nu^{-1}).
\end{aligned}\right\} \tag{16.13}$$

Comparison with (16.6) and (16.8) shows that $\left(\frac{\chi^2}{\nu}\right)^{1/3}$ tends to symmetry, as measured by γ_1, more rapidly than either χ^2 or $\sqrt{(2\chi^2)}$. To order ν^{-2} the variance is, from (16.12), equal to $\frac{2}{9\nu}$ and the mean $1-\frac{2}{9\nu}$. The result may also be expressed by saying that

$$\left\{\left(\frac{\chi^2}{\nu}\right)^{1/3} + \frac{2}{9\nu} - 1\right\}\left(\frac{9\nu}{2}\right)^{1/2} \tag{16.14}$$

is distributed normally about zero mean with unit variance.

16.8 The following table, quoted from Garwood (1936), shows some comparisons of the two approximations with the exact values. In each case there have been worked out the values of $\frac{1}{2}\chi_0^2$ corresponding to $P = \int_{\frac{1}{2}\chi_0^2}^{\infty} dF(\frac{1}{2}\chi^2) = 0{\cdot}01, 0{\cdot}05, 0{\cdot}95$ and $0{\cdot}99$. The

exact values are denoted by m_T, those given by Fisher's approximation by m_F, and those by Wilson and Hilferty's approximation by m_W.

Table 16.1—Comparison of approximations to the χ^2 integral (*Garwood* (1936))

	ν	m_T	m_F	$m_T - m_F$	m_W	$m_T - m_W$
$P = 0{\cdot}99$	40	11·082	10·764	0·318	11·070	0·012
	60	18·742	18·414	0·328	18·732	0·010
	80	26·770	26·436	0·334	26·761	0·009
	100	35·032	34·694	0·338	35·025	0·007
$P = 0{\cdot}95$	40	13·255	13·116	0·139	13·254	0·001
	60	21·594	21·455	0·139	21·594	0·000
	80	30·196	30·056	0·140	30·196	0·000
	100	38·965	38·825	0·140	38·965	0·000
$P = 0{\cdot}05$	42	29·062	28·919	0·143	29·060	0·002
	62	40·691	40·548	0·143	40·689	0·002
	82	52·069	51·926	0·143	52·068	0·001
	102	63·287	63·144	0·143	63·286	0·001
$P = 0{\cdot}01$	42	33·103	32·700	0·403	33·113	−0·010
	62	45·401	45·003	0·398	45·409	−0·008
	82	57·347	56·953	0·394	57·355	−0·008
	102	69·067	68·676	0·391	69·074	−0·007

The m_W approximation is evidently very good and the m_F approximation is fair.

Slutsky (1950) tabulates the χ^2 probabilities using the approximate normality of $\{\sqrt{(2\chi^2)} - \sqrt{(2\nu)}\}$ for $\sqrt{(2/\nu)} = 0$ to 0·25, i.e. from $\nu = \infty$ down to $\nu = 32$, to 5 decimal places.

16.9 A third method of approximation may be obtained by the method of **6.25** and **6.26**, and in fact our Example 6.4 was virtually based on the χ^2 distribution. From equation (6.48) with $l_1 = l_2 = 0$, $l_3 = \kappa_3 = \sqrt{\frac{8}{\nu}}$, $l_4 = \kappa_4 = \frac{12}{\nu}$, $l_5 = \frac{48\sqrt{2}}{\nu^{3/2}}$, $l_6 = \frac{480}{\nu^2}$, we find from (6.54) or (6.56) a normal deviate whose distribution function approximates to that of χ^2.

"Student's" *t*-distribution

16.10 In Example 11.8, we have seen that, in samples from a normal population, the ratio of the difference between the sample and population means to its estimated standard error,

$$t = (\bar{x}-\mu)\sqrt{\{n(n-1)\}}/\sqrt{\Sigma(x-\bar{x})^2} = (\bar{x}-\mu)\sqrt{(n/s^2)},$$

where $s^2 = \Sigma(x-\bar{x})^2/(n-1)$, has a sampling distribution given by

$$dF = dt \Big/ \left\{\nu^{\frac{1}{2}} B(\tfrac{1}{2}, \tfrac{1}{2}\nu)\left(1+\frac{t^2}{\nu}\right)^{\frac{1}{2}(\nu+1)}\right\}, \qquad \nu > 0\,; \quad -\infty \leqslant t \leqslant \infty. \tag{16.15}$$

As before we write ν for the *degrees of freedom*, here equal to $n-1$. The origin of this term requires a few words of explanation.

It has been seen (Example 11.7) that the variance of a normal sample is distributed like the sum of $(n-1)$ squares of independent variates, and generally, that if there are k linear relations connecting the original variates, the sum of squares of the originals is distributed as the sum of $n-k$ independent normal standardized variates. Each linear relation reduces the dimension of the variation, as it were, by unity. It is thus natural to speak of the number of degrees of freedom, ν, of a function such as χ^2, meaning thereby that it is distributed as the sum of squares of ν independent standardized normal variates. The expression only has this natural meaning where normal variation is concerned.

It so happens that the quantity t depends on a quantity ν which is convenient for tabulating its distribution function and is also the number of degrees of freedom of the statistic s^2 entering into the denominator of t. ν may thus, by an extension of the term, be called the number of degrees of freedom of t.

(16.15) is a distribution of the Pearson Type VII. It is evidently symmetrical about the origin and unimodal, and extends to infinity in both directions. It is now universally known as "Student's" t-distribution, the name "Student" having been the pseudonym of its discoverer, W. S. Gosset.

It is clear, from the fact that the estimated standard error of a sample mean converges to the true standard error in large samples, that the distribution (16.15) must tend with large ν to the standardized normal distribution, for the mean of a normal sample is exactly normally distributed (Example 11.12). Compare Example 4.8.

Properties of "Student's" t-distribution

16.11 The characteristic function of the distribution has already been examined in Example 3.13, where for our present notation we replace m by $\frac{1}{2}(\nu+1)$. The moments μ_r' of the distribution exist only for $r<\nu$, and are then equal to zero by symmetry for odd-order moments, while for even moments (Example 3.3)

$$\mu_{2r} = \nu^r \frac{\Gamma(r+\frac{1}{2})\,\Gamma(\frac{1}{2}\nu-r)}{\Gamma(\frac{1}{2})\,\Gamma(\frac{1}{2}\nu)}, \qquad 2r<\nu. \tag{16.16}$$

The distribution function of t may be obtained from Tables of the Incomplete Beta-function by making the transformation

$$\xi = \left(1+\frac{t^2}{\nu}\right)^{-1}$$

in (16.15). Thus we have

$$F(\tau) = \int_{-\infty}^{\tau} dF = \tfrac{1}{2} + \int_0^{\tau} \frac{\Gamma\{\frac{1}{2}(\nu+1)\}}{\sqrt{(\nu\pi)}\,\Gamma\left(\frac{\nu}{2}\right)} \frac{dt}{\left(1+\frac{t^2}{\nu}\right)^{\frac{1}{2}(\nu+1)}}$$

whence

$$2F-1 = \frac{2}{B\left(\frac{\nu}{2},\frac{1}{2}\right)} \int_0^{\tau} \frac{1}{\left(1+\frac{t^2}{\nu}\right)^{\frac{1}{2}(\nu+1)}} \frac{dt}{\sqrt{\nu}}$$

$$= \frac{1}{B\left(\frac{\nu}{2}, \frac{1}{2}\right)} \int_{\xi(\tau)}^{1} \xi^{\frac{1}{2}\nu - 1}(1-\xi)^{-\frac{1}{2}} d\xi = 1 - I_{\xi}\left(\frac{\nu}{2}, \frac{1}{2}\right),$$

whence
$$F = 1 - \tfrac{1}{2} I_{\xi}\left(\frac{\nu}{2}, \frac{1}{2}\right). \tag{16.17}$$

The values of the argument for which I has the values 0·50, 0·25, 0·10, 0·05, 0·025, 0·01, 0·005 and $\nu = 1(1)30$, 40, 60, 120, ∞, have been tabled to five significant figures by C. M. Thompson and others (1941) and can hence be used to derive the values of t corresponding to those probability levels. These tables are reproduced in the *Biometrika Tables*.

16.12 Except for special purposes, however, the use of the B-function is unnecessary, since the distribution function of t itself and tables based thereon are available.

We have

$$-\log\left(1+\frac{t^2}{\nu}\right) = -\frac{t^2}{\nu} + \frac{t^4}{2\nu^2} - \ldots + \frac{(-t^2)^j}{j\nu^j} + \ldots$$

and hence

$$-\tfrac{1}{2}(\nu+1)\log\left(1+\frac{t^2}{\nu}\right) = -\tfrac{1}{2}t^2 + \ldots - \frac{j(-t^2)^{j+1} + (j+1)(-t^2)^j}{2j(j+1)\nu^j} + \ldots \tag{16.18}$$

Further, from the expansion for $\log\Gamma(1+x)$ we find

$$\log\left\{\frac{\Gamma\{\frac{1}{2}(\nu+1)\}}{\Gamma(\frac{1}{2}\nu)}\sqrt{\frac{2}{\nu}}\right\} = -\frac{1}{4\nu} + \frac{1}{24\nu^3} - \frac{1}{20\nu^5} + \ldots \tag{16.19}$$

Now as ν tends to infinity, t tends to the standardized normal form, as we have already seen. Writing

$$y = \frac{1}{\sqrt{(2\pi)}} \exp(-\tfrac{1}{2}t^2), \tag{16.20}$$

we use (16.18), (16.19) and (16.20) to find, for the logarithm of the ordinate of (16.15), in descending powers of ν,

$$\log y + \frac{1}{4\nu}(t^4 - 2t^2 - 1) - \frac{1}{12\nu^2}(2t^6 - 3t^4) + \frac{1}{24\nu^3}(3t^8 - 4t^6 + 1) - \frac{1}{40\nu^4}(4t^{10} - 5t^8) + \frac{1}{60\nu^5}(5t^{12} - 6t^{10} - 3) - \ldots \tag{16.21}$$

Taking the exponential of (16.21) and integrating from t to ∞, we find

$$1 - F = \int_t^{\infty} y\,dt + y\left\{\frac{1}{4\nu}t(t^2+1) + \frac{1}{96\nu^2}(3t^6 - 7t^4 - 5t^2 - 3)t + \frac{1}{384\nu^3}(t^{10} - 11t^8 + 14t^6 + 6t^4 - 3t^2 - 15)t + \frac{1}{92160\nu^4}(15t^{14} - 375t^{12} + 2225t^{10} - 2141t^8 - 939t^6 - 213t^4 + 915t^2 + 945)t + \ldots\right\}. \tag{16.22}$$

(16.22) is the expression, due to Fisher, which was used by "Student" himself in calculating the distribution function of t. For values of $\nu > 20$ the first four terms of (16.22) give F with a maximum error of 0·000,005.

16.13 The most easily available tables of the t-distribution are :

(a) The " Student "–Fisher table of the d.f. for $t = 0(0{\cdot}1)6$ and $\nu = 1(1)20$.
(b) The *Biometrika* table (reproduced in the *Biometrika Tables*) gives the d.f. for $t = 0(0{\cdot}1)8$ and $\nu = 1(1)24$; 30 ; 40 ; 60 ; 120 and ∞.
(c) Fisher and Yates' (1953) inverse table of quantiles of the distribution, i.e. the values of t corresponding to the values of the distribution function F satisfying the relation $2(1-F) = 0{\cdot}9(0{\cdot}1)0{\cdot}1$, 0·05, 0·02, 0·01, 0·001 for $\nu = 1(1)30$; 40 ; 60 ; 120 and ∞. This table is given as Appendix Table 5.
(d) Thompson's inverse table, already mentioned in **16.11** above.
(e) Baldwin's (1946) inverse table for $2(1-F) = 0{\cdot}05$, 0·01 with $\nu = 1(1)30(2)100$.
(f) Federighi (1959, *J. Amer. Statist. Ass.*, **54**, 683) gives inverse tables for 20 values of $(1-F)$ between 0·25 and 0·000,000,1 and $\nu = 1(1)30(5)60(10)100$; 200; 500 ; 1000 ; 2000 ; 10,000 and ∞.
(g) Smirnov (1961, *Tables for the distribution and density functions of t-distribution*, Pergamon, Oxford) gives the d.f. and f.f. to 6 d.p. at $t = 0(0{\cdot}01)3{\cdot}00(0{\cdot}02)4{\cdot}50(0{\cdot}05)6{\cdot}50$ for $\nu = 1(1)12$, and at $t = 0(0{\cdot}01)2{\cdot}50(0{\cdot}02)3{\cdot}50(0{\cdot}05)6{\cdot}50$ for $\nu = 13(1)25$; and also the d.f. at $t = 0(0{\cdot}01)2{\cdot}50(0{\cdot}02)3{\cdot}50(0{\cdot}05)5{\cdot}00$ for $\nu = 25(1)35$. There are auxiliary tables for larger values of t and ν. Inverse tables of quantiles of the d.f. are given for $2(1-F) = 0{\cdot}4$, 0·25, 0·10, 0·05, 0·01, 0·005, 0·0025, 0·001, 0·0005 and $\nu = 1(1)30(10)100$, 120, 150(50)500, etc.

16.14 Before leaving the t-distribution, we should mention that in many applications it is the distribution of t^2, rather than that of t, which is of interest. There is, of course, no trouble in passing from t to t^2 in (16.15). It turns out, however, that the distribution of t^2 is merely a special case of a more general distribution, Fisher's F or z, and it is to this distribution, the last of the three basic sampling distributions related to the normal, that we now turn.

Fisher's F- and z-distributions

16.15 In discussing the distribution of a ratio, we saw, in effect (Example 11.20), that in independent samples from two normal populations with equal variances the sampling distribution of the sample variance ratio, which we shall now call

$$F = \frac{\sum_{i=1}^{n_1}(x_{1i}-\bar{x}_1)^2/(n_1-1)}{\sum_{j=1}^{n_2}(x_{2j}-\bar{x}_2)^2/(n_2-1)}, \tag{16.23}$$

is given by $$dG = \frac{\nu_1^{\frac{1}{2}\nu_1}\nu_2^{\frac{1}{2}\nu_2}F^{\frac{1}{2}(\nu_1-2)}dF}{B(\frac{1}{2}\nu_1,\frac{1}{2}\nu_2)(\nu_1F+\nu_2)^{\frac{1}{2}(\nu_1+\nu_2)}}, \qquad \nu_1,\nu_2>0\ ;\ 0\leqslant F\leqslant\infty, \tag{16.24}$$

where we have written $\nu_1 = n_1-1$, $\nu_2 = n_2-1$, for the " degrees of freedom " as before.

On comparing (16.24) with (16.15), it is evident that if we put $\nu_1 = 1$, $\nu_2 = \nu$, (16.24) becomes identical with the distribution of t^2 obtainable immediately from

(16.15). Further, it will be seen that in Example 11.20 the only facts made use of in deriving the equivalent of (16.24) were that $\sum_i (x_{1i}-\bar{x}_1)^2/\sum_j (x_{2j}-\bar{x}_2)^2$ be a ratio of two independently distributed χ^2 variates with ν_1 and ν_2 degrees of freedom for numerator and denominator respectively. It follows that (16.24) holds for the distribution of any such ratio. For example, reverting to (16.23) above, suppose that the normal populations we are sampling have different variances, say σ_1^2 and σ_2^2. It then follows that $\frac{\sigma_2^2}{\sigma_1^2}F$, where F is as at (16.23), has the distribution (16.24). For, in each sample, it is the quantity

$$\sum_i (x_{pi}-\bar{x}_p)^2/\sigma_p^2, \qquad p = 1,2,$$

which is distributed like χ^2, since the standard deviation is the scale factor.

16.16 The discussion of Example 11.20 was largely concerned with the simple transformation of (16.23)

$$z = \tfrac{1}{2}\log F, \tag{16.25}$$

whose distribution is immediately obtained from (16.24) as

$$dH = \frac{2\nu_1^{\frac{1}{2}\nu_1}\nu_2^{\frac{1}{2}\nu_2}}{B(\frac{1}{2}\nu_1, \frac{1}{2}\nu_2)} \cdot \frac{\exp(\nu_1 z)\,dz}{\{\nu_1 \exp(2z)+\nu_2\}^{\frac{1}{2}(\nu_1+\nu_2)}}, \qquad \nu_1,\nu_2>0\,; \quad -\infty \leqslant z \leqslant \infty. \tag{16.26}$$

It was in this form that the distribution was first obtained by R. A. Fisher. Modern practice is almost entirely in favour of using the simpler statistic F, although z was tabulated earlier, and the original mathematical investigations of the distribution were carried out in terms of z.

Properties of the F- and z-distributions

16.17 The general shape of the F-distribution is indicated by the fact that, while it always ranges from zero to infinity, it is J-shaped if $\nu_1 \leqslant 2$, while if $\nu_1 > 2$ it is a unimodal skew distribution with modal value at

$$\hat{F} = \frac{\nu_1-2}{\nu_1}\frac{\nu_2}{\nu_2+2}. \tag{16.27}$$

It is easily verified directly from (16.24) that the mean and variance of F are

$$\left.\begin{aligned} \mu_1' &= \nu_2/(\nu_2-2), \qquad \nu_2>2, \\ \mu_2 &= 2\nu_2^2(\nu_1+\nu_2-2)/\{\nu_1(\nu_2-2)^2(\nu_2-4)\}, \qquad \nu_2>4, \end{aligned}\right\} \tag{16.28}$$

the conditions on ν_2 being required for the existence of these moments.

Comparison of (16.27) and (16.28) shows that for $\nu_1, \nu_2 > 2$ the F-distribution always has its modal value below $F = 1$ and its mean value above $F = 1$, confirming that the distribution is always " positively " skew, as is otherwise obvious.

16.18 The characteristic function of z is proportional to

$$\int_{-\infty}^{\infty} \frac{e^{(\theta+\nu_1)z}\,dz}{(\nu_1 e^{2z}+\nu_2)^{\frac{1}{2}(\nu_1+\nu_2)}}$$

where $\theta = it$, and is thus

$$\phi(t) = \left(\frac{\nu_2}{\nu_1}\right)^{\frac{1}{2}\theta} \frac{\Gamma\{\frac{1}{2}(\nu_2 - \theta)\}\,\Gamma\{\frac{1}{2}(\nu_1 + \theta)\}}{\Gamma(\frac{1}{2}\nu_1)\,\Gamma(\frac{1}{2}\nu_2)}. \tag{16.29}$$

Thus, taking logarithms and using the Stirling expansion

$$\log\Gamma(1+x) = \tfrac{1}{2}\log 2\pi + (x+\tfrac{1}{2})\log x - x + \frac{1}{12x} - \ldots,$$

we find, from (16.29), for large ν_1 and ν_2

$$\log\phi(t) = \frac{\theta}{2}\left(\frac{1}{\nu_2} - \frac{1}{\nu_1}\right) + \frac{\theta^2}{4}\left(\frac{1}{\nu_1} + \frac{1}{\nu_2}\right) - \ldots \tag{16.30}$$

For large ν_1 and ν_2, z is distributed approximately normally with moments

$$\left.\begin{aligned} \mu_1' &= \tfrac{1}{2}\left(\frac{1}{\nu_2} - \frac{1}{\nu_1}\right) \\ \mu_2 &= \tfrac{1}{2}\left(\frac{1}{\nu_1} + \frac{1}{\nu_2}\right) \end{aligned}\right\} \tag{16.31}$$

16.19 The distribution function of F (or z) may be obtained from the Tables of the Incomplete Beta-function by making the transformation

$$\xi = (1 + \nu_1 F/\nu_2)^{-1}$$

in (16.24). We may then proceed, exactly as in **16.11** above, to the relation (16.17) (where it must be made clear that here "F" on the left-hand side refers to the value of the distribution function of F, and not to the value of F itself). However, special tables of the distribution exist which make these procedures unnecessary in the usual applications.

Because there are two "degrees of freedom" entries to any table of F, a full tabulation of the distribution would require a table of triple entry, and it is therefore more convenient to use inverse tables of the quantiles of the distribution. These have been provided by:

(a) Fisher and Yates (1953) who give the values of F (and z) corresponding to the values $(1-\mathrm{P})$ of the distribution function for P = 0·20, 0·10, 0·05, 0·01 and 0·001; $\nu_1 = 1(1)6$; 8, 12, 24, ∞ and $\nu_2 = 1(1)30$; 40, 60, 120, ∞.

(b) Merrington and Thompson (1943), whose tables also appear in the *Biometrika Tables*, give the values of F corresponding to the values $(1-\mathrm{P})$ of the distribution function for P = 0·50, 0·25, 0·10, 0·05, 0·025, 0·01, 0·005 and 0·001; $\nu_1 = 1(1)10$; 12, 15, 20, 24, 30, 40, 60, 120, ∞ and $\nu_2 = 1(1)30$; 40, 60, 120, ∞.

16.20 As is obvious from the definition of F as a ratio, we may interchange its numerator and denominator and obtain a distribution of exactly the same form, but with ν_1 and ν_2 interchanged. This may be verified by making the transformation $F = 1/y$ in (16.24). In terms of the distribution function of F, say $G_{\nu_1,\nu_2}(F)$, this implies

$$G_{\nu_1,\nu_2}(F) = 1 - G_{\nu_2,\nu_1}\left(\frac{1}{F}\right). \tag{16.32}$$

Since there is a one-one correspondence between F and z, and $\log \frac{1}{F} = -\log F$, this is equivalent to

$$H_{\nu_1,\nu_2}(z) = 1 - H_{\nu_2,\nu_1}(-z). \tag{16.33}$$

The relations (16.32) and (16.33) make it unnecessary to tabulate the distributions at both extremes. In fact, if we are interested in the distribution function for fractional values of F (i.e. negative values of z) we have only to invert the ratio (i.e. change the sign of z) and use these relations to obtain the value of the distribution function which we are seeking.

16.21 Various approximations have been given for the case when ν_1 and ν_2 are not large enough to justify the use of the normal approximation of **16.18.**

(a) (Cornish and Fisher, 1937.) The method is that of **6.25** and depends on the expansion of the distribution in a Gram-Charlier series. From the successive derivatives of $\log \Gamma(1+x)$ we can find those of $\phi(t)$, and hence ascertain the cumulants of z. Writing $r_1 = \frac{1}{\nu_1}$ and $r_2 = \frac{1}{\nu_2}$, we find

$$\left.\begin{aligned}
\kappa_1 &= -\tfrac{1}{2}(r_1 - r_2) - \tfrac{1}{6}(r_1^2 - r_2^2) \\
\kappa_2 &= \tfrac{1}{2}(r_1 + r_2) + \tfrac{1}{2}(r_1^2 + r_2^2) + \tfrac{1}{3}(r_1^3 + r_2^3) \\
\kappa_3 &= -\tfrac{1}{2}(r_1^2 - r_2^2) - (r_1^3 - r_2^3) \\
\kappa_4 &= r_1^3 + r_2^3 + 3(r_1^4 + r_2^4) \\
\kappa_5 &= -3(r_1^4 - r_2^4) \\
\kappa_6 &= 12(r_1^5 + r_2^5).
\end{aligned}\right\} \tag{16.34}$$

Hence, putting $\sigma = r_1 + r_2$ and $\delta = r_1 - r_2$, we find for the l's of **6.25** ($m = 0$, variance $= \frac{1}{2}\sigma$) :—

$$l_1 = -\sqrt{\frac{2}{\sigma}}(\tfrac{1}{2}\delta + \tfrac{1}{6}\delta\sigma),$$

$$l_2 = \tfrac{1}{2}\left(\sigma + \frac{\delta^2}{\sigma}\right) + \tfrac{1}{6}(\sigma^2 + 3\delta^2),$$

and so on. After some reduction we find, for the value of z corresponding to a probability α (which in turn corresponds to a normal deviate ξ),—

$$\begin{aligned}
z = \xi\sqrt{\frac{\sigma}{2}} &- \tfrac{1}{6}\delta(\xi^2 + 2) + \sqrt{\frac{\sigma}{2}}\left\{\frac{\sigma}{24}(\xi^3 + 3\xi) + \frac{1}{72}\frac{\delta^2}{\sigma}(\xi^3 + 11\xi)\right\} \\
&- \frac{\delta\sigma}{120}(\xi^4 + 9\xi^2 + 8) + \frac{\delta^3}{3240\sigma}(3\xi^4 + 7\xi^2 - 16) + \sqrt{\frac{\sigma}{2}}\left\{\frac{\sigma^2}{1920}(\xi^5 + 20\xi^3 + 15\xi)\right. \\
&\left. + \frac{\delta^4}{2880}(\xi^5 + 44\xi^3 + 183\xi) + \frac{\delta^4}{155520\sigma^2}(9\xi^5 - 284\xi^3 - 1513\xi)\right\}.
\end{aligned} \tag{16.35}$$

(b) (Fisher, extended by Cochran (1940).) Writing n indifferently for ν_1 and ν_2, we have, from (16.35), to order $n^{-3/2}$:—

$$z = \xi\sqrt{\frac{\sigma}{2}} - \tfrac{1}{6}\delta(\xi^2 + 2) + \sqrt{\frac{\sigma}{2}}\left\{\frac{\sigma}{24}(\xi^3 + 3\xi) + \frac{1}{72}\frac{\delta^2}{\sigma}(\xi^3 + 11\xi)\right\}.$$

Put $h = 2/\sigma$. Then

$$z = \frac{\xi}{\sqrt{h}} - \tfrac{1}{6}\delta(\xi^2+2) + \frac{1}{\sqrt{h}}\left\{\frac{\xi^3+3\xi}{12h} + \frac{\xi^3+11\xi}{144}h\delta^2\right\}. \tag{16.36}$$

Now

$$\frac{\xi}{\sqrt{(h-\lambda)}} = \frac{\xi}{\sqrt{h}} + \frac{\lambda\xi}{2h\sqrt{h}} + O(n^{-2}).$$

Hence, if we put

$$z = \frac{\xi}{\sqrt{(h-\lambda)}} - \tfrac{1}{6}\delta(\xi^2+2), \tag{16.37}$$

the difference of (16.37) from (16.36) is

$$\frac{(\xi^3+11\xi)\,\delta^2\sqrt{h}}{144},$$

provided that we take $\lambda = \dfrac{\xi^2+3}{6}$.

The difference is small in virtue of the large denominator and the factor $\delta^2 = \left(\dfrac{1}{\nu_1} - \dfrac{1}{\nu_2}\right)^2$ which is small if ν_1 and ν_2 are not too different. Thus we may take z as approximately given by (16.37). The values of λ for various values of the d.f. $(1-P)$ are

$100\,P$ %	40%	30%	20%	10%	5%	1%	0·1%
λ	0·51	0·55	0·62	0·77	0·95	1·40	2·09

For the commoner values of P the form taken by (16.37) is

$$20 \text{ per cent.}: \frac{0{\cdot}8416}{\sqrt{(h-\lambda)}} - 0{\cdot}4514\delta \tag{16.38}$$

$$5 \text{ per cent.}: \frac{1{\cdot}6449}{\sqrt{(h-\lambda)}} - 0{\cdot}7843\delta \tag{16.39}$$

$$1 \text{ per cent.}: \frac{2{\cdot}3263}{\sqrt{(h-\lambda)}} - 1{\cdot}235\delta \tag{16.40}$$

$$0{\cdot}1 \text{ per cent.}: \frac{3{\cdot}0902}{\sqrt{(h-\lambda)}} - 1{\cdot}925\delta. \tag{16.41}$$

The accuracy of the approximation for $\nu_1 = 24$, $\nu_2 = 60$ may be judged from the following comparison :—

$100\,P$ %	Value of z from (16.37)	Exact value
20	0·1337	0·1338
1	0·3748	0·3746
0·1	0·4966	0·4955

(c) (Paulson, 1942.) The Wilson–Hilferty approximation to χ^2 of **16.7** indicates that $\left(\dfrac{\chi^2}{\nu}\right)^{1/3}$ is distributed normally about mean $1-\dfrac{2}{9\nu}$ with variance $\dfrac{2}{9\nu}$. The ratio $\dfrac{s_1^2}{s^2}$ itself is the ratio of two independent quantities distributed as χ^2/ν with ν_1 and ν_2

degrees of freedom. Further, in virtue of Geary's theorem (Exercise 11.11) the ratio $\frac{m_1 - m_2 p}{(\sigma_1^2 + \sigma_2^2 p^2)^{\frac{1}{2}}}$ is normally distributed in standard measure, p^2 being the ratio s_1^2/s_2^2.

We may thus regard

$$u = \frac{\left(1 - \frac{2}{9\nu_2}\right)\left(\frac{s_1}{s_2}\right)^{\frac{2}{3}} - \left(1 - \frac{2}{9\nu_1}\right)}{\left\{\frac{2}{9\nu_2}\left(\frac{s_1}{s_2}\right)^{\frac{4}{3}} + \frac{2}{9\nu_1}\right\}^{\frac{1}{2}}} \tag{16.42}$$

as approximately normally distributed in standard measure. The approximation seems remarkably good. For instance, the following shows the exact and approximate values of p^2 for $\nu_1 = 6$, $\nu_2 = 12$.

100 P %	$\left(\frac{s_1}{s_2}\right)^2 = p^2$, from (16.42)	Exact value
20	1·72	1·72
5	3·00	3·00
1	4·85	4·82
0·1	8·58	8·38

16.22 We have now completed our study of the three fundamental sampling distributions arising from the univariate normal. Before proceeding to the development of the sampling theory of the bivariate normal distribution, it will be as well here to gather together explicitly the various relations existing between these three derived distributions and the parent normal distribution. In brief, these are as follows :—

(1) The sum of squares of n standardized normal variates about population (or sample) mean has a χ^2 distribution with $\nu = n$ (or $(n-1)$) degrees of freedom.

(2) The χ^2 distribution itself is approximately of the normal form for increasing degrees of freedom.

(3) The ratio of the square of a normal variate with zero mean to an independent estimate of its variance (distributed like χ^2/ν) has " Student's " t^2-distribution, and its square root a " Student's " t-distribution.

(4) The " Student " t-distribution tends to the standardized normal form as its degrees of freedom increase.

(5) The ratio of two independent variates, each distributed like χ^2/ν, has a Fisher's F-distribution, and half its natural logarithm has the z-distribution. The t^2-distribution is the special case of the F-distribution where the numerator variate has only one degree of freedom—see (3) and (1) above.

(6) As degrees of freedom increase for both numerator and denominator, the F-distribution tends to normality.

(7) If denominator degrees of freedom, ν_2, alone tend to infinity, it is easily seen that $\nu_1 F$ tends to a χ^2 distribution with ν_1 degrees of freedom, for then we effectively return from (5) to (1) above.

Use of these relationships between the distributions often makes it possible to use one set of tables for many purposes.

The bivariate normal distribution

16.23 We have seen (Example 15.1) that the normal distribution in two variates may be written

$$dF = \frac{dx\,dy}{2\pi\sigma_1\sigma_2(1-\rho^2)^{\frac{1}{2}}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{x^2}{\sigma_1^2}-\frac{2\rho xy}{\sigma_1\sigma_2}+\frac{y^2}{\sigma_2^2}\right)\right\}, \tag{16.43}$$

where σ_1^2, σ_2^2 are the variances of the variates, μ_{11} their covariance and $\rho = \mu_{11}/(\sigma_1\sigma_2)$ is called the *correlation coefficient* between x and y in the distribution. In (16.43), we have measured each variate from its mean, a simplification which affects none of the results which we shall discuss.

By rewriting the exponent in (16.43) alternatively as

$$-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x}{\sigma_1}-\rho\frac{y}{\sigma_2}\right)^2+\frac{y^2}{\sigma_2^2}(1-\rho^2)\right\} \tag{16.44}$$

and

$$-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{y}{\sigma_2}-\rho\frac{x}{\sigma_1}\right)^2+\frac{x^2}{\sigma_1^2}(1-\rho^2)\right\}, \tag{16.45}$$

we see from (16.44) that the conditional distribution of x, given any fixed value of y, is univariate normal with mean and variance

$$\left.\begin{aligned}\mathrm{E}(x\,|\,y) &= \rho y\sigma_1/\sigma_2,\\ \mathrm{var}(x\,|\,y) &= \sigma_1^2(1-\rho^2).\end{aligned}\right\} \tag{16.46}$$

From (16.45), analogous results follow for y, given x, which are (16.46) with x, y and suffixes *1*, *2* interchanged.

It follows from (16.46) that the conditional mean of either variate, given the value of the other, is a linear function of that value, and that its conditional variance is constant, independent of that value. We may express this by saying that the *regression* of x on y (and of y on x) is linear and homoscedastic.

In later chapters we shall be discussing the theory of correlation and regression in detail. In the remaining sections of this chapter, we shall be concerned to derive formally the sampling distributions of certain statistics in samples from a bivariate normal population.

16.24 The joint probability of n sample values $(x_1, y_1) \ldots (x_n, y_n)$ from a bivariate normal population with means μ_1 and μ_2 is

$$dF = \frac{1}{(2\pi)^n\sigma_1^n\sigma_2^n(1-\rho^2)^{\frac{1}{2}n}}\exp\left[-\frac{1}{2(1-\rho^2)}\left\{\Sigma\left(\frac{x-\mu_1}{\sigma_1}\right)^2-2\rho\Sigma\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\Sigma\left(\frac{y-\mu_2}{\sigma_2}\right)^2\right\}\right]dx_1\,dy_1\ldots dx_n\,dy_n. \tag{16.47}$$

The exponent in (16.47) may be expressed solely in terms of the five parameters of the distribution $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ and the corresponding sample statistics

$$\left.\begin{aligned} \bar{x} &= \frac{1}{n}\Sigma x,\ \bar{y} = \frac{1}{n}\Sigma y, \\ s_1^2 &= \frac{1}{n}\Sigma(x-\bar{x})^2,\ s_2^2 = \frac{1}{n}\Sigma(y-\bar{y})^2, \\ r &= \frac{1}{n}\Sigma(x-\bar{x})(y-\bar{y})/(s_1 s_2). \end{aligned}\right\} \tag{16.48}$$

For example, leaving aside the factor $-\dfrac{1}{2(1-\rho^2)}$, the first term in the exponent of (16.47) is

$$\Sigma\left(\frac{x-\mu_1}{\sigma_1}\right)^2 = \Sigma\left\{\frac{(x-\bar{x})+(\bar{x}-\mu_1)}{\sigma_1}\right\}^2 = \frac{n}{\sigma^2}\{s_1^2+(\bar{x}-\mu)^2\}, \tag{16.49}$$

the cross-product term vanishing since $\Sigma(x-\bar{x}) = 0$. The other two terms in the exponent of (16.47) can similarly be resolved into forms like (16.49), so we obtain for the complete exponent

$$n\left\{\left(\frac{\bar{x}-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{\bar{x}-\mu_1}{\sigma_1}\right)\left(\frac{\bar{y}-\mu_2}{\sigma_2}\right)+\left(\frac{\bar{y}-\mu_2}{\sigma_2}\right)^2\right\}+n\left(\frac{s_1^2}{\sigma_1^2}-\frac{2\rho r s_1 s_2}{\sigma_1\sigma_2}+\frac{s_2^2}{\sigma_2^2}\right). \tag{16.50}$$

We proceed to find the joint sampling distribution of the five statistics defined at (16.48). To do this, we must transform (16.47) suitably. With the aid of (16.50) we can express the whole of the non-differential part of (16.47) in terms of the five statistics and the five parameters. We now consider the differential element.

Generalizing the geometrical approach of Chapter 11, we consider a sample space of n dimensions for x, and another such space for y, superimposed upon the first. The sample point varies in each of the two spaces, but not independently so. In fact, if P represents the point $(x_1, \ldots, x_n)$ in the x-space and Q the point $(y_1, \ldots, y_n)$ in the y-space, and if O_1, O_2 are the points $(\bar{x}, \ldots, \bar{x})$, $(\bar{y}, \ldots, \bar{y})$, then for any given r we have

$$r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{n s_1 s_2} = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\{\Sigma(x-\bar{x})^2\Sigma(y-\bar{y})^2\}^{\frac{1}{2}}}$$

and thus r is the cosine of the angle, say θ, between O_1P and O_2Q, so that if P and r are fixed, Q varies on the cone in the y-space obtained by rotating O_2Q such that the angle made with O_1P is constant.

The element in the x-space is proportional to $s_1^{n-2}\,ds_1\,d\bar{x}$, as was seen in Example 11.7. For given r, $\bar{y}$ and s_2, the point Q varies on the $(n-2)$-dimensional surface of the hypersphere of radius $s_2\sqrt{n}$, centre $\bar{y}$ and $(n-1)$ dimensions, but another dimension is lost by keeping θ constant, so that Q lies on a surface sphere of $(n-3)$ dimensions. This $(n-3)$-dimensional surface has radius $s_2\sqrt{n}\sin\theta = s_2\sqrt{n}(1-r^2)^{\frac{1}{2}}$ and "width" $s_2\sqrt{n}\,d\theta = \dfrac{s_2\sqrt{n}\,dr}{(1-r^2)^{\frac{1}{2}}}$ and thus its content is proportional to

$$\{s_2\sqrt{n}(1-r^2)^{\frac{1}{2}}\}^{n-3}\frac{s_2\sqrt{n}\,dr}{(1-r^2)^{\frac{1}{2}}},$$

that is, to $s_2^{n-2}(1-r^2)^{\frac{1}{2}(n-4)}$.

Thus the volume element may be written

$$\begin{aligned} dv &\propto s_1^{n-2}\,ds_1\,d\bar{x}\,.\,s_2^{n-2}(1-r^2)^{\frac{1}{2}(n-4)}\,ds_2\,d\bar{y}\,dr \\ &\propto s_1^{n-2}s_2^{n-2}\,ds_1\,ds_2(1-r^2)^{\frac{1}{2}(n-4)}\,dr\,d\bar{x}\,d\bar{y}. \end{aligned} \tag{16.51}$$

From (16.47), (16.50) and (16.51), the joint frequency element of the five variables is proportional to

$$\exp\left(-\frac{n}{2(1-\rho^2)}\left[\left\{\frac{(\bar{x}-\mu_1)^2}{\sigma_1^2}-2\rho\,\frac{(\bar{x}-\mu_1)(\bar{y}-\mu_2)}{\sigma_1\sigma_2}+\frac{(\bar{y}-\mu_2)^2}{\sigma_2^2}\right\}\right.\right.$$
$$\left.\left.+\left\{\frac{s_1^2}{\sigma_1^2}-2\rho r\frac{s_1 s_2}{\sigma_1\sigma_2}+\frac{s_2^2}{\sigma_2^2}\right\}\right]\right)dv. \quad (16.52)$$

This fundamental result is due to R. A. Fisher (1915).

16.25 One important property of (16.52) may be remarked. The distribution may be factorized into two parts, one containing only $\bar{x}$ and $\bar{y}$ and the other only s_1, s_2 and r, namely

$$dF \propto \exp\left[-\frac{n}{2(1-\rho^2)}\left\{\frac{(\bar{x}-\mu_1)^2}{\sigma_1^2}-2\rho\,\frac{(\bar{x}-\mu_1)(\bar{y}-\mu_2)}{\sigma_1\sigma_2}+\frac{(\bar{y}-\mu_2)^2}{\sigma_2^2}\right\}\right]d\bar{x}\,d\bar{y} \quad (16.53)$$

and

$$dF \propto \exp\left[-\frac{n}{2(1-\rho^2)}\left\{\frac{s_1^2}{\sigma_1^2}-\frac{2\rho r s_1 s_2}{\sigma_1\sigma_2}+\frac{s_2^2}{\sigma_2^2}\right\}\right]s_1^{n-2}s_2^{n-2}(1-r^2)^{\frac{1}{2}(n-4)}\,ds_1\,ds_2\,dr. \quad (16.54)$$

Thus we see that in normal samples the distribution of means is entirely independent of that of the variances and covariance. This is a characteristic property of multivariate normality—cf. **15.24.**

Before leaving (16.53), we may also note that the means are themselves distributed exactly in the bivariate normal form, with $\mathrm{E}(\bar{x}) = \mu_1$, $\mathrm{E}(\bar{y}) = \mu_2$, $\mathrm{var}\,\bar{x} = \frac{\sigma_1^2}{n}$, $\mathrm{var}\,\bar{y} = \frac{\sigma_2^2}{n}$ (all of which results are already familiar), and

$$\mathrm{cov}(\bar{x},\bar{y}) = \frac{\sigma_1\sigma_2\rho}{n}, \quad (16.55)$$

so that the correlation between $\bar{x}$ and $\bar{y}$ is ρ, the correlation in the parent population. We indicated a more general result in Exercise 13.2.

16.26 We may now use (16.54) to obtain the distribution of the sample correlation coefficient by integrating with respect to s_1 and s_2 from 0 to ∞. Let us first of all evaluate the constant to be attached to it from the consideration that $\int dF = 1$. Make the variate-transformation

$$\left.\begin{aligned} a &= \frac{s_1^2}{\sigma_1^2}\cdot\frac{n}{2(1-\rho^2)} \\ b &= \frac{rs_1 s_2}{\sigma_1\sigma_2}\cdot\frac{n}{2(1-\rho^2)} \\ c &= \frac{s_2^2}{\sigma_2^2}\cdot\frac{n}{2(1-\rho^2)}. \end{aligned}\right\} \quad (16.56)$$

We have for the reciprocal of the Jacobian of the transformation

$$\frac{\partial(a,b,c)}{\partial(s_1,r,s_2)} = \begin{vmatrix} \frac{2s_1}{\sigma_1^2}\cdot\frac{n}{2(1-\rho^2)} & 0 & 0 \\ \frac{rs_2}{\sigma_1\sigma_2}\cdot\frac{n}{2(1-\rho^2)} & \frac{s_1s_2}{\sigma_1\sigma_2}\cdot\frac{n}{2(1-\rho^2)} & \frac{rs_1}{\sigma_1\sigma_2}\cdot\frac{n}{2(1-\rho^2)} \\ 0 & 0 & \frac{2s_2}{\sigma_2^2}\cdot\frac{n}{2(1-\rho^2)} \end{vmatrix}$$

$$= \frac{s_1^2 s_2^2 n^3}{2\sigma_1^3\sigma_2^3(1-\rho^2)^3} = \frac{2acn}{\sigma_1\sigma_2(1-\rho^2)}$$

and also the relation

$$r^2 = \frac{b^2}{ac}.$$

The integral of (16.54) then becomes

$$\int \exp[-a+2\rho b-c]\cdot\left\{\frac{2a\sigma_1^2(1-\rho^2)}{n}\right\}^{\frac{1}{2}(n-2)}\left\{\frac{2c\sigma_2^2(1-\rho^2)}{n}\right\}^{\frac{1}{2}(n-2)}\left(1-\frac{b^2}{ac}\right)^{\frac{1}{2}(n-4)} \times \frac{\sigma_1\sigma_2(1-\rho^2)}{2acn}\,da\,db\,dc$$

$$= \frac{2^{n-3}\sigma^{n-1}\sigma_2^{n-1}(1-\rho^2)^{n-1}}{n^{n-1}}\int \exp[-a+2\rho b-c](ac-b^2)^{\frac{1}{2}(n-4)}\,da\,db\,dc \qquad (16.57)$$

where the limits of a and c are 0 to ∞ and those of b are $\pm\sqrt{ac}$. This integral may be evaluated in terms of the Γ-function. Putting $\xi = a-\frac{b^2}{c}$ we find

$$\int \exp(-\xi)\exp\left(+2\rho b-c-\frac{b^2}{c}\right)\xi^{\frac{1}{2}(n-4)}c^{\frac{1}{2}(n-4)}\,d\xi\,db\,dc,$$

$$0\leqslant\xi\leqslant\infty,\ -\infty\leqslant b\leqslant\infty,\ 0\leqslant c\leqslant\infty,$$

$$= \Gamma\{\tfrac{1}{2}(n-2)\}\int \exp\left[-\frac{1}{c}\{(b-\rho c)^2+(1-\rho^2)c^2\}\right]c^{\frac{1}{2}(n-4)}\,db\,dc$$

$$= \Gamma\{\tfrac{1}{2}(n-2)\}\sqrt{\pi}\int \exp\{-(1-\rho^2)c\}\,c^{\frac{1}{2}(n-3)}\,dc$$

$$= \sqrt{\pi}\,\frac{\Gamma\{\frac{1}{2}(n-2)\}\Gamma\{\frac{1}{2}(n-1)\}}{(1-\rho^2)^{\frac{1}{2}(n-1)}}$$

$$= \frac{\pi\Gamma(n-2)}{2^{n-3}(1-\rho^2)^{\frac{1}{2}(n-1)}}. \qquad (16.58)$$

Collecting up the terms from (16.57) and (16.58), we find for the joint distribution of s_1, s_2 and r

$$dF = \frac{n^{n-1}}{\pi\sigma_1^{n-1}\sigma_2^{n-1}(1-\rho^2)^{\frac{1}{2}(n-1)}\Gamma(n-2)}\exp\left[-\frac{n}{2(1-\rho^2)}\left\{\frac{s_1^2}{\sigma_1^2}-\frac{2\rho rs_1s_2}{\sigma_1\sigma_2}+\frac{s_2^2}{\sigma_2^2}\right\}\right] \times s_1^{n-2}s_2^{n-2}(1-r^2)^{\frac{1}{2}(n-4)}\,ds_1\,ds_2\,dr. \qquad (16.59)$$

16.27 Now put

$$\zeta = \frac{s_1 s_2}{\sigma_1 \sigma_2}, \qquad \beta = \log \frac{\sigma_2 s_1}{\sigma_1 s_2}, \qquad r = r.$$

We find, for the reciprocal of the Jacobian of the transformation,

$$\frac{\partial(\zeta, \beta, r)}{\partial(s_1, s_2, r)} = \begin{vmatrix} \frac{s_2}{\sigma_1 \sigma_2} & \frac{s_1}{\sigma_1 \sigma_2} & 0 \\ \frac{1}{s_1} & -\frac{1}{s_2} & 0 \\ 0 & 0 & 1 \end{vmatrix} = -\frac{2}{\sigma_1 \sigma_2}.$$

The exponent in (16.59) becomes

$$\exp\left[-\frac{n}{2(1-\rho^2)}\{\zeta e^{\beta} - 2\rho r\zeta + \zeta e^{-\beta}\}\right]$$

and after a little reduction the distribution becomes

$$dF = \frac{n^{n-1}}{2\pi(1-\rho^2)^{\frac{1}{2}(n-1)}\Gamma(n-2)} \exp\left[-\frac{n}{(1-\rho^2)}\zeta(\cosh\beta - \rho r)\right] \zeta^{n-2}\, d\zeta\, d\beta\, (1-r^2)^{\frac{1}{2}(n-4)}\, dr.$$

On integration with respect to ζ over its range $(0, \infty)$, we have

$$dF = \frac{(1-\rho^2)^{\frac{1}{2}(n-1)}}{2\pi\Gamma(n-2)} \frac{(1-r^2)^{\frac{1}{2}(n-4)}\Gamma(n-1)}{(\cosh\beta - \rho r)^{n-1}}\, d\beta\, dr. \tag{16.60}$$

(16.60) is an even function of β, so we integrate out β from 0 to ∞, dropping the factor 2 from the denominator to compensate.

Putting $-\rho r = \cos\theta$ we have, since

$$\int_0^{\infty} \frac{d\beta}{\cosh\beta + \cos\theta} = \frac{\theta}{\sin\theta},$$

$$dF = \frac{(1-\rho^2)^{\frac{1}{2}(n-1)}}{\pi\Gamma(n-2)} (1-r^2)^{\frac{1}{2}(n-4)} \frac{d^{n-2}}{d(-\cos\theta)^{n-2}} \left(\frac{\theta}{\sin\theta}\right) dr$$

$$= \frac{(1-\rho^2)^{\frac{1}{2}(n-1)}}{\pi\Gamma(n-2)} (1-r^2)^{\frac{1}{2}(n-4)} \frac{d^{n-2}}{d(r\rho)^{n-2}} \left\{\frac{\arccos(-\rho r)}{\sqrt{(1-\rho^2 r^2)}}\right\} dr. \tag{16.61}$$

These results are due to Fisher (1915).

16.28 In the particular case $\rho = 0$, (16.61) reduces, as can also be seen directly from (16.52), to

$$dF = \frac{1}{B\{\frac{1}{2}, \frac{1}{2}(n-2)\}} (1-r^2)^{\frac{1}{2}(n-4)}\, dr, \tag{16.62}$$

a form surmised by " Student " in 1908. Its distribution function may be obtained from tables of the Incomplete Beta-function, or more directly by putting

$$t = \{(n-2)r^2/(1-r^2)\}^{\frac{1}{2}}, \tag{16.63}$$

which reduces (16.62) to a " Student's " t-distribution (16.15) with $\nu = n-2$ degrees of freedom.

16.29 The general distribution (16.61) has been studied in some detail. We will here indicate only the more important features of the results.

First, as to the shape of the frequency curves. When $n = 2$ the distribution of (16.61) becomes nugatory because of the factor $\Gamma(n-2)$. This is understandable because, for samples of two, r must be either $+1$ or -1. In such a case, then, we have a discontinuous distribution which we may regard as an extreme case of a U-shaped distribution.

When $n = 3$ we find for the ordinate of the f.f., say y,

$$y = y_0\left\{\frac{1}{\sin^2\theta} - \frac{\theta\cos\theta}{\sin^3\theta}\right\}\frac{1}{(1-r^2)^{\frac{1}{2}}},$$

again a U-shaped distribution. For $n = 4$,

$$y = \frac{y_0}{\sin^3\theta}(\theta - 3\cot\theta + 3\theta\cot^2\theta).$$

If $\rho = 0$ this reduces to the rectangular form $y = \frac{1}{2}$. In other cases the curve is J-shaped.

For $n > 4$ the frequency curves are unimodal and increasingly skew as $|\rho|$ increases, as follows from the fact that r, being the cosine of an angle, satisfies $r^2 \leqslant 1$. For any ρ, as $n \to \infty$, the distribution of r tends to normality, though slowly. Some interesting photographs of models of these curves are given in the *Co-operative Study* (1917).

16.30 We may express the frequency function (16.61) as a hypergeometric series, as Hotelling (1953) pointed out. Consider the integral

$$I_{n-1} = \int_0^\infty \frac{d\beta}{(\cosh\beta - \rho r)^{n-1}}. \tag{16.64}$$

The substitution $\cosh\beta = (1-\rho r z)/(1-z)$ transforms (16.64) to

$$I_{n-1} = \frac{1}{\sqrt{2}(1-\rho r)^{n-3/2}}\int_0^1 z^{-\frac{1}{2}}(1-z)^{n-2}\{1-\tfrac{1}{2}(1+\rho r)z\}^{-\frac{1}{2}}dz,$$

and if we expand the factor in braces under the integral sign into a uniformly convergent series, and integrate the series term-by-term, we obtain

$$I_{n-1} = \frac{B(\frac{1}{2}, n-1)}{\sqrt{2}(1-\rho r)^{n-3/2}}F(\tfrac{1}{2}, \tfrac{1}{2}, n-\tfrac{1}{2}, \tfrac{1}{2}(1+\rho r)) \tag{16.65}$$

in the usual hypergeometric notation, so that (16.60), (16.64) and (16.65) yield the alternative expression for (16.61)

$$dF = \frac{(n-2)\,dr}{(n-1)\sqrt{2}\,B(\frac{1}{2}, n-\frac{1}{2})}(1-\rho^2)^{\frac{1}{4}(n-1)}(1-r^2)^{\frac{1}{4}(n-4)}(1-\rho r)^{3/2-n}F(\tfrac{1}{2}, \tfrac{1}{2}, n-\tfrac{1}{2}, \tfrac{1}{2}(1+\rho r)). \tag{16.66}$$

(16.66) converges rapidly even for small n. For large n, the first term is often enough. The error caused by stopping at any stage is considerably less than $2/(1-\rho r)$ times the last term used.

16.31 Term-by-term integration of the uniformly convergent series (16.66) gives

us an expression for the distribution function of r. For the powers of $(1+\rho r)$ can be expressed as powers of $\{2-(1-\rho r)\}$ and expanded binomially as a power series in $(1-\rho r)$. For $r>\rho$, Hotelling (1953) obtained the result

$$1-F(r) = \frac{(n-2)}{(n-1)\sqrt{2}\,B(\frac{1}{2},n-\frac{1}{2})}\left\{M_0+\frac{2M_0-M_1}{4(2n-1)} + \frac{9(4M_0-4M_1+M_2)}{32(2n-1)(2n+1)}+\ldots\right\} \tag{16.67}$$

where

$$M_k = \int_r^1 (1-\rho^2)^{\frac{1}{2}(n-1)}(1-x^2)^{\frac{1}{2}(n-4)}(1-\rho x)^{3/2+k-n}\,dx, \qquad k = 0,1,2\ldots,$$

and the asymptotic formula, for $r>\rho$,

$$1-F(r) \sim \frac{(n-2)}{(n-1)^2\sqrt{2}\,B(\frac{1}{2},n-\frac{1}{2})}\cdot\frac{(1-\rho^2)^{\frac{1}{2}(n-1)}(1-r^2)^{\frac{1}{2}(n-2)}}{(r-\rho)(1-\rho r)^{n-5/2}}\{1+O(n^{-1})\}. \tag{16.68}$$

The ordinates and distribution function of the correlation coefficient are not expressible in terms of simple mathematical functions. They have, however, been tabulated by David (1938) for values of $n = 3(1)25$, 50, 100, 200 and 400; for $\rho = 0{\cdot}0(0{\cdot}1)0{\cdot}9$; and for $r = -1{\cdot}00(0{\cdot}05)+1{\cdot}00$, with finer intervals in places.

16.32 The moments of the distribution are also expressible in terms of hypergeometric functions. Returning to (16.59), let us write

$$\alpha_1^2 = \frac{\sigma_1^2(1-\rho^2)}{n}, \quad \alpha_2^2 = \frac{\sigma_2^2(1-\rho^2)}{n}.$$

After a little rearrangement (16.59) becomes

$$dF = \frac{(1-\rho^2)^{\frac{1}{2}(n-1)}}{\pi\,\Gamma(n-2)}\exp\left\{-\frac{1}{2}\frac{s_1^2}{\alpha_1^2}-\frac{1}{2}\frac{s_2^2}{\alpha_2^2}+\rho r\frac{s_1 s_2}{\alpha_1\alpha_2}\right\} \left(\frac{s_1}{\alpha_1}\right)^{n-2}\left(\frac{s_2}{\alpha_2}\right)^{n-2}(1-r^2)^{\frac{1}{2}(n-4)}\left(\frac{ds_1}{\alpha_1}\right)\left(\frac{ds_2}{\alpha_2}\right)dr. \tag{16.69}$$

Putting $u_1 = \frac{s_1}{\alpha_1}$, $u_2 = \frac{s_2}{\alpha_2}$ and expanding the term in $\exp\left(\rho r\frac{s_1 s_2}{\alpha_1\alpha_2}\right)$ in (16.69), we have

$$dF = \frac{(1-\rho^2)^{\frac{1}{2}(n-1)}}{\pi\,\Gamma(n-2)}\exp\left(-\tfrac{1}{2}u_1^2-\tfrac{1}{2}u_2^2\right)u_1^{n-2}u_2^{n-2}(1-r^2)^{\frac{1}{2}(n-4)} \times\sum_{j=0}^{\infty}\frac{(\rho r u_1 u_2)^j}{j!}\,du_1\,du_2\,dr. \tag{16.70}$$

Integrating u_2 from 0 to ∞ in (16.70) we find for the distribution of u_1 and r

$$dF = \frac{(1-\rho^2)^{\frac{1}{2}(n-1)}}{\pi\Gamma(n-2)}\exp\left(-\tfrac{1}{2}u_1^2\right)u_1^{n-2}(1-r^2)^{\frac{1}{2}(n-4)} \times\Sigma\frac{(\rho r u_1)^j}{j!}\Gamma\left(\frac{n-1+j}{2}\right)\cdot 2^{\frac{1}{2}(n+j-3)}\,du_1\,dr \tag{16.71}$$

Multiplying the right-hand side of (16.71) by r and integrating from -1 to $+1$ we find, for the mean of r,

$$\frac{(1-\rho^2)^{\frac{1}{2}(n-1)}}{\pi\Gamma(n-2)}\exp(-\tfrac{1}{2}u_1^2)\,u_1^{n-2}\times\sum_{j=0}^{\infty}\frac{(\rho u_1)^{2j+1}}{(2j+1)!}\Gamma\left(\frac{n+2j}{2}\right)B\left(\frac{n-2}{2},\frac{2j+3}{2}\right).2^{\frac{1}{2}(n+2j-2)}\,du_1$$

and finally, integrating with respect to u_1, we obtain

$$\mu_1'(r)=\frac{(1-\rho^2)^{\frac{1}{2}(n-1)}}{\pi\Gamma(n-2)}\sum_{j=0}^{\infty}\frac{\rho^{2j+1}}{(2j+1)!}\Gamma\left(\frac{n+2j}{2}\right)B\left(\frac{n-2}{2},\frac{2j+3}{2}\right)\Gamma\left(\frac{n+2j}{2}\right)2^{n+2j-2}. \quad (16.72)$$

Substituting for the Beta-function in terms of Gamma-functions and remembering that $\Gamma(x)\,\Gamma(x+\frac{1}{2})=\frac{\pi^{\frac{1}{2}}}{2^{2x-1}}\Gamma(2x)$, we find

$$\begin{aligned}\mu_1'(r)&=\frac{\rho(1-\rho^2)^{\frac{1}{2}(n-1)}\Gamma^2(\frac{1}{2}n)}{\Gamma\{\frac{1}{2}(n-1)\}\Gamma\{\frac{1}{2}(n+1)\}}\left\{1+\frac{\frac{1}{2}n.\frac{1}{2}n}{\frac{1}{2}(n+1)}\frac{\rho^2}{1!}+\frac{\frac{1}{2}n(\frac{1}{2}n+1).\frac{1}{2}n(\frac{1}{2}n+1)}{\frac{1}{2}(n+1).\frac{1}{2}(n+3)}\frac{\rho^4}{2!}+\ldots\right\}\\&=\frac{\rho(1-\rho^2)^{\frac{1}{2}(n-1)}\Gamma^2(\frac{1}{2}n)}{\Gamma\{\frac{1}{2}(n-1)\}\Gamma\{\frac{1}{2}(n+1)\}}F(\tfrac{1}{2}n,\tfrac{1}{2}n,\tfrac{1}{2}(n+1),\rho^2)\end{aligned}$$

and since $F(\alpha,\beta,\gamma,x)=(1-x)^{\gamma-\alpha-\beta}F(\gamma-\alpha,\gamma-\beta,\gamma,x)$

$$\begin{aligned}\mu_1'(r)&=\frac{\rho\,\Gamma^2(\frac{1}{2}n)}{\Gamma\{\frac{1}{2}(n-1)\}\Gamma\{\frac{1}{2}(n+1)\}}F(\tfrac{1}{2},\tfrac{1}{2},\tfrac{1}{2}(n+1),\rho^2)\\&=\rho\left\{1-\frac{(1-\rho^2)}{2n}+O\left(\frac{1}{n^2}\right)\right\}. \quad (16.73)\end{aligned}$$

Similarly (cf. Hotelling (1953)) we find

$$\mu_2(r)=\frac{(1-\rho^2)^2}{n}\left(1+\frac{11\rho^2}{2n}\right)+O\left(\frac{1}{n^3}\right), \quad (16.74)$$

and $$\gamma_1=\frac{-6\rho}{n^{\frac{1}{2}}}+o(n^{-\frac{1}{2}}),\qquad \gamma_2=\frac{6(12\rho^2-1)}{n}+o(n^{-1}).$$

Thus the skewness increases with $|\rho|$ and declines only as $n^{-\frac{1}{2}}$.

Fisher's transformation of r

16.33 Fisher (1921*b*) found a remarkable transformation of r which tends to normality very much faster than r, with a variance almost independent of ρ. Putting

$$r=\tanh z,\quad z=\tfrac{1}{2}\log\frac{1+r}{1-r},\quad \rho=\tanh\zeta,\quad \zeta=\tfrac{1}{2}\log\frac{1+\rho}{1-\rho}, \quad (16.75)$$

we may expand the frequency function of r in powers of $z-\zeta$, $=x$ say, and inverse powers of n. Fisher gives the following expansion:

$$\begin{aligned}f=&\frac{n-2}{\sqrt{\{2\pi(n-1)\}}}e^{-\frac{1}{2}(n-1)x^2}\left\{1+\tfrac{1}{2}\rho x+\left(\frac{2+\rho}{8(n-1)}+\frac{4-\rho^2}{8}x^2+\frac{n-1}{12}x^4\right)\right.\\&+\rho x\left(\frac{4-\rho^2}{16(n-1)}+\frac{4+3\rho^2}{48}x^2+\frac{n-1}{24}x^4\right)+\left(\frac{4+12\rho^2+9\rho^4}{128(n-1)^2}+\frac{8-2\rho^2+3\rho^4}{64(n-1)}x^2\right.\\&\left.\left.+\frac{8+4\rho^2-5\rho^4}{128}x^4+\frac{28-15\rho^2}{1440}x^6(n-1)\right)+\frac{(n-1)^2}{288}x^8+\ldots\right\} \quad (16.76)\end{aligned}$$

Taking moments about $x = 0$ we find, on transferring to the mean,(*)

$$\left.\begin{aligned}
\mu_1' &= \frac{\rho}{2(n-1)}\left\{1+\frac{5+\rho^2}{4(n-1)}+\frac{11+2\rho^2+3\rho^4}{8(n-1)^2}+\dots\right\}\\
\mu_2 &= \frac{1}{n-1}\left\{1+\frac{4-\rho^2}{2(n-1)}+\frac{22-6\rho^2-3\rho^4}{6(n-1)^2}+\dots\right\}\\
\mu_3 &= \frac{\rho^3}{(n-1)^3}+\dots\\
\mu_4 &= \frac{1}{(n-1)^2}\left\{3+\frac{14-3\rho^2}{n-1}+\frac{184-48\rho^2-21\rho^4}{4(n-1)^2}+\dots\right\}
\end{aligned}\right\} \tag{16.77}$$

whence

$$\left.\begin{aligned}
\gamma_1 &= \frac{\rho^3}{(n-1)^{3/2}}+\dots\\
\gamma_2 &= \frac{2}{n-1}+\frac{4+2\rho^2-3\rho^4}{(n-1)^2}+\dots
\end{aligned}\right\} \tag{16.78}$$

Thus the variance of $z-\zeta$ is almost independent of ρ, while γ_1 declines as $n^{-\frac{3}{2}}$, and we may take $z-\zeta$ to be approximately normally distributed with mean and variance as given by (16.77). As a slightly rougher approximation we may take

$$\begin{aligned}
\mu_1'(z-\zeta) &= \frac{\rho}{2(n-1)},\\
\operatorname{var}(z-\zeta) &= \frac{1}{n-1}+\frac{4-\rho^2}{2(n-1)^2},
\end{aligned} \tag{16.79}$$

which is approximately equal for small ρ to

$$\begin{aligned}
&\frac{1}{n-1}+\frac{2}{(n-1)^2}\\
&= \frac{1}{n-3} \text{ approximately.}
\end{aligned} \tag{16.80}$$

When n is moderate we may take a still rougher approximation by assuming $z-\zeta$ to be normally distributed about zero mean with variance $\dfrac{1}{n-3}$. Some comparisons of the various approximations are given in the introduction to David's tables, and it appears that for $n>50$ the forms (16.79) and (16.80) are adequate. The approximation given by (16.77) appears to hold satisfactorily for values of n as low as 11.

Hotelling (1953) has investigated the possibility of improving upon the z transformation. He found that

$$z^* = z-(3z+r)/(4n) \tag{16.81}$$

has moments

$$\left.\begin{aligned}
\mu_1' &= \left(\zeta-\frac{3\zeta+\rho}{4n}\right)+\frac{\rho}{2(n-1)}+\frac{3\rho}{8(n-1)^2}+O(n^{-3})\\
\mu_2 &= \frac{1}{n-1}+O(n^{-3})
\end{aligned}\right\} \tag{16.82}$$

and is more closely normally distributed. A further improvement z^{**} is also discussed.

(*) Equations (16.77), as given in Fisher's original paper, contained some errors which were corrected by Gayen (1950). The main result, equation (16.80), remains [illegible]

16.34 We have now completed our discussion of the sampling distribution of r in normal samples. Before leaving the bivariate normal distribution, however, we have to consider a closely related statistic.

Just as r is defined as the ratio of sample covariance to the product of sample standard deviations, and ρ as the corresponding population quantity, we now define (see (16.48))

$$\left.\begin{aligned} b_1 &= \frac{1}{n}\Sigma(x-\bar{x})(y-\bar{y})/s_2^2, \\ b_2 &= \frac{1}{n}\Sigma(x-\bar{x})(y-\bar{y})/s_1^2, \end{aligned}\right\} \tag{16.83}$$

the population analogues being $\beta_1 = \mu_{11}/\sigma_2^2$ and $\beta_2 = \mu_{11}/\sigma_1^2$ respectively. b_1 and b_2 are called the sample *regression coefficients*, of x on y and of y on x respectively. β_1 and β_2 are the corresponding population regression coefficients.

Referring to (16.46), we see that

$$\mathrm{E}(x|y) = \beta_1 y, \qquad \mathrm{E}(y|x) = \beta_2 x.$$

We proceed to consider the sampling distribution of b_2. That for b_1 is immediately obtainable from this by permutation of suffixes.

Distribution of regression coefficients in normal samples

16.35 Turning again to equation (16.59) we have, substituting $b_2 = \dfrac{r s_2}{s_1}$, the joint frequency function of s_1, s_2 and b_2,

$$dF \propto \exp\left[-\frac{n}{2(1-\rho^2)}\left\{\frac{s_1^2}{\sigma_1^2}-\frac{2\rho s_1^2 b_2}{\sigma_1\sigma_2}+\frac{s_2^2}{\sigma_2^2}\right\}\right] s_1^{n-1} s_2^{n-3}\left(1-\frac{s_1^2 b_2^2}{s_2^2}\right)^{\frac{1}{2}(n-4)} ds_1\, ds_2\, db_2. \tag{16.84}$$

Integration of (16.84) with respect to s_2 gives, for the distribution of s_1 and b_2,

$$dF \propto \exp\left[-\frac{n s_1^2}{2(1-\rho^2)}\left\{\frac{1}{\sigma_1^2}-\frac{2\rho b_2}{\sigma_1\sigma_2}+\frac{b_2^2}{\sigma_2^2}\right\}\right] s_1^{n-1}\, ds_1\, db_2. \tag{16.85}$$

A further integration of (16.85) with respect to s_1 gives for the distribution of b_2

$$dF \qquad \frac{db_2}{\left(1-\dfrac{2\rho\sigma_1}{\sigma_2}b_2+\dfrac{\sigma_1^2}{\sigma_2^2}b_2^2\right)^{\frac{1}{2}n}}$$

or, on evaluation of the constant,

$$dF = \frac{\Gamma(\frac{1}{2}n)\,\sigma_2^{n-1}(1-\rho^2)^{\frac{1}{2}(n-1)}}{\sqrt{\pi}\,\Gamma\left(\dfrac{n-1}{2}\right)\sigma_1^{n-1}} \frac{db_2}{\left\{\dfrac{\sigma_2^2}{\sigma_1^2}(1-\rho^2)+\left(b_2-\dfrac{\rho\sigma_2}{\sigma_1}\right)^2\right\}^{\frac{1}{2}n}}. \tag{16.86}$$

The form (16.86) is a Pearson Type VII distribution, symmetrical about the point $\beta_2 = \dfrac{\rho\sigma_2}{\sigma_1}$, the population regression coefficient. It tends to normality fairly rapidly, and the use of the standard error for regressions is therefore valid for lower values of n than in the case of the correlation coefficient. For small samples, however, (16.86) is not of much use since it depends on the unknown quantities σ_1, σ_2 and ρ, i.e. the population variances and covariance.

16.36 It is possible to find statistics other than b_1 and b_2 which will provide a test of the regressions. Write

$$u = \frac{(b_2-\beta_2)s_1}{(s_2^2-b_2^2s_1^2)^{\frac{1}{2}}}. \tag{16.87}$$

We now return to the distribution of the quantities a, b, c of equation (16.56), namely, from (16.57),

$$dF \propto \exp[-a+2\rho b-c](ac-b^2)^{\frac{1}{2}(n-4)}\,da\,db\,dc. \tag{16.88}$$

We have from (16.87)

$$u = \frac{b-\rho a}{\sqrt{(ac-b^2)}},$$

and on substituting for c in (16.88) we have, after a little reduction,

$$dF \propto \frac{\exp\left[-a\left(1+\frac{\rho^2}{u^2}\right)\right]da\,du}{au^{n-1}} \cdot \exp\left[\frac{u^2+1}{u^2}\left(2\rho b-\frac{b^2}{a}\right)\right](b-\rho a)^{n-2}\,db. \tag{16.89}$$

Integration of the second part on the right of (16.89) will be found to give a factor proportional to

$$\exp\left\{\frac{a\rho^2(u^2+1)}{u^2}\right\}\cdot\left(\frac{au^2}{u^2+1}\right)^{\frac{1}{2}(n-1)} \tag{16.90}$$

and hence for the distribution of a and u we find, from (16.89) and (16.90),

$$dF \propto a^{\frac{1}{2}(n-3)}\exp\{a(\rho^2-1)\}\,da.\frac{du}{(1+u^2)^{\frac{1}{2}(n-1)}}. \tag{16.91}$$

Hence the distributions of a and u are independent, and for that of u we have

$$dF \propto \frac{du}{(1+u^2)^{\frac{1}{2}(n-1)}}. \tag{16.92}$$

This distribution does not contain any of the parent parameters. If we put

$$t = u\sqrt{(n-2)} = \frac{(b_2-\beta_2)s_1\sqrt{(n-2)}}{(s_2^2-b_2^2s_1^2)^{\frac{1}{2}}}, \tag{16.93}$$

then t is distributed in "Student's" form (16.15) with $\nu = n-2$ degrees of freedom.

16.37 In later chapters, we shall be dealing with the applications of the distributions of r and b which have been formally derived in the preceding sections, and also with the generalization of our results to multivariate normal distributions with more than two variates.

EXERCISES

16.1 Show that the moments of the variance-ratio distribution at (16.24) are

$$\mu_r' = \left(\frac{\nu_2}{\nu_1}\right)^r \frac{\Gamma(\frac{1}{2}\nu_1+r)\,\Gamma(\frac{1}{2}\nu_2-r)}{\Gamma(\frac{1}{2}\nu_1)\,\Gamma(\frac{1}{2}\nu_2)}$$

if $2r<\nu_2$, and do not exist if $2r\geqslant\nu_2$.

16.2 If $u_1, u_2, \ldots, u_n$ are independently distributed as χ^2 show, by reference to Exercise 11.15, that $\sum_{i=1}^{n} a_i u_i$, where a_i is either $+1$ or -1, is distributed either as a χ^2 variate, or in the form of a Bessel function, or as a difference of a χ^2 and a Bessel-function variate.

16.3 Show that, for the χ^2 distribution with ν degrees of freedom, the moments about the mean obey the relation

$$\mu_{j+1} = 2j(\mu_j + \nu\mu_{j-1}), \quad j \geqslant 1,$$

and hence verify (16.5).

16.4 n variates $x_1, \ldots, x_n$ are independently distributed in the rectangular form $dF = dx$, $0 \leqslant x \leqslant 1$. Show that $-2\log(x_1 x_2 \ldots x_n)$ is distributed as χ^2 with $2n$ degrees of freedom.

16.5 X_i $(i = 1, \ldots, n)$ are a given set of values with zero mean and ε_i $(i = 1, \ldots, n)$ a set of values of a normal random variable with zero mean and variance σ^2. Show that if

$$y_i = \beta_0 + \beta_1 X_i + \varepsilon_i,$$

and the constants β_0, β_1 are estimated as b_0, b_1 where

$$b_0 = \bar{y}, \quad b_1 = \Sigma(y_i - \bar{y})X_i / \Sigma X_i^2,$$

then if

$$y_i' = y_i - (b_0 + b_1 X_i),$$

we have

$$\Sigma(y_i - \bar{y})^2 = \Sigma(y_i')^2 + \frac{\{\Sigma X_i(y_i - \bar{y})\}^2}{\Sigma X_i^2}.$$

Hence show that $\Sigma(y_i')^2$ is distributed as $\sigma^2\chi^2$ with $(n-2)$ degrees of freedom independently of $b_1^2 \Sigma X_i^2$ which is distributed as $\sigma^2\chi^2$ with one degree of freedom.

If x and y are distributed normally and independently, deduce that

$$(n-2)r^2/(1-r^2)$$

is distributed as "Student's" t^2 with $(n-2)$ degrees of freedom.

16.6 Two independent machines are turning out wire in a continuous length with λ_1, λ_2 mean number of faults per unit length. Show that if t_i is the interval between the $(i-1)$th fault and the ith in some continuous period of inspection of both machines, t_i is distributed in the form $(\lambda_1 + \lambda_2)\exp\{-(\lambda_1 + \lambda_2)t_i\}$, and hence that if $T = \sum_{i=1}^{n} t_i$, $2(\lambda_1 + \lambda_2)T$ is distributed as χ^2 with $2n$ degrees of freedom.

16.7 Show that the distribution function of equation (16.2) is given, for odd integers ν, by

$$1 - F(\chi^2) = \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \left\{ \int_{\chi}^{\infty} \exp(-\tfrac{1}{2}y^2)\, dy + \exp(-\tfrac{1}{2}\chi^2) \sum_{s=1}^{\frac{1}{2}(\nu-1)} \frac{\chi^{2s-1} 2^{s-1} (s-1)!}{(2s-1)!} \right\}$$

and for even integers by

$$1 - F(\chi^2) = \exp(-\tfrac{1}{2}\chi^2) \sum_{s=0}^{\frac{1}{2}(\nu-2)} (\tfrac{1}{2}\chi^2)^s / s!.$$

Hence show that in the latter case the distribution function of χ^2 may be used to evaluate the cumulative sums of terms in a Poisson distribution.

16.8 By the method of **16.7** show that

$$\left[\left(\frac{13\chi^2 - \nu}{12\nu}\right)^{5/13} + \frac{5}{18\nu}\left(1 + \frac{7}{48\nu}\right) - 1\right] \frac{6\nu\sqrt{2}}{5}\left(1 - \frac{1}{18\nu}\right)$$

is approximately a normal variate with zero mean and unit variance.

(Haldane, 1937)

16.9 Show directly from (16.15) that "Student's" t tends to normality as ν tends to infinity.

16.10 By the substitution $q = (1+t^2/\nu)^{-1}$ show that the distribution function $F(t)$ of "Student's" t is, for even ν, given by

$$2F(t)-1 = (1-q)^{\frac{1}{2}}\left\{1+\sum_{s=1}^{\frac{1}{2}(\nu-2)}\frac{q^s}{2^{2s-1}.s.B(s,s)}\right\},$$

the expression in braces on the right-hand side being the cumulative sum of terms of the binomial expansion $(1-q)^{-\frac{1}{2}}$.

(Fisher, 1935)

16.11 Verify equations (16.27) and (16.28) for the F-distribution.

16.12 Derive (16.29) as the characteristic function of the z-distribution. By examining the lower-order terms in (16.30), verify that the distribution tends to normality for large degrees of freedom.

16.13 Using the methods of Chapter 10, show that in large samples from a bivariate normal population the variance of a sample regression coefficient is

$$\operatorname{var}\left(\frac{m_{11}}{m_{20}}\right) = \frac{\sigma_2^2(1-\rho^2)}{\sigma_1^2 \quad n}$$

approximately.

16.14 By differentiation of (16.64) with respect to (ρr) show that

$$\{1-(\rho r)^2\}I_2 = 1+\rho r I_1$$

and hence that

$$n\{1-(\rho r)^2\}I_{n+1}-(2n-1)\rho r I_n-(n-1)I_{n-1} = 0.$$

Hence show that the frequency function, f, of r satisfies the recurrence formula

$$(n-1)\{1-(\rho r)^2\}f_{n+1} = (2n-1)\rho r(1-\rho^2)^{\frac{1}{2}}(1-r^2)^{\frac{1}{2}}f_n+(n-1)^2(1-\rho^2)(1-r^2)f_{n-1}/(n-2).$$

(Co-operative Study, 1917; Hotelling, 1953)

16.15 Verify that the frequency function of r satisfies the relation

$$r\frac{\partial f_n}{\partial r}+\frac{(n-3)r^2 f_n}{1-r^2} = \rho\frac{\partial f_n}{\partial \rho}+\frac{n\rho^2 f_n}{1-\rho^2}.$$

(Hotelling, 1953)

16.16 From (16.60), show that the distribution of $v = \frac{s_1}{\sigma_1}\Big/\frac{s_2}{\sigma_2}$ and r is given by

$$dF \propto \frac{v^{n-2}(1-r^2)^{\frac{1}{2}(n-4)}}{(1-2\rho r v+v^2)^{n-1}}\,dr\,dv.$$

Integrating for r from -1 to $+1$ by putting

$$r = \frac{u(\lambda+\mu)-(1-u)(\lambda-\mu)}{u(\lambda+\mu)+(1-u)(\lambda-\mu)}, \quad \lambda = 1+v^2,\ \mu = 2\rho v,$$

show that the distribution of v is

$$dF = \frac{2(1-\rho^2)^{\frac{1}{2}(n-1)}}{B\left(\frac{n-1}{2},\frac{n-1}{2}\right)}\frac{v^{n-2}}{(1+v^2)^{n-1}}\left\{1-\frac{4\rho^2v^2}{(1+v^2)^2}\right\}^{-\frac{1}{2}n}dv.$$

(Bose, 1935; Finney, 1938)

16.17 If x and y are independent continuous variates, show that in samples of size n from their joint distribution the sample correlation coefficient r has zero mean and variance $(n-1)^{-1}$.

(Pitman, 1937)

16.18 If t is a statistic which in large samples tends to a parent parameter θ, and the variance of t is some function of θ, say $f(\theta)/n+o(n^{-1})$, show that to order n^{-1} the variate

$$z = \int^t \{f(\theta)\}^{-\frac{1}{2}}\, d\theta$$

has variance $1/n+o(n^{-1})$.

Hence, from the large sample result

$$\text{var } r = (1-\rho^2)^2/n,$$

obtain the transformation $z = \tanh^{-1} r$ as one having variance $1/n+o(n^{-1})$.

16.19 Noting that, from (16.77), the variance of z is

$$\frac{1}{n}\left(1+\frac{4-\rho^2}{2n}\right)+o(n^{-2}),$$

show that z^*, defined by

$$z^* \propto \int^r \frac{dr}{(1-r^2)\left(1+\dfrac{4-r^2}{n}\right)^{\frac{1}{2}}} = z-\frac{3z+r}{n}+o(n^{-1}),$$

has variance

$$\text{var } z^* = \frac{1}{n-1}+o(n^{-2}).$$

(Hotelling, 1953; Kendall, 1953)

16.20 $x_1, x_2, \ldots, x_n$ are a random sample from a normal population. The deviations from the sample mean, typified by $x_i-\bar{x}$, are formed and one of these is chosen at random, say $x_k-\bar{x}$. Find the distribution of $(x_k-\bar{x})/s$, where $s = \Sigma(x-\bar{x})^2/(n-1)$, and show that if x_j, x_k are chosen at random, $(x_j-x_k)/s$ is distributed as

$$\frac{t\{2(n-1)\}^{\frac{1}{2}}}{\{t^2+(n-2)\}^{\frac{1}{2}}}$$

where t is "Student's" t.

(H. A David *et al.*, 1954)

16.21 When $\rho = 0$, show that the exact distribution of Fisher's z defined at (16.75) is the variance-ratio z-distribution (16.26) with $\nu_1 = \nu_2 = n-2$. From its c.f. (16.29), use the relation for integral p

$$\frac{\Gamma''(p)}{\Gamma(p)}-\left\{\frac{\Gamma'(p)}{\Gamma(p)}\right\}^2 = \frac{\pi^2}{6}-\sum_{r=1}^{p-1}\frac{1}{r^2}$$

to show that for even $n \geqslant 4$ the exact variance of z is

$$\mu_2 = \pi^2/3-\frac{1}{2}\sum_{r=1}^{\frac{1}{2}n-2}\frac{1}{r^2},$$

which for $n = 6$ equals 0·32 against the value 0·33 obtained from (16.80).

16.22 When $\rho = 0$, show from (16.59) that the distribution of rs_i/σ_i is normal with mean zero and variance $\dfrac{1}{n}$ $(i = 1, 2)$.

APPENDIX TABLES

1 The frequency function of the normal distribution

2 The distribution function of the normal distribution

3 Quantiles of the d.f. of χ^2

4a The distribution function of χ^2 for one degree of freedom, $0 \leqslant \chi^2 \leqslant 1$

4b The distribution function of χ^2 for one degree of freedom, $1 \leqslant \chi^2 \leqslant 10$

5 Quantiles of the d.f. of t

6 5 per cent. points of z

7 5 per cent. points of F

8 1 per cent. points of z

9 1 per cent. points of F

10 Symmetric functions. Augmented symmetrics in terms of power-sums and vice versa

11 Multiple k-statistics. Tables of l-statistics in terms of augmented symmetric functions and vice versa

Appendix Table 1 – Frequency function of the normal distribution $y = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2}$ with first and second differences

x	y	$\Delta^1(-)$	Δ^2	x	y	$\Delta^1(-)$	Δ^2
0·0	0·39894	199	−392	2·5	0·01753	395	+79
0·1	0·39695	591	−374	2·6	0·01358	316	+66
0·2	0·39104	965	−347	2·7	0·01042	250	+53
0·3	0·38139	1312	−308	2·8	0·00792	197	+45
0·4	0·36827	1620	−265	2·9	0·00595	152	+36
0·5	0·35207	1885	−212	3·0	0·00443	116	+27
0·6	0·33322	2097	−159	3·1	0·00327	89	+23
0·7	0·31225	2256	−104	3·2	0·00238	66	+17
0·8	0·28969	2360	− 52	3·3	0·00172	49	+13
0·9	0·26609	2412	0	3·4	0·00123	36	+10
1·0	0·24197	2412	+ 46	3·5	0·00087	26	+ 7
1·1	0·21785	2366	+ 84	3·6	0·00061	19	+ 6
1·2	0·19419	2282	+118	3·7	0·00042	13	+ 4
1·3	0·17137	2164	+143	3·8	0·00029	9	+ 2
1·4	0·14973	2021	+161	3·9	0·00020	7	+ 3
1·5	0·12952	1860	+173	4·0	0·00013	4	—
1·6	0·11092	1687	+177	4·1	0·00009	3	—
1·7	0·09405	1510	+177	4·2	0·00006	2	—
1·8	0·07895	1333	+170	4·3	0·00004	2	—
1·9	0·06562	1163	+162	4·4	0·00002	—	—
2·0	0·05399	1001	+150	4·5	0·00002	—	—
2·1	0·04398	851	+137	4·6	0·00001	—	—
2·2	0·03547	714	+120	4·7	0·00001	—	—
2·3	0·02833	594	+108	4·8	0·00000	—	—
2·4	0·02239	486	+ 91				

Appendix Table 2 – Distribution function of the normal distribution

The table shows the area under the curve $y = (2\pi)^{-\frac{1}{2}}e^{-\frac{1}{2}x^2}$ lying to the left of specified deviates x; e.g. the area corresponding to a deviate 1·86 (= 1·5 + 0·36) is 0·9686.

Deviate	0·0 +	0·5 +	1·0 +	1·5 +	2·0 +	2·5 +	3·0 +	3·5 +
0·00	5000	6915	8413	9332	9772	$9^{2}379$	$9^{2}865$	$9^{3}77$
0·01	5040	6950	8438	9345	9778	$9^{2}396$	$9^{2}869$	$9^{3}78$
0·02	5080	6985	8461	9357	9783	$9^{2}413$	$9^{2}874$	$9^{3}78$
0·03	5120	7019	8485	9370	9788	$9^{2}430$	$9^{2}878$	$9^{3}79$
0·04	5160	7054	8508	9382	9793	$9^{2}446$	$9^{2}882$	$9^{3}80$
0·05	5199	7088	8531	9394	9798	$9^{2}461$	$9^{2}886$	$9^{3}81$
0·06	5239	7123	8554	9406	9803	$9^{2}477$	$9^{2}889$	$9^{3}81$
0·07	5279	7157	8577	9418	9808	$9^{2}492$	$9^{2}893$	$9^{3}82$
0·08	5319	7190	8599	9429	9812	$9^{2}506$	$9^{2}897$	$9^{3}83$
0·09	5359	7224	8621	9441	9817	$9^{2}520$	$9^{2}900$	$9^{3}83$
0·10	5398	7257	8643	9452	9821	$9^{2}534$	$9^{3}03$	$9^{3}84$
0·11	5438	7291	8665	9463	9826	$9^{2}547$	$9^{3}06$	$9^{3}85$
0·12	5478	7324	8686	9474	9830	$9^{2}560$	$9^{3}10$	$9^{3}85$
0·13	5517	7357	8708	9484	9834	$9^{2}573$	$9^{3}13$	$9^{3}86$
0·14	5557	7389	8729	9495	9838	$9^{2}585$	$9^{3}16$	$9^{3}86$
0·15	5596	7422	8749	9505	9842	$9^{2}598$	$9^{3}18$	$9^{3}87$
0·16	5636	7454	8770	9515	9846	$9^{2}609$	$9^{3}21$	$9^{3}87$
0·17	5675	7486	8790	9525	9850	$9^{2}621$	$9^{3}24$	$9^{3}88$
0·18	5714	7517	8810	9535	9854	$9^{2}632$	$9^{3}26$	$9^{3}88$
0·19	5753	7549	8830	9545	9857	$9^{2}643$	$9^{3}29$	$9^{3}89$
0·20	5793	7580	8849	9554	9861	$9^{2}653$	$9^{3}31$	$9^{3}89$
0·21	5832	7611	8869	9564	9864	$9^{2}664$	$9^{3}34$	$9^{3}90$
0·22	5871	7642	8888	9573	9868	$9^{2}674$	$9^{3}36$	$9^{3}90$
0·23	5910	7673	8907	9582	9871	$9^{2}683$	$9^{3}38$	$9^{4}04$
0·24	5948	7704	8925	9591	9875	$9^{2}693$	$9^{3}40$	$9^{4}08$
0·25	5987	7738	8944	9599	9878	$9^{2}702$	$9^{3}42$	$9^{4}12$
0·26	6026	7764	8962	9608	9881	$9^{2}711$	$9^{3}44$	$9^{4}15$
0·27	6064	7794	8980	9616	9884	$9^{2}720$	$9^{3}46$	$9^{4}18$
0·28	6103	7823	8997	9625	9887	$9^{2}728$	$9^{3}48$	$9^{4}22$
0·29	6141	7852	9015	9633	9890	$9^{2}736$	$9^{3}50$	$9^{4}25$
0·30	6179	7881	9032	9641	9893	$9^{2}744$	$9^{3}52$	$9^{4}28$
0·31	6217	7910	9049	9649	9896	$9^{2}752$	$9^{3}53$	$9^{4}31$
0·32	6255	7939	9066	9656	9898	$9^{2}760$	$9^{3}55$	$9^{4}33$
0·33	6293	7967	9082	9664	9901	$9^{2}767$	$9^{3}57$	$9^{4}36$
0·34	6331	7995	9099	9671	9904	$9^{2}774$	$9^{3}58$	$9^{4}39$
0·35	6368	8023	9115	9678	9906	$9^{2}781$	$9^{3}60$	$9^{4}41$
0·36	6406	8051	9131	9686	9909	$9^{2}788$	$9^{3}61$	$9^{4}43$
0·37	6443	8078	9147	9693	9911	$9^{2}795$	$9^{3}62$	$9^{4}46$
0·38	6480	8106	9162	9699	9913	$9^{2}801$	$9^{3}64$	$9^{4}48$
0·39	6517	8133	9177	9706	9916	$9^{2}807$	$9^{3}65$	$9^{4}50$
0·40	6554	8159	9192	9713	9918	$9^{2}813$	$9^{3}66$	$9^{4}52$
0·41	6591	8186	9207	9719	9920	$9^{2}819$	$9^{3}68$	$9^{4}54$
0·42	6628	8212	9222	9726	9922	$9^{2}825$	$9^{3}69$	$9^{4}56$
0·43	6664	8238	9236	9732	9925	$9^{2}831$	$9^{3}70$	$9^{4}58$
0·44	6700	8264	9251	9738	9927	$9^{2}836$	$9^{3}71$	$9^{4}59$
0·45	6736	8289	9265	9744	9929	$9^{2}841$	$9^{3}72$	$9^{4}61$
0·46	6772	8315	9279	9750	9931	$9^{2}846$	$9^{3}73$	$9^{4}63$
0·47	6808	8340	9292	9756	9932	$9^{2}851$	$9^{3}74$	$9^{4}64$
0·48	6844	8365	9306	9761	9934	$9^{2}856$	$9^{3}75$	$9^{4}66$
0·49	6879	8389	9319	9767	9936	$9^{2}861$	$9^{3}76$	$9^{4}67$

Note—Decimal points in the body of the table are omitted. Repeated 9's are indicated by powers, e.g. $9^{3}71$ stands for 0·99971.

Appendix Table 3 – Quantiles of the d.f. of χ^2

(Reproduced from Table III of Sir Ronald Fisher's *Statistical Methods for Research Workers*, Oliver and Boyd Ltd., Edinburgh, by kind permission of the author and publishers)

$P = 1 - F$	0·99	0·98	0·95	0·90	0·80	0·70	0·50	0·30	0·20	0·10	0·05	0·02	0·01
$\nu = 1$	$0{\cdot}0^3157$	$0{\cdot}0^3628$	$0{\cdot}0^2393$	0·0158	0·0642	0·148	0·455	1·074	1·642	2·706	3·841	5·412	6·635
2	0·0201	0·0404	0·103	0·211	0·446	0·713	1·386	2·408	3·219	4·605	5·991	7·824	9·210
3	0·115	0·185	0·352	0·584	1·005	1·424	2·366	3·665	4·642	6·251	7·815	9·837	11·345
4	0·297	0·429	0·711	1·064	1·649	2·195	3·357	4·878	5·989	7·779	9·488	11·668	13·277
5	0·554	0·752	1·145	1·160	2·343	3·000	4·351	6·064	7·289	9·236	11·070	13·388	15·086
6	0·872	1·134	1·635	2·204	3·070	3·828	5·348	7·231	8·558	10·645	12·592	15·033	16·812
7	1·239	1·564	2·167	2·833	3·822	4·671	6·346	8·383	9·803	12·017	14·067	16·622	18·475
8	1·646	2·032	2·733	3·490	4·594	5·527	7·344	9·524	11·030	13·362	15·507	18·168	20·090
9	2·088	2·532	3·325	4·168	5·380	6·393	8 343	10·656	12·242	14·684	16·919	19·679	21·666
10	2·358	3·059	3·940	4·865	6·179	7·267	9·342	11·781	13·442	15·987	18·307	21·161	23·209
11	3·053	3·609	4·575	5·578	6·989	8·148	10·341	12·899	14·631	17·275	19·675	22·618	24·725
12	3·571	4·178	5·226	6·304	7·807	9·034	11·340	14·011	15·821	18·549	21·026	24·054	26·217
13	4·107	4·765	5·892	7·042	8·634	9·926	12·340	15·119	16·985	19·812	22·362	25·472	27·688
14	4·660	5·368	6·571	7·790	9·467	10·821	13·339	16·222	18·151	21·064	23·685	26·873	29·141
15	5·229	5·985	7·261	8·547	10·307	11·721	14·339	17·322	19·311	22·307	24·996	28·259	30·578
16	5·812	6·614	7·962	9·312	11·152	12·624	15·338	18·418	20·465	23·542	26·296	29·633	32·000
17	6·408	7·255	8·672	10·085	12·002	13·531	16·338	19·511	21·615	24·769	27·587	30·995	33·409
18	7·015	7·906	9·390	10·865	12·857	14·440	17·338	20·601	22·760	25·989	28·869	32·346	34·805
19	7·633	8·567	10·117	11·651	13·716	15·352	18·338	21·689	23·900	27·204	30·144	33·687	36·191
20	8·260	9·237	10·851	12·443	14·578	16·266	19·337	22·775	25·038	28·412	31·410	35·020	37·566
21	8·897	9·915	11·591	13·240	15·445	17·182	20·337	23·858	26·171	29·615	32·671	36·343	38·932
22	9·542	10·600	12·338	14·041	16·314	18·101	21·337	24·939	27·301	30·813	33·924	37·659	40·289
23	10·196	11·293	13·091	14·848	17·187	19·021	22·337	26·018	28·429	32·007	35·172	38·968	41·638
24	10·856	11·992	13·848	15·659	18·062	19·943	23·337	27·096	29·553	33·196	36·415	40·270	42·980
25	11·524	12·697	14·611	16·473	18·940	20·867	24·337	28·172	30·675	34·382	37·652	41·566	44·314
26	12·198	13·409	15·379	17·292	19·820	21·792	25·336	29·246	31·795	35·563	38·885	42·856	45·642
27	12·879	14·125	16·151	18·114	20·703	22·719	26·336	30·319	32·912	36·741	40·113	44·140	46·963
28	13·565	14·847	16·928	18·939	21·588	23·647	27·336	31·391	34·027	37·916	41·337	45·419	48·278
29	14·256	15·574	17·708	19·768	22·475	24·577	28·336	32·461	35·139	39·087	42·557	46·693	49·588
30	14·953	16·306	18·493	20·599	23·364	25·508	29·336	33·530	36·250	40·256	43·773	47·962	50·892

Note—For values of ν greater than 30 the quantity $\sqrt{(2\chi^2)}$ may be taken to be distributed normally about mean $\sqrt{(2\nu - 1)}$ with unit variance.

Appendix Table 4a – Distribution function of χ^2 for one degree of freedom for values $\chi^2 = 0$ to $\chi^2 = 1$ by steps of 0·01

χ^2	P	Δ	χ^2	P	Δ
0	1·00000	7966	0·50	0·47950	436
0·01	0·92034	3280	0·51	0·47514	430
0·02	0·88754	2505	0·52	0·47084	423
0·03	0·86249	2101	0·53	0·46661	418
0·04	0·84148	1842	0·54	0·46243	411
0·05	0·82306	1656	0·55	0·45832	406
0·06	0·80650	1516	0·56	0·45426	400
0·07	0·79134	1404	0·57	0·45026	395
0·08	0·77730	1312	0·58	0·44631	389
0·09	0·76418	1235	0·59	0·44242	384
0·10	0·75183	1169	0·60	0·43858	379
0·11	0·74014	1111	0·61	0·43479	374
0·12	0·72903	1060	0·62	0·43105	369
0·13	0·71843	1015	0·63	0·42736	365
0·14	0·70828	974	0·64	0·42371	360
0·15	0·69854	938	0·65	0·42011	355
0·16	0·68916	905	0·66	0·41656	351
0·17	0·68011	874	0·67	0·41305	346
0·18	0·67137	845	0·68	0·40959	343
0·19	0·66292	820	0·69	0·40616	338
0·20	0·65472	795	0·70	0·40278	334
0·21	0·64677	773	0·71	0·39944	330
0·22	0·63904	752	0·72	0·39614	326
0·23	0·63152	731	0·73	0·39288	322
0·24	0·62421	713	0·74	0·38966	318
0·25	0·61708	696	0·75	0·38648	315
0·26	0·61012	679	0·76	0·38333	311
0·27	0·60333	663	0·77	0·38022	308
0·28	0·59670	648	0·78	0·37714	304
0·29	0·59022	634	0·79	0·37410	301
0·30	0·58388	620	0·80	0·37109	297
0·31	0·57768	607	0·81	0·36812	294
0·32	0·57161	595	0·82	0·36518	291
0·33	0·56566	583	0·83	0·36227	287
0·34	0·55983	572	0·84	0·35940	285
0·35	0·55411	560	0·85	0·35655	281
0·36	0·54851	551	0·86	0·35374	278
0·37	0·54300	540	0·87	0·35096	276
0·38	0·53760	530	0·88	0·34820	272
0·39	0·53230	521	0·89	0·34548	270
0·40	0·52709	512	0·90	0·34278	267
0·41	0·52197	503	0·91	0·34011	264
0·42	0·51694	495	0·92	0·33747	261
0·43	0·51199	487	0·93	0·33486	258
0·44	0·50712	479	0·94	0·33228	256
0·45	0·50233	471	0·95	0·32972	253
0·46	0·49762	463	0·96	0·32719	251
0·47	0·49299	457	0·97	0·32468	248
0·48	0·48842	449	0·98	0·32220	246
0·49	0·48393	443	0·99	0·31974	243
0·50	0·47950	436	1·00	0·31731	241

Appendix Table 4b – Distribution function of χ^2 for one degree of freedom for values of χ^2 from 1 to 10 by steps of 0·1

χ^2	P	Δ	χ^2	P	Δ
1·0	0·31731	2304	5·5	0·01902	106
1·1	0·29427	2095	5·6	0·01796	99
1·2	0·27332	1911	5·7	0·01697	94
1·3	0·25421	1749	5·8	0·01603	89
1·4	0·23672	1605	5·9	0·01514	83
1·5	0·22067	1477	6·0	0·01431	79
1·6	0·20590	1361	6·1	0·01352	74
1·7	0·19229	1258	6·2	0·01278	71
1·8	0·17971	1163	6·3	0·01207	66
1·9	0·16808	1078	6·4	0·01141	62
2·0	0·15730	1000	6·5	0·01079	59
2·1	0·14730	929	6·6	0·01020	56
2·2	0·13801	864	6·7	0·00964	52
2·3	0·12937	803	6·8	0·00912	50
2·4	0·12134	749	6·9	0·00862	47
2·5	0·11385	699	7·0	0·00815	44
2·6	0·10686	651	7·1	0·00771	42
2·7	0·10035	609	7·2	0·00729	39
2·8	0·09426	568	7·3	0·00690	38
2·9	0·08858	532	7·4	0·00652	35
3·0	0·08326	497	7·5	0·00617	33
3·1	0·07829	465	7·6	0·00584	32
3·2	0·07364	436	7·7	0·00552	30
3·3	0·06928	408	7·8	0·00522	28
3·4	0·06520	383	7·9	0·00494	26
3·5	0·06137	359	8·0	0·00468	25
3·6	0·05778	337	8·1	0·00443	24
3·7	0·05441	316	8·2	0·00419	23
3·8	0·05125	296	8·3	0·00396	21
3·9	0·04829	279	8·4	0·00375	20
4·0	0·04550	262	8·5	0·00355	19
4·1	0·04288	246	8·6	0·00336	18
4·2	0·04042	231	8·7	0·00318	17
4·3	0·03811	217	8·8	0·00301	16
4·4	0·03594	205	8·9	0·00285	15
4·5	0·03389	192	9·0	0·00270	14
4·6	0·03197	181	9·1	0·00256	14
4·7	0·03016	170	9·2	0·00242	13
4·8	0·02846	160	9·3	0·00229	12
4·9	0·02686	151	9·4	0·00217	12
5·0	0·02535	142	9·5	0·00205	10
5·1	0·02393	134	9·6	0·00195	11
5·2	0·02259	126	9·7	0·00184	10
5·3	0·02133	119	9·8	0·00174	9
5·4	0·02014	112	9·9	0·00165	8
5·5	0·01902	106	10·0	0·00157	8

Appendix Table 5 – Quantiles of the d.f. of t

(Reproduced from Sir Ronald Fisher and Dr. F. Yates: *Statistical Tables for Biological, Medical and Agricultural Research*, Oliver and Boyd Ltd., Edinburgh, by kind permission of the authors and publishers)

$P=2(1-F)$	0·9	0·8	0·7	0·6	0·5	0·4	0·3	0·2	0·1	0·05	0·02	0·01	0·001
$\nu=$ 1	0·158	0·325	0·510	0·727	1·000	1·376	1·963	3·078	6·314	12·706	31·821	63·657	636·619
2	0·142	0·289	0·445	0·617	0·816	1·061	1·386	1·886	2·920	4·303	6·965	9·925	31·598
3	0·137	0·277	0·424	0·584	0·765	0·978	1·250	1·638	2·353	3·182	4·541	5·841	12·924
4	0·134	0·271	0·414	0·569	0·741	0·941	1·190	1·533	2·132	2·776	3·747	4·604	8·610
5	0·132	0·267	0·408	0·559	0·727	0·920	1·156	1·476	2·015	2·571	3·365	4·032	6·869
6	0·131	0·265	0·404	0·553	0·718	0·906	1·134	1·440	1·943	2·447	3·143	3·707	5·959
7	0·130	0·263	0·402	0·549	0·711	0·896	1·119	1·415	1·895	2·365	2·998	3·499	5·408
8	0·130	0·262	0·399	0·546	0·706	0·889	1·108	1·397	1·860	2·306	2·896	3·355	5·041
9	0·129	0·261	0·398	0·543	0·703	0·883	1·100	1·383	1·833	2·262	2·821	3·250	4·781
10	0·129	0·260	0·397	0·542	0·700	0·879	1·093	1·372	1·812	2·228	2·764	3·169	4·587
11	0·129	0·260	0·396	0·540	0·697	0·876	1·088	1·363	1·796	2·201	2·718	3·106	4·437
12	0·128	0·259	0·395	0·539	0·695	0·873	1·083	1·356	1·782	2·179	2·681	3·055	4·318
13	0·128	0·259	0·394	0·538	0·694	0·870	1·079	1·350	1·771	2·160	2·650	3·012	4,221
14	0·128	0·258	0·393	0·537	0·692	0·868	1·076	1·345	1·761	2·145	2·624	2·977	4·140
15	0·128	0·258	0·393	0·536	0·691	0·866	1·074	1·341	1·753	2·131	2·602	2·947	4·073
16	0·128	0·258	0·392	0·535	0·690	0·865	1·071	1·337	1·746	2·120	2·583	2·921	4·015
17	0·128	0·257	0·392	0·534	0·689	0·863	1·069	1·333	1·740	2·110	2·567	2·898	3·965
18	0·127	0·257	0·392	0·534	0·688	0·862	1·067	1·330	1·734	2·101	2·552	2·878	3·922
19	0·127	0·257	0·391	0·533	0·688	0·861	1·066	1·328	1·729	2·093	2·539	2·861	3·883
20	0·127	0·257	0·391	0·533	0·687	0·860	1·064	1·325	1·725	2·086	2·528	2·845	3·850
21	0·127	0·257	0·391	0·532	0·686	0·859	1·063	1·323	1·721	2·080	2·518	2·831	3·819
22	0·127	0·256	0·390	0·532	0·686	0·858	1·061	1·321	1·717	2·074	2·508	2·819	3·792
23	0·127	0·256	0·390	0·532	0·685	0·858	1·060	1·319	1·714	2·069	2·500	2·807	3·767
24	0·127	0·256	0·390	0·531	0·685	0·857	1·059	1·318	1·711	2·064	2·492	2·797	3·745
25	0·127	0·256	0·390	0·531	0·684	0·856	1·058	1·316	1·708	2·060	2·485	2·787	3·725
26	0·127	0·256	0·390	0·531	0·684	0·856	1·058	1·315	1·706	2·056	2·479	2·779	3·707
27	0·127	0·256	0·389	0·531	0·684	0·855	1·057	1·314	1·703	2·052	2·473	2·771	3·690
28	0·127	0·256	0·389	0·530	0·683	0·855	1·056	1·313	1·701	2·048	2·467	2·763	3·674
29	0·127	0·256	0·389	0·530	0·683	0·854	1·055	1·311	1·699	2·045	2·462	2·756	3·659
30	0·127	0·256	0·389	0·530	0·683	0·854	1·055	1·310	1·697	2·042	2·457	2·750	3·646
40	0·126	0·255	0·388	0·529	0·681	0·851	1·050	1·303	1·684	2·021	2·423	2·704	3·551
60	0·126	0·254	0·387	0·527	0·679	0·848	1·046	1·296	1·671	2·000	2·390	2·660	3·460
120	0·126	0·254	0·386	0·526	0·677	0·845	1·041	1·289	1·658	1·980	2·358	2·617	3·373
∞	0·126	0·253	0·385	0·524	0·674	0·842	1·036	1·282	1·645	1·960	2·326	2·576	3·291

Appendix Table 6 – 5 per cent. points of the distribution of z (values at which the d.f. = 0·95)

(Reprinted from Table VI of Sir Ronald Fisher's *Statistical Methods for Research Workers*, Oliver and Boyd Ltd., Edinburgh, by kind permission of the author and the publishers)

Values of ν_2	Values of ν_1: 1	2	3	4	5	6	8	12	24	∞
1	2·5421	2·6479	2·6870	2·7071	2·7194	2·7276	2·7380	2·7484	2·7588	2·7693
2	1·4592	1·4722	1·4765	1·4787	1·4800	1·4808	1·4819	1·4830	1·4840	1·4851
3	1·1577	1·1284	1·1137	1·1051	1·0994	1·0953	1·0899	1·0842	1·0781	1·0716
4	1·0212	0·9690	0·9429	0·9272	0·9168	0·9093	0·8993	0·8885	0·8767	0·8639
5	0·9441	0·8777	0·8441	0·8236	0·8097	0·7997	0·7862	0·7714	0·7550	0·7368
6	0·8948	0·8188	0·7798	0·7558	0·7394	0·7274	0·7112	0·6931	0·6729	0·6499
7	0·8606	0·7777	0·7347	0·7080	0·6896	0·6761	0·6576	0·6369	0·6134	0·5862
8	0·8355	0·7475	0·7014	0·6725	0·6525	0·6378	0·6175	0·5945	0·5682	0·5371
9	0·8163	0·7242	0·6757	0·6450	0·6238	0·6080	0·5862	0·5613	0·5324	0·4979
10	0·8012	0·7058	0·6553	0·6232	0·6009	0·5843	0·5611	0·5346	0·5035	0·4657
11	0·7889	0·6909	0·6387	0·6055	0·5822	0·5648	0·5406	0·5126	0·4795	0·4387
12	0·7788	0·6786	0·6250	0·5907	0·5666	0·5487	0·5234	0·4941	0·4592	0·4156
13	0·7703	0·6682	0·6134	0·5783	0·5535	0·5350	0·5089	0·4785	0·4419	0·3957
14	0·7630	0·6594	0·6036	0·5677	0·5423	0·5233	0·4964	0·4649	0·4269	0·3782
15	0·7568	0·6518	0·5950	0·5585	0·5326	0·5131	0·4855	0·4532	0·4138	0·3628
16	0·7514	0·6451	0·5876	0·5505	0·5241	0·5042	0·4760	0·4428	0·4022	0·3490
17	0·7466	0·6393	0·5811	0·5434	0·5166	0·4964	0·4676	0·4337	0·3919	0·3366
18	0·7424	0·6341	0·5753	0·5371	0·5099	0·4894	0·4602	0·4255	0·3827	0·3253
19	0·7386	0·6295	0·5701	0·5315	0·5040	0·4832	0·4535	0·4182	0·3743	0·3151
20	0·7352	0·6254	0·5654	0·5265	0·4986	0·4776	0·4474	0·4116	0·3668	0·3057
21	0·7322	0·6216	0·5612	0·5219	0·4938	0·4725	0·4420	0·4055	0·3599	0·2971
22	0·7294	0·6182	0·5574	0·5178	0·4894	0·4679	0·4370	0·4001	0·3536	0·2892
23	0·7269	0·6151	0·5540	0·5140	0·4854	0·4636	0·4325	0·3950	0·3478	0·2818
24	0·7246	0·6123	0·5508	0·5106	0·4817	0·4598	0·4283	0·3904	0·3425	0·2749
25	0·7225	0·6097	0·5478	0·5074	0·4783	0·4562	0·4244	0·3862	0·3376	0·2685
26	0·7205	0·6073	0·5451	0·5045	0·4752	0·4529	0·4209	0·3823	0·3330	0·2625
27	0·7187	0·6051	0·5427	0·5017	0·4723	0·4499	0·4176	0·3786	0·3287	0·2569
28	0·7171	0·6030	0·5403	0·4992	0·4696	0·4471	0·4146	0·3752	0·3248	0·2516
29	0·7155	0·6011	0·5382	0·4969	0·4671	0·4444	0·4117	0·3720	0·3211	0·2466
30	0·7141	0·5994	0·5362	0·4947	0·4648	0·4420	0·4090	0·3691	0·3176	0·2419
60	0·6933	0·5738	0·5073	0·4632	0·4311	0·4064	0·3702	0·3255	0·2654	0·1644
∞	0·6729	0·5486	04·787	0·4319	0·3974	0·3706	0·3309	0·2804	0·2085	0

Appendix Table 7 – 5 per cent. points of the variance ratio F (values at which the d.f. = 0·95)

(Reproduced from Sir Ronald Fisher and Dr. F. Yates: *Statistical Tables for Biological, Medical and Agricultural Research*, Oliver and Boyd Ltd., Edinburgh, by kind permission of the authors and publishers)

ν_2 \ ν_1	1	2	3	4	5	6	8	12	24	∞
1	161·40	199·50	215·70	224·60	230·20	234·00	238·90	243·90	249·00	254·30
2	18·51	19·00	19·16	19·25	19·30	19·33	19·37	19·41	19·45	19·50
3	10·13	9·55	9·28	9·12	9·01	8·94	8·84	8·74	8·64	8·53
4	7·71	6·94	6·59	6·39	6·26	6·16	6·04	5·91	5·77	5·63
5	6·61	5·79	5·41	5·19	5·05	4·95	4·82	4·68	4·53	4·36
6	5·99	5·14	4·76	4·53	4·39	4·28	4·15	4·00	3·84	3·67
7	5·59	4·74	4·35	4·12	3·97	3·87	3·73	3·57	3·41	3·23
8	5·32	4·46	4·07	3·84	3·69	3·58	3·44	3·28	3·12	2·93
9	5·12	4·26	3·86	3·63	3·48	3·37	3·23	3·07	2·90	2·71
10	4·96	4·10	3·71	3·48	3·33	3·22	3·07	2·91	2·74	2·54
11	4·84	3·98	3·59	3·36	3·20	3·09	2·95	2·79	2·61	2·40
12	4·75	3·88	3·49	3·26	3·11	3·00	2·85	2·69	2·50	2·30
13	4·67	3·80	3·41	3·18	3·02	2·92	2·77	2·60	2·42	2·21
14	4·60	3·74	3·34	3·11	2·96	2·85	2·70	2·53	2·35	2·13
15	4·54	3·68	3·29	3·06	2·90	2·79	2·64	2·48	2·29	2·07
16	4·49	3·63	3·24	3·01	2·85	2·74	2·59	2·42	2·24	2·01
17	4·45	3·59	3·20	2·96	2·81	2·70	2·55	2·38	2·19	1·96
18	4·41	3·55	3·16	2·93	2·77	2·66	2·51	2·34	2·15	1·92
19	4·38	3·52	3·13	2·90	2·74	2·63	2·48	2·31	2·11	1·88
20	4·35	3·49	3·10	2·87	2·71	2·60	2·45	2·28	2·08	1·84
21	4·32	3·47	3·07	2·84	2·68	2·57	2·42	2·25	2·05	1·81
22	4·30	3·44	3·05	2·82	2·66	2·55	2·40	2·23	2·03	1·78
23	4·28	3·42	3·03	2·80	2·64	2·53	2·38	2·20	2·00	1·76
24	4·26	3·40	3·01	2·78	2·62	2·51	2·36	2·18	1·98	1·73
25	4·24	33·8	2·99	2·76	2·60	2·49	2·34	2·16	1·96	1·71
26	4·22	3·37	2·98	2·74	2·59	2·47	2·32	2·15	1·95	1·69
27	4·21	3·35	2·96	2·73	2·57	2·46	2·30	2·13	1·93	1·67
28	4·20	3·34	2·95	2·71	2·56	2·44	2·29	2·12	1·91	1·65
29	4·18	3·33	2·93	2·70	2·54	2·43	2·28	2·10	1·90	1·64
30	4·17	3·32	2·92	2·69	2·53	2·42	2·27	2·09	1·89	1·62
40	4·08	3·23	2·84	2·61	2·45	2·34	2·18	2·00	1·79	1·51
60	4·00	3·15	2·76	2·52	2·37	2·25	2·10	1·92	1·70	1·39
120	3·92	3·07	2·68	2·45	2·29	2·17	2·02	1·83	1·61	1·25
∞	3·84	2·99	2·60	2·37	2·21	2·09	1·94	1·75	1·52	1·00

Lower 5 per cent. points are found by interchange of ν_1 and ν_2, i.e. ν_1 must always correspond to the greater mean square.

Appendix Table 8 – 1 per cent. points of the distribution of z (values at which the d.f. = 0·99)

(Reprinted from Table VI of Sir Ronald Fisher's *Statistical Methods for Research Workers* Oliver and Boyd Ltd., Edinburgh, by kind permission of the author and the publishers)

Values of ν_2	Values of ν_1: 1	2	3	4	5	6	8	12	24	∞
1	4·1535	4·2585	4·2974	4·3175	4·3297	4·3379	4·3482	4·3585	4·3689	4·3794
2	2·2950	2·2976	2·2984	2·2988	2·2991	2·2992	2·2994	2·2997	2·2999	2·3001
3	1·7649	1·7140	1·6915	1·6786	1·6703	1·6645	1·6569	1·6489	1·6404	1·6314
4	1·5270	1·4452	1·4075	1·3856	1·3711	1·3609	1·3473	1·3327	1·3170	1·3000
5	1·3943	1·2929	1·2449	1·2164	1·1974	1·1838	1·1656	1·1457	1·1239	1·0997
6	1·3103	1·1955	1·1401	1·1068	1·0843	1·0680	1·0460	1·0218	0·9948	0·9643
7	1·2526	1·1281	1·0672	1·0300	1·0048	0·9864	0·9614	0·9335	0·9020	0·8658
8	1·2106	1·0787	1·0135	0·9734	0·9459	0·9259	0·8983	0·8673	0·8319	0·7904
9	1·1786	1·0411	0·9724	0·9299	0·9006	0·8791	0·8494	0·8157	0·7769	0·7305
10	1·1535	1·0114	0·9399	0·8954	0·8646	0·8419	0·8104	0·7744	0·7324	0·6816
11	1·1333	0·9874	0·9136	0·8674	0·8354	0·8116	0·7785	0·7405	0·6958	0·6408
12	1·1166	0·9677	0·8919	0·8443	0·8111	0·7864	0·7520	0·7122	0·6649	0·6061
13	1·1027	0·9511	0·8737	0·8248	0·7907	0·7652	0·7295	0·6882	0·6386	0·5761
14	1·0909	0·9370	0·8581	0·8082	0·7732	0·7471	0·7103	0·6675	0·6159	0·5500
15	1·0807	0·9249	0·8448	0·7939	0·7582	0·7314	0·6937	0·6496	0·5961	0·5269
16	1·0719	0·9144	0·8331	0·7814	0·7450	0·7177	0·6791	0·6339	0·5786	0·5064
17	1·0641	0·9051	0·8229	0·7705	0·7335	0·7057	0·6663	0·6199	0·5630	0·4879
18	1·0572	0·8970	0·8138	0·7607	0·7232	0·6950	0·6549	0·6075	0·5491	0·4712
19	1·0511	0·8897	0·8057	0·7521	0·7140	0·6854	0·6447	0·5964	0·5366	0·4560
20	1·0457	0·8831	0·7985	0·7443	0·7058	0·6768	0·6355	0·5864	0·5253	0·4421
21	1·0408	0·8772	0·7920	0·7372	0·6984	0·6690	0·6272	0·5773	0·5150	0·4294
22	1·0363	0·8719	0·7860	0·7309	0·6916	0·6620	0·6196	0·5691	0·5056	0·4176
23	1·0322	0·8670	0·7806	0·7251	0·6855	0·6555	0·6127	0·5615	0·4969	0·4068
24	1·0285	0·8626	0·7757	0·7197	0·6799	0·6496	0·6064	0·5545	0·4890	0·3967
25	1·0251	0·8585	0·7712	0·7148	0·6747	0·6442	0·6006	0·5481	0·4816	0·3872
26	1·0220	0·8548	0·7670	0·7103	0·6699	0·6392	0·5952	0·5422	0·4748	0·3784
27	1·0191	0·8513	0·7631	0·7062	0·6655	0·6346	0·5902	0·5367	0·4685	0·3701
28	1·0164	0·8481	0·7595	0·7023	0·6614	0·6303	0·5856	0·5316	0·4626	0·3624
29	1·0139	0·8451	0·7562	0·6987	0·6576	0·6263	0·5813	0·5269	0·4570	0·3550
30	1·0116	0·8423	0·7531	0·6954	0·6540	0·6226	0·5773	0·5224	0·4519	0·3481
60	0·9784	0·8025	0·7086	0·6472	0·6028	0·5687	0·5189	0·4574	0·3746	0·2352
∞	0·9462	0·7636	0·6651	0·5999	0·5522	0·5152	0·4604	0·3908	0·2913	0

Appendix Table 9 – 1 per cent. points of the variance ratio F (values at which the d.f. = 0·99)

(Reproduced from Sir Ronald Fisher and Dr. F. Yates: *Statistical Tables for Biological, Medical and Agricultural Research*, Oliver and Boyd Ltd., Edinburgh, by kind permission of the authors and publishers)

ν_2 \ ν_1	1	2	3	4	5	6	8	12	24	∞
1	4052	4999	5403	5625	5764	5859	5981	6106	6234	6366
2	98·49	99·00	99·17	99·25	99·30	99·33	99·36	99·42	99·46	99·50
3	34·12	30·81	29·46	28·71	28·24	27·91	27·49	27·05	26·60	26·12
4	21·20	18·00	16·69	15·98	15·52	15·21	14·80	14·37	13·93	13·46
5	16·26	13·27	12·06	11·39	10·97	10·67	10·27	9·89	9·47	9·02
6	13·74	10·92	9·78	9·15	8·75	8·47	8·10	7·72	7·31	6·88
7	12·25	9·55	8·45	7·85	7·46	7·19	6·84	6·47	6·07	5·65
8	11·26	8·65	7·59	7·01	6·63	6·37	6·03	5·67	5·28	4·86
9	10·56	8·02	6·99	6·42	6·06	5·80	5·47	5·11	4·73	4·31
10	10·04	7·56	6·55	5·99	5·64	5·39	5·06	4·71	4·33	3·91
11	9·65	7·20	6·22	5·67	5·32	5·07	4·74	4·40	4·02	3·60
12	9·33	6·93	5·95	5·41	5·06	4·82	4·50	4·16	3·78	3·36
13	9·07	6·70	5·74	5·20	4·86	4·62	4·30	3·96	3·59	3·16
14	8·86	6·51	5·56	5·03	4·69	4·46	4·14	3·80	3·43	3·00
15	8·68	6·36	5·42	4·89	4·56	4·32	4·00	3·67	3·29	2·87
16	8·53	6·23	5·29	4·77	4·44	4·20	3·89	3·55	3·18	2·75
17	8·40	6·11	5·18	4·67	4·34	4·10	3·79	3·45	3·08	2·65
18	8·28	6·01	5·09	4·58	4·25	4·01	3·71	3·37	3·00	2·57
19	8·18	5·93	5·01	4·50	4·17	3·94	3·63	3·30	2·92	2·49
20	8·10	5·85	4·94	4·43	4·10	3·87	3·56	3·23	2·86	2·42
21	8·02	5·78	4·87	4·37	4·04	3·81	3·51	3·17	2·80	2·36
22	7·94	5·72	4·82	4·31	3·99	3·76	3·45	3·12	2·75	2·31
23	7·88	5·66	4·76	4·26	3·94	3·71	3·41	3·07	2·70	2·26
24	7·82	5·61	4·72	4·22	3·90	3·67	3·36	3·03	2·66	2·21
25	7·77	5·57	4·68	4·18	3·86	3·63	3·32	2·99	2·62	2·17
26	7·72	5·53	4·64	4·14	3·82	3·59	3·29	2·96	2·58	2·13
27	7·68	5·49	4·60	4·11	3·78	3·56	3·26	2·93	2·55	2·10
28	7·64	5·45	4·57	4·07	3·75	3·53	3·23	2·90	2·52	2·06
29	7·60	5·42	4·54	4·04	3·73	3·50	3·20	2·87	2·49	2·03
30	7·56	5·39	4·51	4·02	3·70	3·47	3·17	2·84	2·47	2·01
40	7·31	5·18	4·31	3·83	3·51	3·29	2·99	2·66	2·29	1·80
60	7·08	4·98	4·13	3·65	3·34	3·12	2·82	2·50	2·12	1·60
120	6·85	4·79	3·95	3·48	3·17	2·96	2·66	2·34	1·95	1·38
∞	6·64	4·60	3·78	3·32	3·02	2·80	2·51	2·18	1·79	1·00

Lower 1 per cent. points are found by interchange of ν_1 and ν_2, i.e. ν_1 must always correspond to the greater mean square.

Appendix Table 10 – Augmented symmetric functions in terms of power-sums and vice versa

(Reproduced from David and Kendall (1949) by kind permission of Dr. David and the editors of *Biometrika*)

weight 1

$$(1) = [1]$$

weight 2

	[2]	$[1^2]$
(2)	1	−1
$(1)^2$	1	1

weight 3

	[3]	[21]	$[1^3]$
(3)	1	−1	2
(2) (1)	1	1	−3
$(1)^3$	1	3	1

weight 4

	[4]	[31]	$[2^2]$	$[21^2]$	$[1^4]$
(4)	1	−1	−1	2	−6
(3) (1)	1	1	·	−2	8
$(2)^2$	1	·	1	−1	3
(2) $(1)^2$	1	2	1	1	−6
$(1)^4$	1	4	3	6	1

weight 5

	[5]	[41]	[32]	$[31^2]$	$[2^21]$	$[21^3]$	$[1^5]$
(5)	1	−1	−1	2	2	−6	24
(4) (1)	1	1	·	−2	−1	6	−30
(3) (2)	1	·	1	−1	−2	5	−20
(3) $(1)^2$	1	2	1	1	·	−3	20
$(2)^2$ (1)	1	1	2	·	1	−3	15
(2) $(1)^3$	1	3	4	3	3	1	−10
$(1)^5$	1	5	10	10	15	10	1

weight 6

	[6]	[51]	[42]	$[41^2]$	$[3^2]$	[321]	$[31^3]$	$[2^3]$	$[2^21^2]$	$[21^4]$	$[1^6]$
(6)	1	−1	−1	2	−1	2	−6	2	−6	24	−120
(5) (1)	1	1	·	−2	·	−1	6	·	4	−24	144
(4) (2)	1	·	1	−1	·	−1	3	−3	5	−18	90
(4) $(1)^2$	1	2	1	1	·	·	−3	·	−1	12	−90
$(3)^2$	1	·	·	·	1	−1	2	·	2	−8	40
(3) (2) (1)	1	1	1	·	1	1	−3	·	−4	20	−120
(3) $(1)^3$	1	3	3	3	1	3	1	·	·	−4	40
$(2)^3$	1	·	3	·	·	·	·	1	−1	3	−15
$(2)^2(1)^2$	1	2	3	1	2	4	·	1	1	−6	45
(2) $(1)^4$	1	4	7	6	4	16	4	3	6	1	−15
$(1)^6$	1	6	15	15	10	60	20	15	45	15	1

To express the [ ] functions in terms of ( ), read downwards up to and including the main diagonal, e.g. $[41^2] = 2\,(6) - 2\,(5)\,(1) - (4)\,(2) + (4)\,(1)^2$. To express the ( ) functions in terms of [ ], read across up to and including the main diagonal. e.g. $(4)\,(1)^2 = [6] + 2\,[51] + [42] + [41^2]$.

Appendix Table 11 – Table of multiple *k*-statistics in terms of augmented symmetric functions and vice versa

(Reproduced from Wishart (1952) by kind permission of the late author and the editors of *Biometrika*)

1st order

$l_1 = [1]/n$

2nd order

	l_{11}	l_2
$[1^2]/n^{[2]}$	1	−1
$[2]/n$	1	1

3rd order

	l_{111}	l_{21}	l_3
$[1^3]/n^{[3]}$	1	−1	2
$[21]/n^{[2]}$	1	1	−3
$[3]/n$	1	3	1

4th order

	l_{1111}	l_{211}	l_{22}	l_{31}	l_4
$[1^4]/n^{[4]}$	1	−1	1	2	−6
$[21^2]/n^{[3]}$	1	1	−2	−3	12
$[2^2]/n^{[2]}$	1	2	1	·	−3
$[31]/n^{[2]}$	1	3	·	1	−4
$[4]/n$	1	6	3	4	1

5th order

	l_{11111}	l_{2111}	l_{221}	l_{311}	l_{32}	l_{41}	l_5
$[1^5]/n^{[5]}$	1	−1	1	2	−2	−6	24
$[21^3]/n^{[4]}$	1	1	−2	−3	5	12	−60
$[2^21]/n^{[3]}$	1	2	1	·	−3	−3	30
$[31^2]/n^{[3]}$	1	3	·	1	−1	−4	20
$[32]/n^{[2]}$	1	4	3	1	1	·	−10
$[41]/n^{[2]}$	1	6	3	4	·	1	−5
$[5]/n$	1	10	15	10	10	5	1

6th order

	l_{111111}	l_{21111}	l_{2211}	l_{3111}	l_{222}	l_{321}	l_{411}	l_{33}	l_{42}	l_{51}	l_6
$[1^6]/n^{[6]}$	1	−1	1	2	−1	−2	−6	4	6	24	−120
$[21^4]/n^{[5]}$	1	1	−2	−3	3	5	12	−12	−18	−60	360
$[2^21^2]/n^{[4]}$	1	2	1	·	−3	−3	−3	9	15	30	−270
$[31^3]/n^{[4]}$	1	3	·	1	·	−1	−4	4	4	20	−120
$[2^3]/n^{[3]}$	1	3	3	·	1	·	·	·	−3	·	30
$[321]/n^{[3]}$	1	4	3	1	·	1	·	−6	−4	−10	120
$[41^2]/n^{[3]}$	1	6	3	4	·	·	1	·	−1	−5	30
$[3^2]/n^{[2]}$	1	6	9	2	·	6	·	1	·	·	−10
$[42]/n^{[2]}$	1	7	9	4	3	4	1	·	1	·	−15
$[51]/n^{[2]}$	1	10	15	10	·	10	5	·	·	1	−6
$[6]/n$	1	15	45	20	15	60	15	10	15	6	1

To express l's in terms of symmetrics read downwards as far as and including the main diagonal, e.g. $l_{222} = -[1^6]/n^{[6]} + 3\,[21^4]/n^{[5]} - 3\,[2^21^2]/n^{[4]} + [2^3]/n^{[3]}$.

To express symmetrics in terms of l's read horizontally up to and including the diagonal, e.g. $[2^21^2]/n^{[4]} = l_{111111} + 2l_{21111} + l_{2211}$.

REFERENCES

ABDEL ATY, S. H. (1954). Tables of generalised k-statistics. *Biom.*, **41**, 253.

AITKEN, A. C. (1950). Statistical independence of quadratic forms in normal variates. *Biom.*, **37**, 93.

ANIS, A. A., and LLOYD, E. H. (1953). Range of partial sums of normal variates. *Biom.*, **40**, 35.

BAKER, G. A. (1930). Distribution of the means of samples of n drawn at random from a population represented by the Gram-Charlier Series. *Ann. Math. Statist.*, **1**, 199.

BALDWIN, E. M. (1946). Percentage points of the t distribution. *Biom.*, **33**, 362.

BARTLETT, M. S. (1934). The vector representation of a sample. *Proc. Camb. Phil. Soc.*, **30**, 327.

BOSE, S. S. (1935). On the distribution of the ratio of variances of two samples drawn from a given normal bivariate correlated population. *Sankhyā*, **2**, 65.

BURR, I. W. (1942). Cumulative frequency functions. *Ann. Math. Statist.*, **13**, 215.

CARLEMAN, T. (1925). *Les fonctions quasi-analytiques.* Gauthier-Villars, Paris.

CARLTON, G. A. (1946). Estimating the parameters of a rectangular distribution. *Ann. Math. Statist.*, **17**, 355.

CHARLIER, C. V. L. (1931). *Applications* [*de la théorie des probabilités*] *à l'astronomie.* (Part of the *Traité* edited by Borel.) Gauthier-Villars, Paris.

CHU, J. T. (1955). The 'inefficiency' of the sample median for many familiar symmetric distributions. *Biom.*, **42**, 520.

COCHRAN, W. G. (1934). The distribution of quadratic forms in a normal system, with applications to the analysis of covariance. *Proc. Camb. Phil. Soc.*, **30**, 178.

COCHRAN, W. G. (1940). Note on an approximative formula for significance levels of z. *Ann. Math. Statist.*, **11**, 93.

COOK, M. B. (1951). Bivariate k-statistics and cumulants of their joint sampling distribution. *Biom.*, **38**, 179.

Co-operative Study (1917). On the distribution of the correlation coefficient in small samples. *Biom.*, **11**, 328.

CORNISH, E. A., and FISHER, R. A. (1937). Moments and cumulants in the specification of distributions. *Rev. Inst. Int. Statist.*, **5**, 307.

COX, D. R. (1948). Asymptotic distribution of the range. *Biom.*, **35**, 310.

COX, D. R. (1954). Mean and coefficient of variation of range in small samples from non-normal populations. *Biom.*, **41**, 469; correction, **42**, 277.

CRAIG, A. T. (1943). Note on the independence of certain quadratic forms. *Ann. Math. Statist.*, **14**, 195.

CRAIG, C. C. (1936). Sheppard's corrections for a discrete variable. *Ann. Math. Statist.*, **7**, 55.

CRAMÉR, H. (1926). On some classes of series used in mathematical statistics. *Skandinaviske Matematikercongres*, Copenhagen.

CRAMÉR, H. (1928). On the composition of elementary errors. *Skand. Akt.*, **11**, 13 and 141.

CRAMÉR, H. (1937). *Random Variables and Probability Distributions.* Cambridge University Press.

CRAMÉR, H., and WOLD, H. (1936). Some theorems on distribution functions. *J. Lond. Math. Soc.*, **11**, 290.

DANIELS, H. E. (1954). Saddlepoint approximations in statistics. *Ann. Math. Statist.*, **25**, 631.

DAS, S. C. (1956). The numerical evaluation of a class of integral. II. *Proc. Camb. Phil. Soc.*, **52**, 442.

DAVID, F. N. (1938). *Tables of the Correlation Coefficient.* Cambridge University Press.

DAVID, F. N. (1949*a*). Note on the application of Fisher's *k*-statistics. *Biom.*, **36**, 383.

DAVID, F. N. (1949*b*). Moments of the z and F distributions. *Biom.*, **36**, 394.

DAVID, F. N., and JOHNSON, N. L. (1951). The effect of non-normality on the power function of the F-test. *Biom.*, **38**, 43.

DAVID, F. N., and JOHNSON, N. L. (1954). Statistical treatment of censored data. Part I. Fundamental Formulae. *Biom.*, **41**, 225.

DAVID, F. N., and KENDALL, M. G. (1949, 1951, 1953, 1955). Tables of symmetric functions. *Biom.*, **36**, 431 ; **38**, 435 ; **40**, 427 ; **42**, 223.

DAVID, H. A., HARTLEY, H. O., and PEARSON, E. S. (1954). Distribution of the ratio, in a single normal sample, of range to standard deviation. *Biom.*, **41**, 482.

DYSON, F. J. (1943). A note on kurtosis. *J. Roy. Statist. Soc.*, **106**, 360.

EDGEWORTH, F. Y. (1904). The Law of Error. *Trans. Camb. Phil. Soc.*, **20**, 36 and 113 (with an Appendix not printed in the *T.C.P.S.* but issued with reprints).

ELDERTON, SIR W. P. (1938*a*). *Frequency Curves and Correlation*, 3rd edn. Cambridge University Press.

ELDERTON, SIR W. P. (1938*b*). Correzioni dei momenti quando la curva è simmetrica. *Giorn. Ist. Ital. Att.*, **16**, 145.

ELFVING, G. (1947). Asymptotical distribution of range in samples from a normal population. *Biom.*, **34**, 111.

FELLER, W. (1950). *An Introduction to Probability Theory and its Applications.* John Wiley and Sons, New York. Chapman and Hall, London.

FINNEY, D. J. (1938). The distribution of the ratio of estimates of the two variances in a sample from a normal bivariate population. *Biom.*, **30**, 190.

FISHER, R. A. (1915). Frequency-distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biom.*, **10**, 507.

FISHER, R. A. (1920). A mathematical examination of the methods of determining the accuracy of an observation by the mean error and by the mean square error. *Month. Not. R. Astr. Soc.*, **80**, 758.

FISHER, R. A. (1921*a*). On the mathematical foundations of theoretical statistics. *Phil. Trans. Roy. Soc.*, **A, 222,** 309.

FISHER, R. A. (1921*b*). On the probable error of a coefficient of correlation deduced from a small sample. *Metron*, **1**, No. 4, 1.

FISHER, R. A. (1928). Moments and product-moments of sampling distributions. *Proc. Lond. Math. Soc.*, (2), **30**, 199.

FISHER, R. A. (1935). The mathematical distributions used in the common tests of significance. *Econometrica*, **3**, 353.

FISHER, R. A., and TIPPETT, L. H. C. (1928). Limiting forms of the frequency-distribution of the largest or smallest member of a sample. *Proc. Camb. Phil. Soc.*, **24**, 180.

FISHER, R. A., and WISHART, J. (1931). The derivation of the pattern formulae of two-way partitions from those of simpler patterns. *Proc. Lond. Math. Soc.*, **33**, 195.

FISHER, R. A., and YATES, F. (1953). *Statistical Tables for use in Biological, Agricultural and Medical Research.* 4th edition. Oliver and Boyd, Edinburgh.

FISHER, R. A., CORBET, A. S., and WILLIAMS, C. B. (1943). The relation between the number of species and the number of individuals. *J. Animal Ecology*, **12**, 42.

FOSTER, F. G., and STUART, A. (1954). Distribution-free tests in time-series based on the breaking of records. *J. Roy. Statist. Soc.*, **B, 16,** 1.

FRÉCHET, M. (1927). Sur la loi de probabilité de l'écart maximum. *Annales de la Soc. Polonaise de Math.*, **6**, 92.

FRÉCHET, M. (1937). *Recherches théoriques modernes.* (Part of the *Traité* edited by Borel.) Gauthier-Villars, Paris.

FRISCH, R. (1926). Sur les semi-invariants et moments employés dans l'étude des distributions statistiques. *Oslo, Skrifter af det Norske Videnskaps Academie, II, Hist.-Filos. Klasse*, No. 3.

GARWOOD, F. (1936). Fiducial limits for the Poisson distribution. *Biom.*, **28**, 437.

GAYEN, A. K. (1950). The variance ratio in random samples of any size drawn from non-normal universes. *Biom.*, **37**, 236.

GEARY, R. C. (1930). The frequency distribution of the quotient of two normal variables. *J. Roy. Statist. Soc.*, **93**, 442.

GEARY, R. C. (1936). The distribution of ' Student's ' ratio for non-normal samples. *Supp. J. Roy. Statist. Soc.*, **3**, 178.

GINI, C. (1912). Variabilità e Mutabilità, contributo allo studio delle distribuzioni e relazioni statistiche. Studi Economico-Giuridici della R. Università di Cagliari.

GODWIN, H. J. (1949). Some low moments of order statistics. *Ann. Math. Statist.*, **20**, 279.

GODWIN, H. J. (1955). Generalisations of Tchebycheff's inequality. *J. Amer. Statist. Ass.*, **50**, 923.

GODWIN, H. J. (1945). On the distribution of the estimate of mean deviation obtained from samples from a normal population, *and* HARTLEY, H. O. (1945). Probability integral and percentage points of the mean deviation in normal samples. *Biom.*, **33**, 254, 257.

GREENWOOD, M., and YULE, G. U. (1920). An inquiry into the nature of frequency-distributions of multiple happenings, etc. *J. Roy. Statist. Soc.*, **83**, 255.

GULDBERG, S. (1935). Recurrence formulae for the semi-invariants of some discontinuous frequency distributions of n variables. *Skand. Akt.*, **18**, 270.

GUMBEL, E. J. (1934). Les valeurs extrêmes des distributions statistiques. *Ann. Inst. H. Poincaré*, **5**, 115.

GUMBEL, E. J. (1947). The distribution of the range. *Ann. Math. Statist*, **18**, 384.

GUMBEL, E. J. (1949). Probability tables for the range. *Biom.*, **36**, 142.

GUMBEL, E. J. (1954). *Statistical Theory of Extreme Values and some Practical Applications.* (National Bureau of Standards, A.M.S. 33.)

HALDANE, J. B. S. (1937). The approximate normalisation of a class of frequency-distributions. *Biom.*, **29**, 392.

HALDANE, J. B. S. (1939). Cumulants and moments of the binomial distribution. *Biom.*, **31**, 392.

HALDANE, J. B. S. (1942). Mode and median of a nearly normal distribution with given cumulants. *Biom.*, **32**, 294.

HALDANE, J. B. S. (1948). Note on the median of a multivariate distribution. *Biom.*, **35**, 414.

HAMBURGER, H. (1920, 1921). Über eine Erweiterung des Stieltjesschen Momentproblems. *Math. Ann.*, **81**, 235; **82**, 120 and 168.

HARTLEY, H. O. (1945), *see* GODWIN, H. J. (1945).

HARTLEY, H. O., and DAVID, H. A. (1954). Universal bounds for mean range and extreme observation. *Ann. Math. Statist.*, **25**, 85.

HASTINGS, C., MOSTELLER, F., TUKEY, J. W., and WINSOR, C. P. (1947). Low moments for small samples. *Ann. Math. Statist.*, **18**, 413.

HATKE, M. A. (1949). A certain cumulative probability function. *Ann. Math. Statist.*, **20**, 461.

HELMERT, F. R. (1876). Die Genauigkeit der Formel von Peters zur Berechnung des wahrscheinlichen Beobachtungsfehlers direkter Beobachtungen gleicher Genauigkeit. *Astronomische Nachrichten*, **88**, No. 2096.

HOJO, T. (1931, 1933). Distribution of the median, quartiles and interquartile distance in samples from a normal population. *Biom.*, **23**, 315; and: A further note on the relation between the median and the quartiles in small samples from a normal population. *Biom.*, **25**, 79.

HOTELLING, H. (1953). New light on the correlation coefficient and its transforms. *J. Roy. Statist. Soc.*, **B, 15**, 193.

HOTELLING, H., and SOLOMONS, L. M. (1932). The limits of a measure of skewness. *Ann. Math. Statist.*, **3**, 141.

HSU, C. T., and LAWLEY, D. N. (1939). The derivation of the fifth and sixth moments of b_2 in samples from a normal population. *Biom.*, **31**, 238.

IRWIN, J. O. (1927). On the frequency-distribution of the means of samples from a population having any law of frequency with finite moments, etc. *Biom.*, **19**, 225, and **21**, 431.

IRWIN, J. O. (1937). The frequency-distribution of the difference between two independent variates following the same Poisson distribution. *J. Roy. Statist. Soc.*, **100**, 415.

IRWIN, J. O., and KENDALL, M. G. (1944). Sampling moments of moments for a finite population. *Ann. Eug. Lond.*, **12**, 138.

JACKSON, D. (1921). Note on the median of a set of numbers. *Bull. Amer. Math. Soc.*, **27**, 160.

JAMES, G. S. (1952). Notes on a theorem of Cochran. *Proc. Camb. Phil. Soc.*, **48**, 443.

JENKINSON, A. F. (1955). The frequency-distribution of the annual maximum (or minimum) values of meteorological elements. *Quart. Jour. Roy. Met. Soc.*, **81**, 158.

JOHNSON, N. L. (1949*a*). Systems of frequency curves generated by methods of translation. *Biom.*, **36**, 149.

JOHNSON, N. L. (1949*b*). Bivariate distributions based on simple translation systems. *Biom.*, **36**, 297.

JOHNSON, N. L. (1957). A note on the mean deviation of the binomial distribution. *Biom.*, **44**, 532.

JØRGENSEN, N. R. (1916). *Undersøgelser over Frequensflader og Korrelation.* Busck, Copenhagen.

KAPLAN, E. L. (1952). Tensor notation and the sampling cumulants of *k*-statistics. *Biom.*, **39**, 319.

KAPLANSKY, I. (1945). A common error concerning kurtosis. *J. Amer. Statist. Ass.*, **40**, 259.

KENDALL, D. G., and RAO, K. S. (1950). On the generalized second limit theorem in the theory of probabilities. *Biom.*, **37**, 224.

KENDALL, M. G. (1938). The conditions under which Sheppard's corrections are valid. *J. Roy. Statist. Soc.*, **101**, 592.

KENDALL, M. G. (1940*a*, *b*, *c*). Some properties of *k*-statistics. *Ann. Eug. Lond.*, **10**, 106; Proof of Fisher's rules for ascertaining the sampling semi-invariants of *k*-statistics. *Ibid.*, **10**, 215; The derivation of multivariate sampling formulae from univariate formulae by symbolic operation. *Ibid.*, **10**, 392.

KENDALL, M. G. (1941). Relations connected with the tetrachoric series and its generalisation. *Biom.*, **32**, 196.

KENDALL, M. G. (1942). On seminvariant statistics. *Ann. Eug. Lond.*, **11**, 300.

KENDALL, M. G. (1949*a*). Reconciliation of theories of probability. *Biom.*, **36**, 101.

KENDALL, M. G. (1949*b*). Rank and product-moment correlation. *Biom.*, **36**, 177.

KENDALL, M. G. (1952). Moment-statistics in samples from a finite population. *Biom.*, **39**, 14.

KENDALL, M. G. (1953). Discussion of Hotelling (1953).

KENDALL, M. G. (1954). Two problems in sets of measurements. *Biom.*, **41**, 560.

KENDALL, M. G., and SMITH, B. B. (1938–9). Randomness and random sampling numbers. *J. Roy. Statist. Soc.*, **101**, 147, and *Supp. J. Roy. Statist. Soc.*, **6**, 51.

KULLBACK, S. (1934). An application of characteristic functions to the distribution problem of statistics. *Ann. Math. Statist.*, **5**, 264.

LANCASTER, H. O. (1954). Traces and cumulants of quadratic forms in normal variables. *J. Roy. Statist. Soc.*, **B, 16,** 247.

LIAPOUNOFF, A. (1901). Nouvelle forme du théorème sur la limite de probabilité. *Mem. Acad. Sci. St. Pét.*, (8), **12,** No. 5.

LOMNICKI, Z. A. (1952). The standard error of Gini's mean difference. *Ann. Math. Statist.*, **23,** 635.

LUKACS, E. (1942). A characterisation of the normal distribution. *Ann. Math. Statist.*, **13,** 91.

LUKACS, E. (1952). An essential property of the Fourier transforms of distribution functions. *Proc. Amer. Math. Soc.*, **3,** 508.

LUKACS, E., and SZÁSZ, O. (1952). On analytic characteristic functions. *Pacific Jour. Maths.*, **2,** 615.

McKAY, A. T. (1935). Sampling distribution of the difference between the extreme observation and the mean. *Biom.*, **27,** 466.

MALLOWS, C. L. (1956). Generalisations of Tchebycheff's Inequalities. *J. Roy. Statist. Soc.*, **B, 18,** 139.

MARCINKIEWICZ, J. (1938). Sur une propriété de la loi de Gauss. *Math. Zeitschr.*, **44,** 612.

MARTIN, E. S. (1934). On the correction for the moment coefficients of frequency-distributions when the start of the frequency is one of the characteristics to be determined. *Biom.*, **26,** 12.

MATÉRN, B. (1949). Independence of non-negative quadratic forms in normally correlated variables. *Ann. Math. Statist.*, **20,** 119.

MERRINGTON, M., and THOMPSON, C. M. (1943). Tables of the percentage points of the inverted beta (*F*) distribution. *Biom.*, **33,** 73.

MILLER, J. C. P. (1954). Tables of binomial coefficients. *Roy. Soc. Math. Tables*, vol.3.

MOLINA, E. C. (1942). *Tables of Poisson's Exponential Limit.* Van Nostrand Co., Inc., New York.

MORAN, P. A. P. (1956). The numerical evaluation of a class of integrals. *Proc. Camb. Phil. Soc.*, **52,** 230.

MORIGUTI, S. (1951). Extremal properties of extreme value distributions. *Ann. Math. Statist.*, **22,** 523.

MURTY, V. N. (1955). Distribution of the quotient of maximum values in samples from a rectangular distribution. *J. Amer. Statist. Ass.*, **50,** 1136.

NAIR, U. S. (1936). The standard error of Gini's mean difference. *Biom.*, **28,** 428.

PAIRMAN, E., and PEARSON, K. (1919). On the corrections for moment coefficients of limited-range frequency-distributions when there are finite or infinite ordinates and any slopes at the terminals of the range. *Biom.*, **12,** 231.

PAULSON, E. (1942). An approximate normalisation of the analysis of variance distribution. *Ann. Math. Statist.*, **13,** 233.

PEARSON, E. S. (1926). A further note on the distribution of range in samples taken from a normal population. *Biom.*, **18,** 173.

PEARSON, E. S. (1930). A further development of tests for normality. *Biom.*, **22**, 239.

PEARSON, E. S. (1932). The percentage limits for the distribution of range in samples from a normal population. *Biom.*, **24**, 404.

PEARSON, E. S., and HARTLEY, H. O. (1942). The probability integral of the range in samples of n observations from a normal population. *Biom.*, **32**, 301.

PEARSON, E. S., and HARTLEY, H. O. (1950). Tables of the χ^2 integral and of the cumulative Poisson distribution. *Biom.*, **37**, 313.

PEARSON, K. (1900). On a criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen in random sampling. *Phil. Mag.*, (5), **50**, 157.

PEARSON, K. (1919). On generalised Tchebycheff theorems in the mathematical theory of statistics. *Biom.*, **12**, 284.

PEARSON, K. (1924*a*). On the moments of the hypergeometrical series. *Biom.*, **16**, 157.

PEARSON, K. (1924*b*). On a certain double hypergeometrical series and its representation by continuous frequency surfaces. *Biom.*, **16**, 172.

PEARSON, K. (1931). Appendix to a paper by Professor Tokishige Hojo. On the standard error of the median to a third approximation, etc. *Biom.*, **23**, 361.

PEARSON, K., STOUFFER, S. A., and DAVID, F. N. (1932). Further applications in statistics of the $T_m(x)$ Bessel function. *Biom.*, **24**, 293.

PITMAN, E. J. G. (1937). Significance tests which may be applied to samples from any population. II. The correlation coefficient test. *Supp. J. Roy. Statist. Soc.*, **4**, 225.

PLACKETT, R. L. (1947). Limits of the ratio of mean range to standard deviation. *Biom.*, **34**, 120.

PLACKETT, R. L. (1954). A reduction formula for normal multivariate integrals. *Biom.*, **41**, 351.

POLYÀ, G. (1945). Remarks on characteristic functions. *Proceedings of the First Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press.

RAFF, M. S. (1956). On approximating the point binomial. *J. Amer. Statist. Ass.*, **51**, 293.

RIDER, P. R. (1955). The distribution of the product of maximum values in samples from a rectangular distribution. *J. Amer. Statist. Ass.*, **50**, 1142.

ROMANOVSKY, V. (1925). On the moments of the hypergeometrical series. *Biom.*, **17**, 57.

RUBEN, H. (1954). On the moments of order statistics in samples from normal populations. *Biom.*, **41**, 200.

SAMPFORD, M. (1953). Some inequalities on Mill's ratio and related functions. *Ann. Math. Statist.*, **24**, 130.

SANDON, F. (1924). Note on the simplification of the calculation of abruptness coefficients to correct crude moments. *Biom.*, **16**, 193.

SARHAN, A. E., and GREENBERG, B. G. (1956). Estimation of location and scale parameters by order statistics from singly and doubly censored samples. *Ann. Math. Statist.*, **27**, 427.

SHENTON, L. R. (1954). Inequalities for the normal integral, including a new continued fraction. *Biom.*, **41**, 177.

SHEPPARD, W. F. (1898). On the application of the theory of error to cases of normal distributions and normal correlations. *Phil. Trans*, **A, 192**, 101, and *Proc. Roy. Soc.*, **62**, 170.

SHEPPARD, W. F. (1939, posthumous). *The Probability Integral.* British Ass. Math. Tables, vol. 7. Cambridge University Press.

SICHEL, H. S. (1949). The method of frequency moments and its application to Type VII distributions. *Biom.*, **36**, 404.

SLUTSKY, E. E. (1950). Tables for the calculation of the Incomplete Gamma Function and the χ^2 probability function. Moscow, *Izdat. Akad. Nauk. S.S.S.R.*

STIELTJES, J. (1918). Recherches sur les fractions continues. *Œuvres*, Groningen.

" STUDENT " (1908). On the probable error of a mean. *Biom.*, **6**, 1.

" STUDENT " (1919). An explanation of deviations from Poisson's law in practice. *Biom.*, **12**, 211.

TEICHROEW, D. (1956). Tables of expected values of order statistics and products of order statistics. *Ann. Math. Statist.*, **27**, 410.

THIELE, T. N. (1903). *Theory of Observations.* Reprint in *Ann. Math. Statist.*, **2**, 165, of the English version published in 1903.

THOMPSON, C. M. (1941). Tables of percentage points of the χ^2-distribution. *Biom.*, **32**, 187.

THOMPSON, C. M., PEARSON, E. S., COMRIE, L. J., and HARTLEY, H. O. (1941). Tables of percentage points of the incomplete beta-function. *Biom.*, **32**, 151.

TIPPETT, L. H. C. (1925). On the extreme individuals and the range of samples taken from a normal population. *Biom.*, **17**, 364.

TUKEY, J. W. (1950). Some sampling simplified. *J. Amer. Statist. Ass.*, **45**, 501.

VON MISES, R. (1936). La distribution de la plus grande de n valeurs. *Revue de l'Union Interbalkanique*, **1**, 1.

WICKSELL, S. D. (1917). On logarithmic correlation with an application to the distribution of ages at first marriage. *Medd. Lunds Astr. Obs.*, No. 84.

WILLIAMS, J. D. (1946). An approximation to the probability integral. *Ann. Math. Statist.*, **17**, 363.

WILSON, E. B., and HILFERTY, M. M. (1931). The distribution of chi-square. *Proc. Nat. Acad. Sci.*, **17**, 694.

WISE, M. E. (1954). A quickly convergent expansion for cumulative hypergeometric probabilities, direct and inverse. *Biom.*, **41**, 317.

WISHART, J. (1929). The correlation between product-moments of any order in samples from a normal population. *Proc. Roy. Soc. Edin.*, **49**, 1.

WISHART, J. (1949). Cumulants of multivariate multinomial distributions. *Biom.*, **36**, 47.

WISHART, J. (1952). Moment coefficients of the k-statistics in samples from a finite population. *Biom.*, **39**, 1.

WOLD, H. (1934*a*). Sulle correzione di Sheppard. *Giorn. Ist. Ital. Att.*, **4**, 304.

WOLD, H. (1934*b*). Sheppard's correction formulæ in several variables. *Skand. Akt.*, **17**, 248.

YASUKAWA, K. (1926). On the probable error of the mode of frequency-distributions. *Biom.*, **18**, 263.

YATES, F. (1935). Some examples of biassed sampling. *Ann. Eug. Lond.*, **6**, 202.

YUAN, P. T. (1933). On the logarithmic frequency distribution. *Ann. Math. Statist.*, **4**, 30.

YULE, G. U. (1910). On the distribution of deaths with age when the causes of death act cumulatively. *J. Roy. Statist. Soc.*, **73**, 26.

YULE, G. U. (1927). On reading a scale. *J. Roy. Statist. Soc.*, **90**, 570.

YULE, G. U. (1938). On some properties of normal distributions, univariate and bivariate, based on sums of squares of frequencies. *Biom.*, **30**, 1.

ZIA UD-DIN, M. (1954). Expression of the k-statistics, k_9 and k_{10}, in terms of power sums and sample moments. *Ann. Math. Statist.*, **25**, 800.

INDEX

(References are to page numbers)

Abdel Aty, S. H., tables of multiple k-statistics, 303.

Abortion, distribution of women by term of, (Table 1.22) 27.

Absolute moments, 62–3 ; Liapounoff's inequality for, 63 and (Exercise 3.15) 91.

Accidents, London bus drivers', distribution of, (Table 1.21) 26 ; and the Poisson and negative binomial distributions, (Table 5.3) 128–30.

Aeppli, *see* Pólya-Aeppli.

Aitchison, J., and Brown, J. A. C., *The Lognormal Distribution*, 169.

Aitken, A. C., independence of quadratic forms in correlated normal variates, 362 and (Exercise 15.13) 367.

Anis, A. A., multivariate normal dispersion matrix, (Exercise 15.11) 367.

Anthropometric Committee of the British Association, data from Report of, (Table 1.7) 8 ; (Table 1.10) 9.

Antimode, 39.

Approximations to sampling distributions, 274–324.

Arithmetic mean, *see* Mean, arithmetic.

Asymmetrical distributions, 7–9 ; *see also* Skewness.

Attributes, sampling of, 219–26 ; (Exercises 9.1–9.12 and 9.14) 226–7 ; (Exercise 10.4) 242.

Augmented symmetric functions, 276.

Australian marriages, distribution of, (Table 1.8) 8 ; moments of, (Example 3.1) 57–8 ; skewness and kurtosis of, (Example 3.18) 86.

Average, *see* Mean.

Average grouping corrections to moments, 78–9 ; *see also* Corrections ; Sheppard's corrections.

b_1, b_2, β_1, β_2, *see* Skewness, Kurtosis.

Babington Smith, B., data on telephone directory digits, (Table 1.4) 6 ; data on randomness, (Example 9.3) 209–10 ; tables of Random Sampling Numbers, 214, 218.

Baker, G. A., distribution of sums from Type A series, (Exercise 11.10) 270.

Baldwin, E. M., table of percentage points of t-distribution, 377.

Bartlett, M. S., characterization of normal distribution, 364–5.

Barton, D. E., and Dennis, K. E. R., Type A series, 162.

Bayes' theorem, 198–9 ; postulate, 199–203 ; and maximum likelihood, 202–3 and (Exercises 8.1–2) 204 ; in sampling of attributes, (Exercise 9.3) 226.

Beans, distribution of, (Table 1.15) 20 ; histogram of, (Fig. 1.6) 21 ; Pearson distribution fitted, (Example 6.1) 153–4 ; Gram-Charlier series fitted, (Example 6.2) 160 ; lognormal distribution fitted, (Example 6.5) 170.

Bernoulli numbers and polynomials, 80–1 and (Exercise 4.20) 119.

Beta distribution, of first kind, arithmetic mean, (Example 2.2) 35 ; geometric and harmonic means, (Example 2.4) 36–7 ; variance, (Example 2.8) 45 ; of first kind and second kind, 150–1 ; Jacobi polynomials for expansion in series, 163 ; Johnson transformations, (Exercises 6.14–15) 177–8 ; *see also* Type I, Type VI distributions.

Bias, in sampling, 208–11 ; in choosing plants, (Example 9.1) 208 ; in scale reading, (Example 9.2) 208–9 ; in reading randomizing machine, (Example 9.3) 209–10 ; in crop forecasting, (Example 9.4) 210 ; in estimation, 225 and (Exercise 9.13) 227.

Bienaymé-Tchebycheff inequality, 88, 104, 192 ; in sampling of attributes, (Exercise 9.4) 226 ; used with standard error, 234.

Binomial distribution, frequency-generating function, 24 ; moments, (Example 2.3) 35 ; (Example 2.9) 45–6 ; (Example 3.2) 58–9 ; m.g.f. (c.f.), (Example 3.5) 61 ; factorial moments (Example 3.8) 67 and (Exercise 3.7) 90 ; factorial cumulants, (Example 3.14) 75 ; skewness and kurtosis, (Example 3.19) 86 ; limiting normality,

(Example 4.6) 106–7 ; generally, 120–5 ; c.f., moments and cumulants, 121–2 and (Exercise 5.1) 143 ; d.f., 123–5 ; tables of, 124–5 ; error in normal approximation, 125 ; mixed binomial population, 126–8 and (Exercise 5.6) 144 ; bivariate, 141–3 ; incomplete moments, (Exercises 5.2–3) 143–4 ; mean deviation of, (Exercise 5.4) 144 ; in sampling of attributes, 219–20 ; (Exercise 9.14) 227 ; with negative index, *see* Negative binomial distribution ; distribution of means of, (Example 11.10) 259–60.

Birth-rates, distribution of in Local Government Areas, (Table 1.1) 3 ; frequency polygon of, (Fig. 1.1) 4.

Bivariate binomial distribution, 141–3.

— distributions, 20–4, (Exercise 1.22) 31 ; fitting of, 174–5.

— k-statistics, 308–12.

— moments and cumulants, 81–4 ; standard errors of, 235–6.

— normal distribution, (Example 1.1) 23 ; (Example 3.17) 82–3 ; moments, (Example 3.19) 91 ; uniquely determined by moments, (Exercise 4.17) 119 ; joint c.f. of squares of variates, (Exercise 4.19) 119 ; as limit of bivariate binomial, 142–3 ; bivariate Gram-Charlier series and a differential equation, (Exercises 6.16–17) 178 ; c.f. of sum of variates, (Exercise 7.13) 196 ; distribution of ratio of variates, (Exercise 11.22) 273 ; moments of sample covariance, (Example 13.3) 322 ; variances and covariances of k-statistics, (Exercises 13.1–2 and 13.7–8) 322–3 ; (Example 15.1), 349–50 ; Sheppard's theorem on median dichotomy, 351 ; (Exercises 15.7–8), 367 ; covariance conditional on a linear constraint, (Exercise 15.18) 368 ; univariate normal margins not sufficient for, (Exercise 15.20) 368 ; sampling from, 383–93 ; distribution of sample correlation coefficient, 383–92 ; distribution of sample regression coefficients, 392–3.

— Poisson distribution, (Exercise 5.11) 145.

Booker, H. S., data from, (Table 1.21) 26.

Bortkiewicz, L. von, data on suicides, (Table 1.6) 7.

Bose, S. S., distribution of variance-ratio for correlated variates, (Exercise 16.16) 395.

Box, G. E. P., transforming normal to rectangular variates, (Exercise 1.21) 31.

Bridge, probabilities at, (Example 7.2) 183–4.

Brown, J. A. C., *see* Aitchison, J.

Burr, I. W., fitting of d.f.'s, 173–4 and (Exercise 6.19) 178.

Carleman, T., uniqueness criterion for problem of moments, 111.

Carlton, G. A., sampling from rectangular distribution, (Exercises 14.17–18) 346.

Cauchy distribution, moments infinite, (Example 3.3) 59–60 ; c.f., (Example 3.13) 73–4 ; inversion of c.f., (Example 4.2) 98 ; distribution of sample mean the same, (Example 11.1) 248–9 and (Example 11.17) 264 ; as ratio of squares of normal variates, (Example 11.21) 268 ; as ratio of non-normal variables, (Exercise 11.23) 273.

Cauchy-Schwarz inequality, 38.

Cauchy-type distributions in extreme-value theory, 333.

Central Limit theorem, 193–5, 223–4, 234 ; failure of for Cauchy distribution, (Example 11.1) 248–9 ; for rectangular variates, (Example 11.9) 259.

Centre of a distribution, 40 ; *see also* Mid-range.

C.f., *see* Characteristic function.

C.g.f., cumulant-generating functions, 67–8.

Chakravarti, M. C., mean deviation and standard deviation, 45 footnote.

Chambers, E. G., data from, (Table 1.21) 26.

Change of variables, 23–4, 247–8.

Characteristic function (c.f.), as m.g.f., 61–2 ; in general, 94–119 ; relation to d.f., 94–101 and 116–17 ; conditions for a function to be a c.f., 101–2 and (Exercises 4.5, 4.10–11) 117–18 ; limiting properties, 102–6 ; multivariate, 107–9 ; of a function of a variate, 108–9 ; expansion of, 112–13 ; derivatives and existence of moments, 61 and 112–14 ; of a non-negative function inverted, (Exercise 4.14) 118 ; relations between derivatives, (Exercise 4.15) 119 ; of a sum, 188–9 and (Example 7.6) 190 ; in obtaining exact sampling distributions, 259–63 ; in method of steepest descents, 269.

Characterizations, of normal distributions, 362–6 and (Exercise 15.16) 368 ; of Gamma distribution, (Exercise 15.22) 368.

Charlier, C. V. L., Tchebycheff-Hermite polynomials, 156 ; *see also* Gram-Charlier series.

Chi-square, *see* χ^2 (indexed at x).

Chu, J. T., bound for variance of median, (Exercise 14.7) 345.
Class-frequency, 2.
Class-interval, 2 ; ambiguities in, 5.
Cloudiness, data on intensities of, (Table 1.11) 10 ; Johnson distributions and Pearson Type I distribution fitted, (Example 6.6) 171–2.
Clusters, distributions of numbers in, (Exercises 5.7–8) 144–5 and (Exercises 5.21–2) 147.
Cochran, W. G., random permutations, 216–17 footnote ; independence of quadratic forms, 357 ; theorem on decomposition of sum of squares, 360–2 and (Exercise 15.17) 368 ; approximation to z-distribution, 380–1.
Cochran's theorem, 360–2.
Colonies, distributions of numbers in, *see* Clusters
Compatible events, probabilities of realization, (Exercises 7.1–5) 195.
Concentration, coefficient of, 48 ; curve, 48–9.
Conditional distributions, 21–2.
Contagious distribution, *see* Type A.
Continuous distributions, 7, 12 ; obtained by inversion of c.f., 96–7 ; sampling from, 217–19.
Convolution of random variables, *see* Sum.
Cook, M. B., multivariate moments and cumulants, 84 footnote ; sampling cumulants of multivariate k-statistics, 311, 322 and (Exercise 13.4) 323.
Co-operative Study on distribution of correlation coefficient, 388 and (Exercise 16.14) 395.
Corbet, A. S., logarithmic distribution, (Table 5.4) 132.
Cornish, E. A., normalization of distributions, 166 ; approximation to z-distribution, 380.
Corrections for grouping, 34, 46, 75–81, 84 and (Exercises 3.10, 3.13) 90–1 ; *see also* Sheppard's corrections.
Correlation coefficient, standard error, (Example 10.6) 235–6, 243 ; variance-stabilizing transformation, (Exercise 10.17) 245, 383 ; sampling distribution for bivariate normal parent, 383–90 ; tables of, 389 ; Fisher's z-transformation, 390–1, (Exercise 16.21) 396 ; recurrence relation for f.f., (Exercise 16.14) 395 ; differential equation for f.f., (Exercise 16.15) 395 ; mean and variance if parent variates independent, (Exercise 16.17) 396.
Covariance, 82 ; moments of, (Example 133) 322 ; matrix, *see* Dispersion matrix.
Cox, D. R., distribution of range, 339 and (Exercise 14.15) 345–6 ; reduced range, 342.
Cox, G. M., random permutations, 216–17 footnote.
Craig, A. T., quadratic forms, 358, 362, (Exercise 15.16) 368.
Craig, C. C., corrections to moments of discrete distributions, 79 and (Exercise 3.13) 91.
Craig, R., data from, (Table 1.5) 6.
Cramér, H., conditions for a function to be a c.f., 102 ; multivariate Carleman criterion for moments, 111 ; convergence of Gram-Charlier series, 161–2, 167 ; Central Limit theorem, 194 ; distribution of a ratio, 266–7.
Crop-forecasting, bias in, (Example 9.4) 210.
Cumulants, definition, 67 ; generating function for, 67–8 ; invariantive properties, 68 ; relations with moments, 68–71 ; existence of, 71–2 ; calculation of, 72–4 ; multivariate, 81–4 ; grouping corrections, 81, 84 and (Exercise 3.13) 91 ; can only all vanish after the first ((Example 4.3) 99) or the second, (Exercise 4.11) 118 ; standard errors, (Example 10.4) 233 and (Exercise 10.12) 244 ; structural properties of, 318–20 ; *see also* k-statistics.
Cumulant-generating function (c.g.f.) or cumulative function, 68.

Dairy herds, distribution of milk-production costs, (Table 1.9) 9.
Daniels, H. E., distribution of ratios, 266, (Exercise 11.24) 273 ; method of steepest descents, 269.
Das, S., Mills' ratio, (Exercise 5.13) 145.
David, F. N., tables of symmetric functions, 69 footnote, 276 ; distribution of difference of Type III variates, (Exercise 11.15) 271–2 ; moments of symmetric functions, 299 ; moments of variance-ratio, 300 ; moments of order-statistics, 329 and (Exercises 14.1, 14.3) 344 ; tables of correlation coefficient, 389, 391.
David, H. A., upper bounds for means of order-statistics, 336–7 ; bounds for mean of range, 340 ; distribution of ratio of random deviation from sample mean to the sample standard deviation, (Exercise 16.20) 396.

Deciles, 40 ; interdecile range, 42 ; standard errors of, 243 ; *see also* Order-statistics.
Decomposition of sum of squares, 360–2 and (Exercises 15.16–17) 368.
Degrees of freedom, 268 footnote, 369, 375, 377.
Deming, W. E., matching of lists, (Exercise 10.21) 245.
Demoivre, A., discoverer of normal distribution, 135 footnote.
D.f., *see* Distribution function.
Dice-throws, Weldon's data, (Table 1.14) 19 ; (Table 1.16) 25 ; (Table 5.1) 121 ; (Example 9.10) 221.
Digits, distribution in telephone directory, (Table 1.4) 6.
Discontinuous distributions, 6–7, 11, 14–15 ; inversion of c.f., 97.
Discrete, *see* Discontinuous.
Dispersion, measures of, 42–51.
Dispersion matrix, 349 ; (Exercises 15.1–3, 15.9–11) 366–7.
Distribution function (d.f.), 11–15 ; graph of, for determining quantiles, 41–2 ; determined by c.f., 94–7 ; limiting properties, 102–6 ; and moments, 109–12 ; dual role with c.f., 116–17 ; and probability, 186–7.
Double exponential (Laplace) distribution, c.f. inverted, (Exercise 4.3) 117 ; variance of midrange, (Example 14.5) 341.
Dronkers, J. J., asymptotic distributions of extreme values, 336.
Durbin, J., Jacobi polynomials in Beta distribution expansion, 163.
Dyson, F. J., kurtosis, (Exercise 3.20) 92.

Edgeworth, F. Y., citing Weldon's dice data, (Table 1.14) 19 ; form of Gram-Charlier series, 159.
Edgeworth series, 157–9, 163 ; relation between mean, median and mode, (Exercise 6.20) 179 ; method of steepest descents more powerful than, 269.
Elderton, Sir William P., grouping corrections for moments, 79 and (Exercise 3.10) 90 ; fitting of Pearson distributions, 152.
Elfving, G., asymptotic distribution of range, 341–2 and (Exercise 14.20) 346.
Envelopes and letters, matching problem, (Example 7.3) 184–5 and (Exercise 7.4) 195.
Error, *see* Standard error.
Estimation, in sampling of attributes, 219–22 ; in large samples generally, 223–4.
Euler-Maclaurin sum formula, 75.
Expectations, Expected values, 51–2.
Expected largest value, 332 footnote.
Exponential distribution, mean, standard deviation, mean difference and interquartile range, (Exercise 2.6) 52 ; cumulants, (Exercise 3.6) 89 ; moments of mean difference, 242 ; distribution of order-statistics, 325 ; variance of smallest value, (Exercise 14.8) 345 ; of time-intervals, (Exercise 16.6) 394 ; *see also* Gamma distribution, Double exponential distribution.
Exponential-type distributions in extreme-value theory, 332, 342–3.
Extreme values, 329 ; asymptotic distributions of, 330–6 ; tables for, 343.

F-distribution, *see* Variance-ratio distribution.
Factorial cumulants, 75.
Factorial moments, 63–7 ; in terms of ordinary moments, 64 ; calculation of, 64–6 ; Sheppard's corrections to, 80.
Farmer, E., data from, (Table 1.21) 26.
Fecundity, distribution of brood-mares by, (Table 1.19) 26.
Federighi, E. T., tables of t-distribution, 377.
Feller, W., *Introduction to Probability Theory*, 194.
F.f., *see* Frequency-function.
Fieller, E. C., tables of random correlated normal deviates, 215 ; distribution of a ratio of correlated normal variates, (Exercise 11.22) 273.
Finite populations, 207 ; k-statistics in sampling from, 300–3 ; bivariate, 311–12.
Finney, D. J., distribution of variance ratio for correlated normal variates, (Exercise 16.16) 395.
First Limit theorem, 104–5 ; converse, 105–6.
Fisher, Sir Ronald A., Sheppard's corrections, (Exercise 3.17) 92 ; logarithmic distribution, 132 ; normalization of distributions, 166 ; tables of Random Sampling Numbers, 214 ; variance of mean deviation, 240 ; variance of function of multinomial frequencies, (Exercise 10.2) 242 ; variance-ratio (F and z) distributions, *see* Variance-ratio ; k-statistics, 277 ; sampling distribution of $\sqrt{b_1}$, 298 ; pattern functions, (Exercises 12.4–5, 12.7) 305 ; m.g.f. of moments of k-statistics, (Exercise 12.14) 306 ; combinatorial rules for k-statistics in normal case, (Exercises 13.10–12) 323–4 ; asymptotic distributions of

extreme values, 330–3 ; table of χ^2 percentage points, 371, 400 ; approximation to χ^2 distribution, 371–4 ; tables of t-distribution, 377 ; tables of percentage points of F- and z-distributions, 379 ; approximations to z-distribution, 380–1 ; distribution of correlation coefficient, 385 ; z-transformation of correlation coefficient, 390–1 ; d.f. of t-distribution, (Exercise 16.10) 395.

F.m.g.f., factorial moment-generating function, 66–7 ; inversion, (Exercise 3.24) 93.

Foster, F. G., dispersion matrix, (Exercise 15.9) 367.

Fourier transforms and c.f., 94, 117.

Franel, J., distribution of digits in mathematical tables, 214 footnote.

Fréchet, M., conventions regarding mean values, 59 footnote and (Exercise 7.19) 197 ; asymptotic distribution of extreme values, 333.

Frequency (class-frequency), 2.

Frequency-distributions, 1–31 ; discontinuous, 6–7 ; continuous, 7–11 ; symmetrical, 7–8 ; asymmetrical, 7–9 ; J-shaped, 9–10 ; U-shaped, 10 ; genesis of, 19 ; multivariate, 20–4 ; conditional, 21–2 ; marginal, 22.

Frequency functions (f.f.), 11–15 ; determined by c.f., 94–7 ; normalization of, 163–7.

Frequency-generating functions, 24 ; determining mean ((Example 2.3) 35) and variance, (Example 2.9) 45–6.

Frequency-moments, of normal distribution, (Exercise 5.10) 145 ; fitting distributions by, 174.

Frequency polygon, 3–5 ; bivariate form, 21.

Frisch, R., factorial moments, 64 ; cumulants and moments of binomial distribution, (Exercises 5.1, 5.3) 143–4.

Functions of random variables, c.f. of, 108–9 ; standard errors of, 231–3 ; covariances of, 232.

γ_1, γ_2, *see* Skewness, Kurtosis.

Galbrun, H., convergence of Gram-Charlier series, 161.

Gamma distribution, cumulants and inversion of c.f., (Example 4.4) 99–100 ; moments determining d.f., (Exercise 4.6) 117 ; square of a normal variate, (Exercise 4.18) 119 ; Type A Gram-Charlier expansion, (Example 6.3) 162–3 ; Laguerre polynomials for series expansion, 163 ; normalization of, (Example 6.4) 166–7 ; logarithmic (Johnson) transformation, (Exercise 6.13) 177 ; characterization of, (Exercise 15.22) 368 ; *see also* Type III distribution.

Garwood, F., approximations to χ^2 distribution, (Table 16.1) 373–4.

Gauss, C., and normal distribution, 135 footnote.

Gauss-Winckler inequality, (Exercise 3.18) 92.

Gayen, A. K., z-transformation of correlation coefficient, 391 footnote.

Geary, R. C., distribution of a ratio, 266, (Exercise 11.11) 270–1 ; independence of mean and variance in normal samples, (Exercise 11.19) 272 and (Example 12.7) 289–90.

Generating functions, *see* c.g.f., frequency-generating functions, f.m.g.f., m.g.f.

Geometric mean, *see* Mean, geometric.

Geometric probabilities, (Example 7.4) 185–6 and (Exercises 7.9–11) 196.

Gini, C., coefficient of mean difference, 46, 240–3 ; coefficient of concentration, 48.

Glasser, G. J., matching of lists, (Exercise 10.21) 245.

Godwin, H. J., Tchebycheff-type inequalities, 89 ; distribution of mean deviation, 240 ; moments of order-statistics, 328.

Goodman, L. A., matching of lists, (Exercise 10.21) 245 ; variance of product, (Exercise 10.23) 245.

Goodness-of-fit χ^2 distribution, (Example 15.3) 355–6 ; (Exercise 15.21) 368.

Gosset, W. S., *see* " Student."

Gram-Charlier series, Type A, 156–63 ; d.f. of, 159 ; fitted to bean data, (Example 6.2) 160 ; conditions for convergence, 161–2 and (Exercise 6.9) 176 ; bivariate, 175 and (Exercise 5.19) 178 ; (Exercises 6.10–11), 176 ; distribution of sums from, (Exercise 11.10) 270 ; *see also* Edgeworth series.

— — — Types B and C, 163 and (Exercise 6.8) 176.

Graphical representation of frequency-distributions, *see* Frequency polygon, Histogram.

Greenberg, B., table of variances of normal order-statistics, 329.

Greenwood, M., data on industrial accidents, (Table 5.3) 129.

Grouping corrections, *see* Corrections; Sheppard's corrections.

Guldberg, S., cumulants of multinomial distributions, (Exercise 5.19) 147.
Gumbel, E. J., bivariate d.f., (Exercise 1.22) 31 ; theory of extreme values, 330 footnote, 332 footnote, 335–6 ; reduced range and mid-range, 342 ; tables for extreme-value distributions, 343 ; distribution function of range, (Exercise 14.19) 346.

Haldane, J. B. S., cumulants of binomial distribution, (Exercise 3.1) 143 ; relation between mean, median and mode in Edgeworth expansion, (Exercise 6.20) 179 ; approximation to χ^2 distribution, (Exercise 16.8) 394.
Hall, A. D., data on grain-yields, (Table 1.20) 26.
Hall, P., distribution of mean in samples from rectangular distribution, (Example 11.9) 257–9.
Hamburger, H., problem of moments, 109 footnote.
Hardy, G. H., Littlewood and Pólya, *Inequalities*, 38.
Harmonic mean, *see* Mean, harmonic.
Harter, H. L., table of means of normal order-statistics, 329 ; tables of range in normal samples, 339.
Hartley, H. O., distribution of mean deviation, 240 ; upper bounds for means of order-statistics, 336–7 ; tables of distribution of range in normal samples, 339, 343 ; bounds for expected value of range, 340 ; table of χ^2, 370 ; distribution of ratio of extreme deviation from sample mean to the sample standard deviation, (Exercise 16.20) 396.
Hastings, C., moments of order-statistics, 328.
Hatke, M. A., fitting of d.f.'s, 173–4.
Height, distribution of men by, (Table 1.7) 8 ; frequency polygon of, (Fig. 1.3) 7 ; mean, (Example 2.1) 33–4 ; median, (Example 2.5) 38–9 ; quartiles, (Example 2.6) 40–1 ; distribution curve, (Fig. 2.1) 41 ; mean deviation and standard deviation, (Example 2.7) 43–5 ; mean difference, (Example 2.10) 50–1 ; factorial and ordinary moments, (Example 3.7) 65–6 ; cumulants, (Example 3.9) 72 ; normal distribution fitted, (Table 5.6) 140 ; standard error of mean, (Example 10.1) 230–1 ; sampling variation of higher moments, (Exercise 10.1) 242.
Helmert, W., variance of mean deviation, 240 ; orthogonal transformation, (Example 11.3) 250–1.
Hermite polynomials, 155–6.
Hilferty, M. M., approximation to χ^2 distribution, 371–4.
Histogram, 5 ; bivariate, 21.
Hogg, R. V., decomposition of quadratic forms, (Exercise 15.16) 368.
Hojo, T., moments of order-statistics, 327–8.
Hölder's inequality, 38 and (Exercise 3.15) 91.
Homoscedastic, 383.
Hotelling, H., measure of skewness, (Exercise 3.22) 93 ; distribution of correlation coefficient, 388–90 ; transformations of correlation coefficient, 391 and (Exercise 16.19) 396 ; relations for f.f. of correlation coefficient, (Exercises 16.14–15) 395.
Hsu, C. T., sampling cumulants of normal k-statistics, 297.
Hypergeometric distribution, 133–5 ; c.f. and moments, 134 ; limiting forms of, 140–1 ; factorial moments, (Exercise 5.26) 147 ; in sampling of attributes, (Example 9.12) 224.
Hypotheses, 201–3.
Hypothetical population, 19–20, 187 ; sampling from, 207–8.

Income, distribution of persons by, 3, 4, 14.
Incomplete moments, 48 ; of binomial distribution, (Exercises 5.2–3) 143–4.
Independence, 22 ; in bivariate binomial distribution, 142–3.
Inequalities for moments, Liapounoff's, 63 and (Exercise 3.15) 91–2 ; Gauss-Winckler, (Exercise 3.18) 92 ; (Exercise 3.19) 92.
Interdecile range, 42.
Interquartile range, 42 ; standard error of semi-interquartile range, (Example 10.8) 239 ; standard error in normal case, 327.
Interval, *see* Class-interval.
Inverse probability, *see* Bayes' theorem.
Inverse sampling for attributes, (Example 9.13) 225–6.
Inversion theorem on c.f.'s, 94–101, (Exercise 4.22) 119.
Irwin, J. O., distribution of mean from Type II distribution, (Exercise 11.14) 271 ; difference of two Poisson variates, (Exercise 11.16) 272 ; sampling from finite populations, 301.

J-shaped distributions, 9–10.
Jackson, D., on median, 39.
Jacobi polynomials, for Beta distribution expansions, 163.
Jacobian (determinant) of transformation, 23–4, 247–8.
James, G. S., transformation of variate, (Exercise 13.6) 323 ; decomposition of sum of squares, 360–1.
Jeffreys, Sir Harold, and Bayes' postulate, 199 footnote.
Jenkinson, A. F., asymptotic distribution of extreme values, 330.
Johannsen, W., bean data cited by Pretorius, (Table 1.15) 20.
Johnson, N. L., mean deviation of binomial distribution, (Exercise 5.4) 144 ; system of variate-transformations, 167, 171, 172 ; fitting bivariate distributions, 174–5 ; moments of variance-ratio, 300 ; moments of order-statistics, 329 and (Exercises 14.1 and 14.3) 344.
Johnson transformations, 167–73 ; fitted to bean data, (Examples 6.5, 6.7) 170, 172–3 ; fitted to cloudiness data, (Example 6.6) 171–2 ; applied to Gamma distribution, (Exercise 6.13) 177.
Jørgensen, N. R., tables of Tchebycheff-Hermite polynomials, 156.

k-statistics, 277–324 ; definition, 277–8 ; as seminvariants, 278 ; in terms of symmetric products and sums, 280–1 ; rules for obtaining sampling cumulants of, 281–9 ; rules for sampling cumulants when parent distribution normal, (Example 12.5) 288–9 ; sampling cumulants formulae, 290–7 ; sampling cumulants formulae for normal parent, 295–7 ; in sampling from finite population, 300–3 ; multiple k-statistics for estimating non-linear functions of cumulants, 303–4 and (Exercises 12.18–19) 307 ; m.g.f. of moments of k-statistics, (Exercises 12.14–16) 306–7 ; bivariate case, 308–12 ; proof of rules for sampling cumulants, 313–18 ; structural properties of, 318–20 ; summary of formulae in tensor form, 320–2.
κ criterion of Type in Pearson distributions, 151 ; infinite for Type III, (Exercise 6.5) 175 ; unity for Type V, (Exercise 6.6) 175.
Kaplan, E. L., tensor formulae for k-statistics, 320–2 and (Exercise 13.9) 323.
Kaplansky, I., kurtosis, (Exercise 3.21) 93.
Kawata, T., characterization of normal distribution, 364.
Kendall, D. G., and Rao, K. S., sequences of d.f.'s., 103 ; Second Limit theorem, 112–16 ; moments becoming infinite as d.f. tends to normality, (Exercise 4.12) 118.
Kendall, M. G., digits from telephone directories, (Table 1.4) 6 ; tables of symmetric functions, 69 footnote and 276 ; Sheppard's corrections, 79 ; bivariate Gram-Charlier series, (Exercises 6.16–17) 178 ; theories of probability, 180 ; data on randomness, (Table 9.3), 209–10 ; table of Random Sampling Numbers, 214, 218 ; random permutations, 216 footnote ; sampling from finite populations, 301 ; multiple k-statistics, 303 ; m.g.f. of moments of k-statistics, (Exercises 12.15–16) 307 ; symbolic derivation of multivariate k-statistic formulae from the univariate, 311 ; variance of observation nearest to true mean in normal samples, (Exercises 14.10–11) 345 ; multivariate normal integral, 351–2 ; transformations of correlation coefficient, (Exercise 16.19) 396.
Kullback, S., distributions of geometric means, (Exercises 11.3, 11.9) 269–70.
Kurtosis, 85-6 ; (Exercises 3.20–1) 92–3 ; standard error of b_2, 243 ; moments of b_2, (Exercise 12.10) 306–7.

Laguerre polynomials and Gamma distribution expansions, 163.
Laha, R. G., characterization of Gamma distribution, (Exercise 15.22) 368.
Lancaster, H. O., independence of quadratic forms in normal variates, 358, 362 ; decomposition of sums of squares, 360 ; characterizations of normal distribution, 365–6 ; condition for quadratic form to be a component of another, (Exercise 15.15) 368.
Laplace, P. S. (Marquis de), and normal distribution, 135 footnote ; continued fraction for normal distribution, 137–8 and (Exercise 5.12) 145 ; succession rule, (Example 8.3) 200–1 and (Exercise 8.4) 205.
Laplace distribution, *see* Double exponential distribution.
Large Numbers, Laws of, Weak, 192–3 and (Exercise 7.20) 197 ; Strong, 193.
Large samples, approximations in, *see* Standard error.

Lawley, D. N., sampling cumulants of normal k-statistics, 297.
Lee, A., data from, (Table 1.19) 26.
Leptokurtosis, 86.
Letters and envelopes, matching, (Example 7.3) 184–5 and (Exercise 7.4) 195.
Lévy, P., c.f., 94.
Liapounoff, A., inequality for moments, 63 and (Exercise 3.15) 91–2.
Lieberman, G. J., hypergeometric tables, 135.
Likelihood, 198–9 ; maximum likelihood principle, 201–4 ; and Bayes' theorem, 202–3 and (Exercises 8.1–2) 204 ; in sampling of attributes, (Exercises 9.5, 9.10) 226–7.
Limit theorems, *see* First Limit theorem, Second Limit theorem, Central Limit theorem, Large Numbers.
Lindeberg, J. W., Central Limit theorem, 194.
Linear functions, of variates, standard error of, 232 ; of independent normal variates, distribution of, (Example 11.2) 249–50, (Example 11.5) 253–4 ; 350 ; independence of quadratic forms, 359.
Lloyd, E. H., multivariate normal dispersion matrix, (Exercise 15.11) 367.
Location, measures of, 32–42.
Logarithmic distribution, 131–3 ; c.f. and moments, 133 ; combined in clustering with Poisson to produce negative binomial, (Exercise 5.21) 147.
Logistic distribution, c.f., (Exercise 4.21) 119 ; d.f. of extreme value, (Exercise 14.2) 344.
Lognormal distribution, 168–70.
Lomnicki, Z. A., standard error of Gini's mean difference, 241.
Lottery sampling, 212–13.
Lukacs, E., c.f.'s, 94 footnote, 102, 117 ; characterization of normal distributions by independence of sample mean and variance, (Exercise 11.19) 272, (Example 12.7) 290 and 362–4 ; characterizations, 366.

mth values, *see* Extreme values, Order-statistics.
Mallows, C. L., Tchebycheff-type inequalities, 89.
Marcinkiewicz, J., conditions for a c.f., 102 and (Exercise 4.11) 118.
Marginal distribution, 22.
Markoff, A., inequality, 88, 192, (Exercise 7.16) 196.
Martin, E. S., grouping corrections, 79.
Matching, of envelopes and letters, (Example 7.3) 184–5 and (Exercise 7.4) 195 ; of lists, (Exercises 10.21–2) 245.
Matérn, B., independence of non-negative quadratic forms, 358.
Mathematical tables, distribution of digits from, 214.
Maximum likelihood, *see* Likelihood.
McKay, A. T., distribution of normal order-statistics, (Exercise 14.14) 345.
Mean, *see* Mean, arithmetic.
Mean, arithmetic, definition, 33 ; calculation of, 33–5 ; greater than geometric and harmonic means, 35–7 ; relation with median and mode, 39–40 and (Exercise 6.20) 179 ; as first moment, 43 ; standard error of, 230, 243 ; distribution from Cauchy parent, (Example 11.1) 248–9 ; distribution from normal parent, (Example 11.7) 255–6 ; independence of sample variance in normal samples, (Example 11.12) 260 ; distribution from rectangular parent, (Examples 11.9, 11.14) 257–9, 261 ; distribution from binomial parent, (Example 11.10) 259–60 ; from Poisson parent, (Example 11.11) 260 ; from Type III parent, (Example 11.13) 260–1 ; moments of, in sampling from finite populations, (Example 12.8) 301–2 ; uncorrelated with any even k-statistic, (Example 12.9) 302–3 ; *see also* Central Limit theorem.
Mean deviation, 42–3 ; calculation of, (Example 2.7) 43–5, (Exercise 2.21) 54 ; (about mean) no greater than standard deviation, 45 footnote ; minimized about median, (Exercise 2.1) 52 ; standard errors, 240, 243.
Mean difference, 46–7 and (Exercises 2.4, 2.9, 2.19) 52–4 ; calculation of, 49–51 and (Exercise 2.10) 53 ; standard error, 241–3.
Mean, geometric, 35–7 ; in value between arithmetic and harmonic means, 37–8 ; distribution of, 261–2 ; from rectangular parent, (Example 11.15) 262 ; from Type III distributions, (Exercises 11.3, 11.9) 269–70.
Mean, harmonic, 36 ; less than arithmetic and geometric means, 35–7.
Mean values, 51–2.
Median, 38 ; calculation, (Example 2.5) 38–9 ; indeterminacy in, 38–9 ; of bivariate distribution, 39 ; relations with mean and mode, 39–40 and (Exercise 6.20) 179 ; as minimizing origin for mean deviation, (Exercise 2.1) 52 ; standard error, (Ex-

ample 10.7) 237, 243, (Example 11.4) 253 ; distribution in normal case, 326–8 ; distribution in Cauchy case, (Exercise14.1) 344 ; bound for variance, (Exercise 14.7) 345.
Median dichotomy, 351.
Mendelian law, sampling of attributes, 219–20.
Mercer, W., data from, (Table 1.20) 26.
Merrington, M., table of percentage points of variance-ratio distribution, 379.
Mesokurtosis, 86, 136.
M.g.f., moment-generating functions, 60-2.
Midrange, 40 ; distribution of in normal case, 327 ; convergence of, to population value, 340–1 ; variance of, in normal case, (Example 14.4) 341 ; in double-exponential case, (Example 14.5) 341 ; reduced midrange, 342–3, (Exercise 14.4) 344.
Mills' ratio, 137 ; (Exercises 5.13–16) 145–6.
Milne-Thomson, L. M., *Calculus of Finite Differences*, 75 footnote.
Mises, R. von, asymptotic distributions of extreme values, 332.
Mode, 39 ; relation with mean and median, 39–40 and (Exercise 6.20) 179.
Moments, 43, 55 ; grouping corrections to, 46, 75–81, 84 ; about one point in terms of those about another, 55–6 ; calculation of, 56–9 ; absolute moments, *see* Absolute ; factorial moments, *see* Factorial ; inequalities for, *see* Inequalities ; in terms of factorial moments, 64 ; relationship with cumulants, 68–71 ; multivariate, 81–3 ; as characteristics of a distribution, 86–7 ; non-uniqueness of, *see* Problem of moments ; in fitting Pearson distributions, 152–4 ; standard errors of, 228–36 and (Exercise 10.5) 243 ; covariances of, (Exercises 10.7, 10.16) 244 ; distribution of, 262.
Moment-generating functions (m.g.f.), 60–2.
Monomial symmetric functions, 276.
Moore, G., data from, (Table 1.19) 26.
Moran, P. A. P., multivariate normal integral, 353–4.
Moriguti, S., bounds for frequency-moments, 174 ; bound for means of order-statistics, 337.
Muller, M., transforming normal to rectangular variates, (Exercise 1.21) 31.
Multimodal, 39.
Multinomial distribution, 141 ; multivariate, cumulants of, 143 ; cumulants, (Exercise 5.19) 147 ; variance of function of frequencies, (Exercise 10.2) 242.
Multivariate, distributions, 20–4 ; moments and cumulants, 81–4; c.f.'s, 107–9 ; sampling distributions, 268–9 ; k-statistics, 308–12, 318–22.
Multivariate normal (multinormal) distribution, rules for k-statistics, (Exercise 13.12) 324 ; generally, 347–68 ; singular, 347 and (Example 15.3) 355–6 ; c.f. and moments, 347–9 ; linear functions of normal variates, 350 ; marginal distributions normal, 350 ; integral, 350–4 and (Exercises 15.6, 15.12) 367 ; quadratic forms, 355–62 and (Exercise 15.13) 367 ; characterizations, 364–5 ; conditional distributions normal, (Exercise 15.1) 366 ; conditions on dispersion matrix, (Exercises 15.2–3) 366 ; examples of, (Exercises 15.9–11) 367 ; covariance conditional on a linear constraint, (Exercise 15.19) 368 ; marginal normal distributions not sufficient for, (Exercise 15.20) 368.
Murty, V. N., distribution of ratio of largest values in samples from rectangular distributions, (Exercise 11.20) 272.
Myller-Lebedeff, V., expansions in series of derivatives, 87 footnote.

Nair, U. S., moments of mean difference, 241–2.
Negative binomial distribution, 129–31 ; produced by Poisson with parameter a Type III random variable, 129–30 ; c.f. and cumulants, 131 ; produced by clustering from Poisson and logarithmic distributions, (Exercise 5.21) 147 ; in sequential sampling for attributes, (Example 9.13) 225-6 ; (Exercise 9.12) 227.
Negative exponential distribution, *see* Exponential distribution.
Newspapers read by students, (Table 1.23) 28.
Neyman, J., Type A contagious distribution produced by clustering from two Poisson variates, (Exercise 5.7) 144–5.
Normal distribution, moments and m.g.f., (Example 3.4) 60 ; cumulants, (Example 3.11) 73 ; mesokurtosis, 85–6, 136 ; c.f. inverted, (Example 4.1) 97 ; as limit of binomial, (Example 4.6) 106–7 ; determined uniquely by moments, (Example 4.7) 111 ; as limit of " Student's " distribution, (Example 4.8) 114–15 ; as limit of Poisson, (Example 4.9) 116 ; c.f. of square of a normal variate, (Exercise 4.18) 119 ; d.f. of, 136–8 and (Exercises 5.12–18)

145–7 ; tables of, 138–9 ; mean deviation, 139 ; fitted to height data, (Table 5.6) 140 ; frequency-moments, (Exercise 5.10) 145 ; with parameters random variables (Exercises 5.24–5) 147 ; as a Pearson distribution, (Exercise 6.1) 175 ; in Central Limit theorem, 193–5 ; joint c.f. of variate and its square, (Exercise 7.15) 196 ; random normal deviates, 215 ; in large-sample theory, 221–4 ; standard error of median, (Example 10.7) 237 ; standard error of semi-interquartile range, (Example 10.8) 239 ; methods of estimating parent standard deviation, 239 ; moments of mean deviation, 240 ; moments of mean difference, 241 ; variance of estimated sampling variance of sample variance, (Exercise 10.9) 244 ; distribution of sum of squares, 246–7, (Example 11.6) 254–5, (Example 11.16) 262–3 and (Example 11.18) 264–5 ; distribution of linear function of independent normal variates normal, (Example 11.2) 249–50 and (Example 11.5) 254–5 ; distribution of sample mean, (Example 11.7) 255–6, (Example 11.12) 260 ; characteristic independence of sample mean and variance, (Example 11.7) 255–6, (Example 12.7) 289–90, (Examples 15.4–6) 359–61, and 362–4 ; distribution of ratio of normal variates, (Example 11.21) 268, (Exercise 11.11) 270–1 and (Exercise 11.22) 273 ; expectations of moment-ratios, (Exercise 11.5) 270 ; all k-statistics uncorrelated, (Example 12.5) 288–9 ; sampling cumulants of k-statistics, 295–7 ; characteristic independence of sample mean and any other k-statistic, and zero correlation of mean and all other k-statistics, (Exercise 12.20) 307 ; combinatorial rules for k-statistics, (Exercises 13.10–11) 323 ; moments of median and other order-statistics, 326–8, (Example 14.1) 328–9, and (Exercise 14.14) 345 ; tables of moments of order-statistics, 329 ; extreme-value distribution, (Example 14.3) 333–4 ; distribution of range and tables of it, 339 ; variance of midrange, (Example 14.4) 341 ; variance of observation nearest to true mean, (Exercises 14.10–11) 345 ; quadratic forms in normal variates, 355–62 ; characterizations of normal distribution, 362–6 ; inequality for integral, (Exercise 15.5) 367 ; associated distributions, 369–96 ; relations with these, 382 ; distribution of the ratio of a random deviation from the sample mean to the sample standard deviation, (Exercise 16.20) 396 ; *see also* Bivariate normal distribution, Multivariate normal distribution.

Normalization of frequency functions, 163–7 ; of Gamma distribution, (Example 6.4) 166–7.

Order-statistics, 251, 325–46 ; distribution of, 251–3, 325 ; joint distribution of, (Exercise 11.7) 270, 325, 337–8, 325–46 ; moments of, in normal case, 326–8 ; tables of, in normal case, 329 ; asymptotic distributions, 329–36 and (Exercise 14.21) 346 ; upper bounds for expected values, 336–7 ; ranges and midranges, 338–43 ; tables of extreme-value distributions, 343 ; recurrence relations among distributions, 336 and (Exercise 14.5) 344 ; variances, (Exercise 14.6) 344.

Orthogonal polynomials, Tchebycheff-Hermite, 155–6 ; Laguerre and Jacobi, 163.

Orthogonal transformation, 249–50 ; Helmert's, (Example 11.3) 250–1 ; characteristic property of normal distributions, 365–6.

Owen, D. B., hypergeometric tables, 135 ; multivariate normal integral, 351.

Pairman, E., grouping corrections, 79.

Parameters, 32.

Pareto curve, and mean difference, (Exercise 2.19) 53–4.

Pattern functions, in determining sampling cumulants of k-statistics, 283–9 and (Exercises 12.4–5, 12.7) 305.

Paulson, E., normalization of variance-ratio distribution, 381–2.

Pearce, T. V., data from, (Table 1.22) 27.

Pearse, G. E., data from, (Table 1.11) 10.

Pearson, E. S., sampling distribution of $\sqrt{b_1}$, 297–8 ; sampling distribution of b_2, (Exercise 12.10) 305–6 ; distribution of range, 337 and 343 ; tables of distribution of range, 339 and 343 ; tables of χ^2, 370 ; distribution of ratio of a random deviation from the mean to the sample standard deviation, (Exercise 16.20) 396.

Pearson, K., data on trypanosomes, (Table 1.13) 11 ; data on fecundity of mares, (Table 1.19) 26 ; coefficient of variation, 47 ; grouping corrections, 79 ; measure of skewness, 85, 149 ; Tchebycheff-type

inequality, 88 ; data on whist deals, (Table 5.5) 135 ; moments of binomial distribution, (Exercise 5.5) 144 ; tetrachoric functions tabulation, 161 ; distribution of difference of Type III variates, (Exercise 11.15) 271–2 ; moments of normal order-statistics, 328 ; expansion of a variate in terms of its d.f., 328–9 ; χ^2 goodness-of-fit distribution, 356.

Pearson distributions, as limit of hypergeometric, 140–1 ; generally, 148–54 ; recurrence relations for moments, 149 ; skewness of, 149 ; criterion for type, 151 ; fitting of, 152–4 ; recurrence relations for c.f. and cumulants, (Exercise 6.7) 176 ; *see also* Types I–VII.

Permutations, random, (Example 9.7) 216.

Pitman, E. J. G., derivatives of c.f. and moments, 114 ; characteristic property of Gamma distributions, (Exercise 15.22) 368 ; moments of correlation coefficient in case of independence, (Exercise 6.17) 396.

Plackett, R. L., bound for expected value of range, 340 ; multivariate normal integral, 351.

Platykurtosis, 86.

Poisson distribution, frequency-generating function, (Exercise 1.13) 30 ; c.f. and cumulants, (Example 3.10) 72–3 and 125–6 ; moments, (Exercise 3.1) 89 ; difference of two Poisson variates, (Exercise 3.8) 90, (Example 4.5) 100–1 and (Exercise 11.16) 272 ; normal limiting form, (Example 4.9) 116 ; moments determining uniquely, (Exercise 4.4) 117 ; when parameter a random variable, (Exercise 4.16) 119, 129–33 and (Exercise 5.23) 147 ; generally, 125–9 ; d.f. and tables, 126 ; mixed distributions, 126–8 ; clustering of two Poissons producing Neyman Type A contagious distribution, (Exercise 5.7) 144–5 ; clustering producing Pólya-Aeppli distribution, (Exercise 5.8) 144–5 ; bivariate form, (Exercise 5.11) 145 ; clustering with logarithmic distribution to produce negative binomial distribution, (Exercise 5.21) 147 ; c.f.'s of clustering distributions, (Exercise 5.22) 147 ; factorial moments, (Exercise 5.26) 147 ; sum of independent Poisson variates a Poisson, (Exercise 7.14) 196 ; sampling of attributes from, (Exercise 9.5) 226 ; distribution of sample mean, (Example 11.11) 260 ; conditional, in obtaining χ^2 goodness-of-fit distribution, (Exercise 15.21) 368 ; distribution of time-intervals, (Exercise 16.6) 394.

Pólya, G., Pólya-Aeppli distribution, (Exercise 5.8) 144–5 ; inequality for normal integral, (Exercise 5.17) 146 ; normal d.f., (Exercise 5.18) 147.

Polygon, *see* Frequency polygon.

Polykays, multiple k-statistics, 303–4.

Populations, as basis of statistical theory, 1 ; existent, 19 ; hypothetical, 19–20 ; types in sampling, 206–8.

Posterior probability, 198–9.

Potatoes, bias in estimating yield, (Example 9.4) 210.

Power-sums, as symmetric functions, 276.

Pretorius, S. J., data on Australian marriages, (Table 1.8) 8 ; data on beans, (Table 1.15) 20 and (Table 6.1) 160 ; bivariate distributions, 174.

Principal value, in expected value, 59 footnote and (Exercise 7.19) 197.

Prior probability, 198–9.

Probability, 180–205 ; calculus of, 180–97 ; basic rules, 181–2 ; importance of reference-set, (Examples 7.1–2) 182–4 and (Exercises 7.6, 7.11) 195–6 ; in a continuum, 185–6 ; geometrical, (Example 7.4) 185–6 and (Exercises 7.9–11) 196 ; and statistical distributions, 186–7 ; prior and posterior, 198–9 ; *see also* Bayes' theorem, Likelihood.

Problem of moments, (Exercise 3.12) 91 ; generally, 109–12 and 115–16 ; (Exercises 4.13, 4.17) 118–19 ; and convergence of Type A series, (Exercise 6.9) 176.

Product, variance of, (Exercise 10.23) 245.

Proportions, sampling for, *see* Attributes.

Pseudo-random numbers, 218.

Quadratic forms in normal variates, 355–62 and (Exercises 15.14–15) 368 ; independence of, 357–9 ; independence of linear forms, 359 and (Exercise 15.13) 367 ; decomposition of sums of squares, 360–2.

Quantiles, 40 ; graphical determination of, 41–2 ; asymptotic normality and standard errors of, 236–9, 243 ; covariances of, 238–9 ; *see also* Extreme values, Order-statistics.

Quartiles, 40 ; interquartile range as measure of dispersion, 42 ; standard errors, 243.

Quenouille, M. H., combination of Poisson

and logarithmic distributions to produce negative binomial, (Exercise 5.21) 147 ; random deviates, 215.

Raff, M. S., error in normal approximation to binomial, 125.
Rand Corporation, table of 1,000,000 random digits, 214–15.
Random normal deviates, 215.
Random permutations, (Example 9.7) 216.
Random sampling, 206–19 ; technique of, 211–19.
Random Sampling Numbers, 213–19 ; tables of, 214–15.
Random variables, 186–90 ; sums of, 187–90, 192–5, 263–5 ; difference of, 188–90.
Randomizing machine, (Example 9.3) 209–10.
Randomness, 208–11.
Range, 42 ; exact distribution, 338–9 and (Exercise 14.19) 346 ; bounds for expected value, 339–40 ; asymptotic distribution, 340–1 and (Exercise 14.20) 346 ; reduced range, 342–3 ; existence of expected value, (Exercise 14.15) 345–6.
Rao, K. S., *see* Kendall, D. G.
Ratio, standard error of, 232 ; distribution of, 265–8 and (Exercises 11.11–13, 11.22) 270–3 ; moments of, 299–300.
Rectangular distribution, transformations to, 18–19, (Exercise 1.21) 31 ; variance, mean deviation and mean difference, (Exercise 2.5) 52 ; c.f. and cumulants, (Exercise 3.23) 93 ; moments of mean difference, 242 ; variance of median, (Exercise 10.15) 244 and (Example 14.2) 329 ; variance of mean deviation, (Exercise 10.19) 245 ; exact distribution of sample mean, (Example 11.9) 257–9 and (Example 11.14) 261 ; distribution of geometric mean, (Example 11.15) 262 ; distribution of product of variates, (Exercise 11.6) 270 and (Exercise 16.4) 394 ; of ratio of largest values, (Exercise 11.20) 272 ; of product of largest values, (Exercise 11.21) 272 ; distributions of order-statistics Beta distributions, 325 ; joint distribution of range and midrange, (Exercise 14.12) 345 ; distribution of midrange, (Exercise 14.13) 345 ; distribution of range, (Exercise 14.16) 346 ; distribution of ratio of midrange to range, (Exercise 14.17) 346 ; distribution of difference of two midranges, (Exercise 14.18) 346.
Rectangular distribution, discrete : c.f. and cumulants, (Exercise 3.23) 93.
Reduced range and midrange, 342–3.
Regression, linear, in bivariate normal distribution, 383 ; coefficients, 392 ; distribution of coefficients in bivariate normal samples, 392–3 ; in linear regression model, (Exercise 16.5) 394 ; standard errors of coefficients, (Exercise 16.13) 395.
Replacement, sampling with and without, 207 ; (Example 9.12) 224 ; (Exercises 9.8–9) 227.
Rider, P. R., distribution of product of largest values in samples from rectangular distributions, (Exercise 11.21) 272.
Romanovsky, V., moments of binomial, (Exercise 5.2) 143.
Ruben, H., moments of largest of a normal sample, 336 ; multivariate normal integral, 354.

Saddlepoint approximations, 269.
Sakamoto, H., characterization of normal distribution, 364.
Saltus, in d.f., 13, 97.
Sampford, M. R., Mills' ratio, (Exercise 5.16) 146.
Sampling, with and without replacement, 207 ; from hypothetical populations, 207–8 ; lottery or ticket, 212–13 ; from continuous populations, 217–19 ; *see also* Random sampling.
Sampling distribution, 191–2 ; role in sampling problems, 223 ; exact, 246–72 ; derived analytically, 246–53 ; derived geometrically, 253–9 ; derived by c.f., 259–63 ; of a sum derived by induction, 263–5 ; of a ratio, 265–8 ; multivariate, 268–9 ; approximations to, 274–324 ; *see under names of parent distributions.*
Sampling moments, *see* k-statistics.
Sandiford, P. J., hypergeometric approximated by binomial, 135.
Sandon, F., grouping corrections, 79.
Sarhan, A. E., table of variances of normal order-statistics, 329.
Scale-reading, bias in, (Example 9.2) 208–9.
Schwarz inequality, 38.
Second Limit theorem, 115–16.
Semi-interquartile range (s.i.q.), *see* Interquartile range.
Seminvariants, *see* Cumulants, k-statistics.
Sen, P. K., variances of order-statistics, (Exercise 14.6) 344.
Sentences, distribution of by length, (Table 1.17) 25.

Sequential sampling for attributes, (Example 9.13) 225–6.
Severo, N. C., bivariate normal integral, 350.
Shenton, L. R., continued fractions for normal distribution, 138 and (Exercises 5.14–15) 145–6.
Sheppard, W. F., tables of the normal distribution, 138; theorem on median dichotomy of bivariate normal distributions, 351.
Sheppard's corrections for grouping, 46; to standard deviation, (Exercise 2.8) 53; generally, 75–81, 84; (Exercise 3.17) 92; to standard deviation compared with sampling fluctuations, 234–5.
Sichel, H. S., frequency-moments, 174.
Signum function (sgn x), 95.
Simple random sampling, 206.
Skewness, 9; measures of, 85–6; (Exercise 3.22) 93; standard error of b_1, 234–43; sampling moments of $\sqrt{b_1}$, 297–8 and (Exercise 12.9) 305.
Slutsky, E. E., table of χ^2 distribution, 374.
Solomons, L. M., measure of skewness, (Exercise 3.22) 93.
Stabilization of variance by transformation, (Exercise 10.17) 245, (Exercises 16.18–19) 396.
Standard deviation, 43; calculation of, (Exercise 2.7) 43–5; relations with mean deviation and mean difference, 45 footnote and (Exercise 2.4) 52; and Sheppard's correction, (Exercise 2.8) 53, 234–5; standard error of, 243; *see also* Variance.
Standard errors, 221–4 and 228–45; table of, 243; compared with Sheppard's corrections, 234–5.
Standard measure, 47–8; effect on cumulants of transformation to, 68; effect on c.f., (Example 4.6) 106–7, (Example 4.9) 116.
Statistic, 2.
Statistical Abstract, data from, (Table 1.2) 3.
Statistical hypotheses, 202–3.
Statistical Review of England and Wales, data from, (Table 1.1) 3; (Table 1.3) 6; (Table 1.12) 11.
Statistical tables, distribution of digits from, 214.
Statistics, definition, 1–2.
Steck, G. P., Cauchy-distributed ratios, (Exercise 11.23) 273; trivariate normal integral, 351.
Steepest descents, method of approximation, 269, 342.
Stereogram, 21.
Stieltjes, J., problem of moments, (Exercise 3.12) 91; 109 footnote; 111.
Stieltjes integrals, 15–17.
Stirling's series, definition, 81.
Stouffer, S. A., on distribution of difference of Type III variates, (Exercise 11.15) 271–2.
Strong Law of Large Numbers, 193.
Stuart, A., dispersion matrix (Exercise 15.9) 367.
" Student " (W. S. Gosset), mixed binomials, (Exercise 5.6) 144; 375; table of t-distribution, 377; distribution of correlation coefficient, 387.
" Student's " distribution, c.f., (Example 3.13) 73–4; normal limit, (Example 4.8) 114–15; derived, (Example 11.8) 256–7 and (Example 11.19) 265; generally, 374–7; tendency to normality, c.f., moments, 375; d.f., 375–7 and (Exercise 16.10) 395; tables, 376–7; relations with normal, χ^2 and F-distributions, 382.
Subsample mean, standard error of difference from overall sample mean, (Exercise 10.20) 245.
Succession rule of Laplace, (Example 8.3) 200–1 and (Exercise 8.4) 205.
Suicides, data on, (Table 1.6) 7; and Poisson distribution, 128–9.
Sum of variates, distribution of, 188–90, 192–5, 263–5.
Sums of squares, decomposition of, 360–2 and (Exercise 15.9) 367.
Symmetric functions, tables of, 69 footnote, 276; generally, 274–81; (Exercise 12.19) 307.
Symmetrical distributions, 7–8; c.f. real, (Exercise 4.1) 117; zero correlation between sample mean and even moments about the mean in large samples, (Example 10.3) 231; zero correlation between any odd-order and even-order moments in large samples, (Exercise 10.6) 244; non-negative correlation in large samples between even sample moments, (Exercise 10.8) 244; zero correlation between odd-order and even-order k-statistics, (Example 12.9) 302–3 and (Exercise 12.13) 306; upper bounds for means of order-statistics, 337; distribution of reduced range and midrange for exponential-type parent, 342–3; bound for variance of median, (Exercise 14.7) 345.

Szász, O., conditions for c.f.'s, 102.
Szegö, G., *Orthogonal polynomials*, 163.

t-distribution, *see* " Student's " distribution.
Tabulation of frequency-distributions, 2–5.
Target, probability of hitting, (Exercise 15.7) 367.
Tchebycheff, P. L., *see* Bienaymé.
Tchebycheff-Hermite polynomials, 155–6.
Tchebycheff-type inequalities, 88–9, 192.
Teichroew, D., table of means of normal order-statistics, 329.
Telephone directory, distribution of digits from, (Table 1.4) 6, 213–14.
Term of abortion among women, data on, (Table 1.22) 27.
Tetrachoric functions, 160–1.
Thiele, T. N., cumulants, 68, 201.
Thompson, C., table of χ^2 percentage points, 371 ; table of percentage points of *t*-distribution, 376–7 ; table of percentage points of *F*-distribution, 379.
Ticket sampling, 212–13.
Tippett, L. H. C., tables of Random Sampling Numbers, 214 ; asymptotic distribution of extreme values, 330–3 ; exact distribution of largest of a normal sample, 336 ; distribution of range in normal samples, 339–343.
Tocher, J. F., data from, (Table 1.24) 28.
Transformations, 17–19, 23–24, 247–8 ; to stabilize variance, (Exercise 10.17) 245 ; orthogonal, 249–51, 365–6 ; leaving orders of magnitude of cumulants unaffected, (Exercise 13.6) 323 ; of correlation coefficient, *see* Correlation coefficient.
Truncated distributions, 10, 174.
Trypanosomes, data on, (Table 1.13) 11.
Tsetse flies, data on, (Table 1.13) 11.
Tukey, J. W., multiple *k*-statistics, 303.
Type A contagious distribution, (Exercise 5.7) 144–5.
Types A, B, C series, *see* Gram-Charlier series.
Type I (Pearson) distribution, grouping corrections, 77 ; generally, 150–1 ; fitted to cloudiness data, (Example 6.6) 171–2 ; *see also* Beta distribution of first kind.
Type II (Pearson) distribution, (Exercise 6.1) 175 ; distribution of mean, (Exercise 11.14) 271.
Type III (Pearson) distribution, c.f. and moments, (Example 3.6) 62 ; cumulants, (Example 3.12) 73 ; of Poisson parameter producing negative binomial, 129–30 and (Exercise 5.23) 147 ; generally, 152 ; κ criterion infinite, (Exercise 6.5) 175 ; distribution of mean, (Example 11.13) 260–1 ; distributions of geometric means, (Exercises 11.3, 11.9) 269–70 ; distribution of difference of, (Exercise 11.15) 271–2 ; *see also* Gamma distribution.
Type IV (Pearson) distribution, moments, (Exercise 3.3) 89 ; generally, 151–2 ; fitted to bean data, (Example 6.1) 153–4 ; tables of, 154.
Type V (Pearson) distribution, moments and cumulants, (Exercise 3.2) 89 ; κ criterion unity, (Exercise 6.6) 175.
Type VI (Pearson) distribution, 151 and (Exercise 6.4) 175 ; *see also* Beta distribution of second kind.
Type VII (Pearson) distribution, moments, (Example 3.3) 59 ; c.f., (Example 3.13) 73–4 ; (Exercise 6.1), 175 ; *see also* " Student's " distribution.

U-shaped distributions, 10.
Unbiassed estimates, 222, 225 ; (Exercises 9.13–14) 227.
Uniform distribution, *see* Rectangular distribution.
Unimodal distributions, 32.
Unit distribution, c.f., (Example 4.3) 99.
Unitary symmetric functions, 276.
Upholt, W. M., data from, (Table 1.5) 6.

Variable, and variate, 2 ; transformations, 17–19, 23–24, 247–8 ; *see also* Random variables.
Variance, 43 ; as half mean-square of differences, 47 ; standard error, 230, 243 ; exact mean and variance of its sampling distribution, (Exercise 10.13) 244, 275, (Example 12.1) 276 and (Example 12.3) 286–7 ; exact distribution in normal samples, (Example 11.7) 255–6 ; its sampling variance in samples from a finite population, (Exercise 12.11) 306 ; *see also* Standard deviation.
Variance-covariance matrix, *see* Dispersion matrix.
Variance-ratio distribution, (Example 11.20) 266–7 ; generally, 377–82 ; relation between *F*- and z-distributions, 378 ; shape of *F*-distribution, 378 ; c.f. of z-distribution, 378–9 ; tendency of z to normality, 379 ; d.f. of *F* and z, and tables

379 ; interchange of arguments of F or z, 379–80 ; approximations to distribution of z, 380–2 ; relations with normal, χ^2 and t-distributions, 382; moments of F, (Exercise 16.1) 393 ; distribution for correlated normal variates, (Exercise 16.6) 395.
Variance-stabilizing transformations, (Exercise 10.17) 245 and (Exercises 16.18–19) 396.
Variate, 2.
Variation, coefficient of, 47 ; standard error of, (Example 10.5) 233, 234, 243.
Von Mises, R., asymptotic distributions of extreme values, 332.

Wallace, D. L., asymptotic expansions, normalization of f.f., 163 footnote.
Watson, G. S., Jacobi polynomials in Beta distribution expansions, 163.
Weak Law of Large Numbers, 192–3 and (Exercise 7.20) 197.
Weight, distribution of men by, (Table 1.10) 9.
Weldon, W. F. R., dice-throwing data, (Table 1.14) 19 ; (Table 1.16) 25 ; (Table 5.1) 121 ; (Table 9.10) 221.
Wheat plants, distribution by height, (Table 9.1) 208.
Whist, distribution of cards at, (Table 5.5) 135.
Wicksell, S. D., lognormal distribution, 169.
Williams, C. B., logarithmic distribution, (Table 5.4) 132.
Williams, J. L., inequality for normal integral, (Exercise 15.5) 367.
Wilson E. B., approximation to χ^2 distribution, 371–4.
Wise, M., hypergeometric d.f., 135.
Wishart, J., cumulants of multivariate multinomial distributions, 143 ; tables of multiple k-statistics, 303 ; pattern functions, (Exercise 12.7) 305 ; variances and covariances of bivariate k-statistics, (Exercises 13.7–8) 323 ; combinatorial rules for k-statistics, (Exercises 13.10–12) 323–4.
Wold, H., Sheppard's corrections, 76, 80, 84 ; multivariate Carleman criterion for moments, 111 ; table of random normal deviates, 215.
Woo, T. L., data on skulls, (Table 1.18) 25.

χ^2 distribution, 369–74 ; shape, 370 ; c.f. and moments, tendency to normality, d.f., 370 and (Exercise 16.7) 394 ; tables of, 370–1, 374 ; approximations to, 371–4 and (Exercise 16.8) 394 ; relation with normal, t- and F-distributions, 382 ; recurrence relation for mean-moments, (Exercise 16.3) 394 ; *see also* Gamma distribution, Type III distribution.
χ^2 goodness-of-fit distribution, (Example 15.3) 355–6 and (Exercise 15.21) 368.
χ^2 variables, 360–2 and (Exercises 15.14–15) 368.

Yates, F., data on height of plants, (Example 9.1) 208 ; tables of Random Sampling Numbers, 214 ; table of χ^2 percentage points, 371 ; table of percentage points of t-distribution, 377 ; tables of percentage points of F- and z-distributions, 379.
Yields of grain, data, (Table 1.20) 26.
Yuan, P. T., lognormal distribution, 169.
Yule, G. U., data on sentence-length, (Table 1.17) 25 ; data on industrial accidents, (Table 5.3) 129 ; negative binomial distributions, 129–30 ; frequency-moments of normal distribution, (Exercise 5.10) 145 ; bias in scale-reading, (Example 9.2) 208–9 ; table of χ^2 for 1 degree of freedom, 371.

z-distribution, *see* Variance-ratio distribution.
z-transformation, *see* Correlation coefficient.
Zelen, M., bivariate normal integral, 350.