PUBLISHERS' NOTE

The series in which this title appears was introduced by the publishers in 1957 and is under the general editorship of Maurice G. Kendall, Sc.D., formerly Professor of Statistics in the University of London. It is intended to fill a need which has been evident for some time and is likely to grow: the need for some form of publication at moderate cost which will make accessible to a group of readers specialized studies in statistics or special courses on particular statistical topics. There are numerous cases where, for example, a monograph on some newly developed field would be very useful, but the subject has not reached the stage where a comprehensive book is possible; or, again, where a course of study is desired in a domain not covered by textbooks but where an exhaustive treatment, even if possible, would be expensive and perhaps too elaborate for the readers' needs.

Considerable attention has been given to the problem of producing these books speedily and economically. Appearing in a cover the design of which will be standard, the contents of each volume will follow a simple, straightforward layout, the text production method adopted being suited to the complexity or otherwise of the subject.

The publishers will be interested in approaches from any authors who have work of importance suitable for the series.

CHARLES GRIFFIN & CO. LTD.

GRIFFIN BOOKS ON STATISTICS, &c.

A statistical primer	F. N. DAVID
An introduction to the theory of statistics	G. U. YULE and M. G. KENDALL
The advanced theory of statistics (three volumes)	M. G. KENDALL and A. STUART
Rank correlation methods	M. G. KENDALL
Exercises in theoretical statistics	M. G. KENDALL
Rapid statistical calculations	M. H. QUENOUILLE
Combinatorial chance	F. N. DAVID and D. E. BARTON
Biomathematics	C. A. B. SMITH
The design and analysis of experiment	M. H. QUENOUILLE
Sampling methods for censuses and surveys	F. YATES
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

A COURSE IN THE GEOMETRY OF n DIMENSIONS

M. G. KENDALL, Sc.D.
formerly Professor of Statistics in the University of London
President, Royal Statistical Society, 1960–62

BEING NUMBER EIGHT OF
GRIFFIN'S STATISTICAL MONOGRAPHS & COURSES
EDITED BY
M. G. KENDALL, Sc.D.

CHARLES GRIFFIN & COMPANY LIMITED
LONDON

PRINTED IN GREAT BRITAIN BY JOHN WRIGHT & SONS LTD., AT THE STONEBRIDGE PRESS, BRISTOL

PREFACE

The geometry of n dimensions, as developed by pure mathematicians, is a somewhat recondite and difficult subject, most of which is remarkable more for its aesthetic appeal than for its utility. Certain branches of it, however, have an immediate application in statistics, partly in clarifying statistical ideas, partly in solving distributional problems. In fact, it is not easy to develop a comprehensive theory of statistics without introducing n-dimensional geometry at a fairly early stage.

In teaching statistics at the advanced level I have found that most students, even those with a good mathematical background, encounter serious difficulty with proofs depending on n-dimensional systems. The following course was given at the London School of Economics in 1960 to fill a gap in my students' knowledge. It does not purport to be a complete account of n-dimensional geometry or to replace the excellent little book by D. M. Y. Sommerville, first published in 1929 as *An Introduction to the Geometry of N Dimensions*. My object was to set out that part of the subject which had statistical applications and to sketch very briefly what those applications were. Since teachers in other parts of the world doubtless encounter similar difficulties, I decided to publish the lecture notes in the hope that they might be found generally useful.

Although I have no direct experience, I suspect that there are other fields of applied mathematics where the ideas of n-dimensional geometry are serviceable; and I hope, accordingly, that the first half of the book, at least, may help students and teachers outside the domain of theoretical statistics.

From the nature of the case, diagrams of n-dimensional situations are difficult, if not impossible, to present. In lecturing, however, I make a free use of two-dimensional drawings of two- and three-dimensional cases, and I would recommend the student to get into the habit of sketching these simpler cases for himself, so as to prepare himself for the visualization of the n-dimensional extensions.

My thanks are due to Mr. T. M. F. Smith and Mr. A. W. Matz, who read the typescript of the book and materially helped to remove obscurities and misprints. Doubtless some remain and I should be grateful to any reader who calls them to my attention.

M. G. K.

LONDON,
May, 1961.

CONTENTS

PART 1

THE GEOMETRY OF *n* DIMENSIONS

PART 2

STATISTICAL APPLICATIONS

PART 1

THE GEOMETRY OF *n* DIMENSIONS

Introduction

1 There are at least two ways in which a geometry may be developed. The first is familiar to anyone who has studied Euclid's *Elements* or its modern equivalent at school. It begins with certain undefined ideas such as "point" and "straight line"; it requires them to obey certain postulates; and it then proceeds to develop a series of propositions by purely deductive logic. As is well known, a remarkably large body of theorems can be deduced from comparatively few primitive ideas in this field.

2 The second approach, originated by Descartes, is to relate the concepts of geometry to the properties of numerical coordinates. A point in a plane is defined by a pair of numbers (x, y). A straight line is the locus of all points obeying a linear relation $lx+my+n = 0$; and so on. Coordinate geometry lacks the elegance and aesthetic appeal of the classical Greek system, but it is much more powerful and enables us to apply to the geometrical domain many of the results of numerical, algebraical and analytical mathematics.

3 The idea of "dimension" itself is, on the axiomatic approach, usually introduced intuitively. We all "know" what is meant

by one, two and three dimensions. In actual fact, we experience only three-dimensional objects, the two- and one-dimensional concepts being abstracts from observation. They do not cause us much difficulty. When we wish to proceed from three to more dimensions, however, we run up against a conceptual barrier so difficult that many people cannot surmount it. Just how far certain gifted individuals can visualize, say, a space of four dimensions is an arguable matter which it would be wise to leave to the psychologists. My own opinion is that nobody can. (The so-called four-dimensional continuum consisting of three spatial and one temporal dimension is quite a different thing and is not relevant here.) But these difficulties of "seeing the situation" do not prevent us from setting up a geometry of n dimensions in the classical sense and of reasoning about it.*

4 In the following treatment we shall rely mainly on the more modern approach through coordinate geometry. A point in n dimensions is defined as an ordered set of quantities $(x_1, x_2, \ldots, x_n)$ and there is no real difficulty in making n as large as we please. The geometry we shall be concerned with is then equivalent to the mathematics of these ordered n-uples, which we can also, if we wish, interpret as vectors (usually column-vectors). From this viewpoint our "geometry" becomes a branch of numerical mathematics couched in a particular language, and it is always possible to express our results in a non-geometrical form. For example, in two dimensions two straight lines always meet in a single point except when they are parallel. This is equivalent to saying that the two equations

$$l_1 x + l_2 y + l_3 = 0$$

$$m_1 x + m_2 y + m_3 = 0$$

* The reader who is interested in the axiomatic approach will find an account of it in Sommerville's book (1958) (see page 63).

have a unique solution unless the determinant

$$\begin{vmatrix} l_1 & l_2 \\ m_1 & m_2 \end{vmatrix} = l_1 m_2 - l_2 m_1 = 0.$$

5 What, then, is the point of having a geometry of n dimensions? Why not express all our results in terms of algebraic quantities which make no demands on our powers of visualization? There are, I suggest, three reasons:

(1) The language of geometry is much simpler and more elegant than the language of algebra. To say that in three dimensions a plane cuts a sphere in a circle is very much simpler, and, in a sense, more immediately informative, than stating the equivalent in terms of coordinates, which would have to go something like this: the surface

$$(x-a)^2+(y-b)^2+(z-c)^2 = r^2$$

determines on the plane

$$l_1 x + l_2 y + l_3 z + l_4 = 0$$

a curve which, with a suitable change of coordinates, can be expressed in the form

$$(\xi-\alpha)^2+(\eta-\beta)^2 = \rho^2.$$

(2) It is possible to carry through quite rigorous trains of reasoning in geometrical terms without translating them into algebra. This gives us considerable economy both in thought and in communication of thought. We shall meet many examples in the sequel. One reason for this possibility is that, although we are working with a coordinate system, most of the results we need are independent of the system and are therefore invariant under certain classes of coordinate transformation. The language of geometry embodies this element of invariance.

(3) Most important, perhaps, is the fact that our geometrical imagery in two or three dimensions suggests results for more

dimensions and offers us a powerful tool of inductive or creative reasoning. For example, in three dimensions two spheres intersect in a circle (if they intersect at all). This immediately suggests that in n dimensions two hyperspheres of $n-1$ dimensions intersect in a hypersphere of $n-2$ dimensions, a proposition which is easy to verify. Some of the results which we shall encounter later were suggested by this kind of analogical argument. Even persons who refuse to recognize its rigour must acknowledge its creative possibilities. Some of the results arrived at in this manner are, indeed, exceedingly difficult to establish by analytical methods.

6 Apart from the simpler undefinables of classical geometry such as point, line and curve, we shall require certain other notions, of which the chief are distance, angle between two lines, and volume (or content). These can be defined analytically, e.g. the distance between two points (x_1, x_2, x_3) and (y_1, y_2, y_3) in three dimensions may be defined as the positive square root of

$$(x_1-y_1)^2+(x_2-y_2)^2+(x_3-y_3)^2.$$

In geometries where distance is defined differently (for example, in Riemannian geometries where spaces are "curved" and a metric is set up in terms of a quadratic differential element) our intuitive ideas may need severe conditioning. But such matters will not concern us here. If we have to deal with curved surfaces they will be immersed in flat or Euclidean spaces, just as the surface of an ordinary sphere is immersed in the ordinary three dimensions of experience. We shall find that in n dimensions our ideas of angle and volume are fairly straightforward generalizations of the three-dimensional concepts. Straight lines are the shortest distances between points, parallels exist, and Pythagoras' theorem holds almost by definition.

Reality

7 Questions of reality, however, sometimes need attention. In axiomatic geometries such as the two-dimensional geometry of the schoolroom we are concerned almost entirely with real points. Thus, for example, we have to say that a straight line meets a circle either in two points, or in one point (as a tangent), or no points. From the analytical viewpoint we can put this more simply at the expense of admitting complex numbers: a linear and a quadratic form in two variables with real coefficients have two values in common, both real, both real and coincident, both imaginary. Attractive as this may be from the point of view of generality, it is a resource which is not always open to us in the applications of n-dimensional geometry which we are going to consider. In integrating over a domain determined by certain boundaries, for example, it is vital to know how those boundaries intersect in the real domain.

Varieties

8 Consider, then, the aggregate of points typified by $(x_1, x_2, \ldots, x_n)$, where the x's can take any real values from $-\infty$ to ∞. We shall refer to this space as an S_n.

Any equation in the variables x determines a subspace in S_n. This is called a *variety*. For the most part we shall be concerned with varieties determined by rational integral algebraic equations. If the degree of such an equation is r we say that the *order* of the variety is r and write it as V^r_{n-1}. It is of $n-1$ dimensions because $n-1$ coordinates are sufficient to fix a point on it. Likewise if we have a set of p equations in $x_1, x_2, \ldots, x_n$ they define a V^h_{n-p}, where h is its order as defined in section 13 below.

Intersections of varieties

9 Two varieties in an S_n may or may not have points in common. For example, a line and a circle in three dimensions

may not meet; a circle and a sphere in three dimensions will meet (if at all in the real plane) in two points unless the circle lies entirely on the sphere; and so on. Generally, if a V_{n-p} and V_{n-q} have points in common, they are said to intersect. By the definition of variety, this intersection is a variety. In general, a V_{n-p} and a V_{n-q} intersect in a $V_{n-(p+q)}$. For the first is determined by p equations and the second by q, and hence their common set of values by $p+q$ equations. If $p+q=n$ the varieties will intersect only in points. If $p+q<n$ they will not, in general, intersect at all.

By extension, a V_{n-p}, V_{n-q} and V_{n-r} intersect in a $V_{n-p-q-r}$. This is easily proved by induction, and the extension to more varieties is immediate.

10 To avoid prolixity we may, at this point, set aside from the main discussion certain degenerate situations. If, of the p equations determining a V_{n-p} in S_n, some are dependent on the others (i.e. can be expressed as a function of them) we have fewer than p *independent* conditions, say p', and the variety is of $n-p'$ dimensions. Likewise if p independent and q independent equations are not independent between themselves, but are together equivalent to $t<p+q$ independent equations, they determine a variety of dimension $n-t$. On occasion we shall have to consider these degeneracies *ad hoc*, but for the most part we shall, unless the contrary is stated, assume that they do not exist or that they have been removed from the situation by prior scrutiny.

11 Linear spaces are of particular importance. A variety of order 1 is called a hyperplane, a prime, or a *flat*. If the p equations in the $(x_1, \ldots, x_n)$ are linear they define an $(n-p)$-flat, which is an S_{n-p}. A p-flat, then, is a flat (Euclidean) space of p dimensions.

It follows from the definition that the intersection of two flats is always flat.

An S_1 is called a (straight) line. In n dimensions $(n-1)$ equations are required to define it. As a matter of convenience in symmetry, however, we sometimes write it in the form (X relating to current coordinates, x to a point on the line)

$$\frac{X_1-x_1}{l_1}=\frac{X_2-x_2}{l_2}=\frac{X_3-x_3}{l_3}=\ldots\frac{X_n-x_n}{l_n} \tag{1}$$

which comprises $(n-1)$ independent equations. We also find it convenient to represent the line with a parameter ρ by putting

$$\begin{aligned} X_1 &= x_1+\rho l_1 \\ X_2 &= x_2+\rho l_2 \\ &\text{etc.} \end{aligned} \tag{2}$$

12 An S_2 is called a plane. Except in three dimensions it is not so easy to write in a symmetrical form similar to equation (1), but we can write it in parametric form

$$\begin{aligned} X_1 &= x_1+\rho l_1+\sigma m_1 \\ X_2 &= x_2+\rho l_2+\sigma m_2 \\ &\text{etc.} \end{aligned} \tag{3}$$

Clearly if we eliminate ρ and σ from the n equations (3) we have $n-2$ linear equations in the current coordinates X which define an S_2. Conversely, given $n-2$ linear equations we can reduce them to the form (3).

Example 1

In S_3 two planes intersect in an $S_{3-1-1}=S_1$, namely a line. Two lines intersect in an $S_{3-2-2}=S_{-1}$, i.e. do not intersect in general. In four dimensions, S_4, two planes S_2 intersect in an $S_{4-2-2}=S_0$, i.e. in a point. In S_5 they do not intersect in general.

In S_5 the three spaces S_4, S_3, S_3 intersect in an $S_{5-1-2-2}=S_0$, i.e. in a point.

Example 2

A straight line in S_n intersects a V_{n-1}^r in r points (some of which may be coincident or imaginary). For the V_{n-1}^r is defined by one equation in the x's of degree r. If we substitute from equations (2) we have a polynomial of degree r in ρ, with r roots.

13 The order of a variety V_p is most conveniently defined as the number of points in which it intersects an arbitrary S_{n-p}. This is consistent with our definition of the order of a V_{n-1}.

A V_p^r and a V_{n-p}^s in an S_n will, in general, intersect in rs points. Consider, in fact, the V_p^r defined by $n-p$ equations, say of degree $\alpha_1, \alpha_2, \ldots, \alpha_{n-p}$. An arbitrary S_{n-p} can be expressed, as in equation (3), in terms of $n-p$ parameters. When we substitute for the x's and eliminate all but one of these parameters we shall reach an equation in the remaining parameter of degree $\alpha_1 \alpha_2 \ldots \alpha_{n-p}$. The original set of $n-p$ equations are thus equivalent to one of such degree which is the order r. Likewise there is one equation of order s for V_{n-p}. These two have, in general, rs solutions.

Interpretation over matters of reality is nevertheless to be carried out with care. In two dimensions two conics (V_1^2) will intersect in four points. But two circles will intersect only in two real points at most, the other two being the imaginary "circular points" at infinity.

Projection

14 In S_3 we can project a curve on to a plane by selecting a point, joining it by straight lines to all the points of the curve so as to obtain a cone, and observing the curve of intersection of the cone and the plane of projection. This is *conical* projection. If the apex of the cone is at infinity all the lines of projection are parallel. In particular, if they are at right angles to the plane of projection we have *orthogonal* projection.

The process can be carried out in n dimensions. Given a V_1 we can select a point O and construct the family of lines through

O and V_1. These lines will trace out another V_1 by intersection with a given S_{n-1}. We cannot, of course, project a V_1 by these means on to a flat of dimension less than $n-1$. We could construct more general methods of projection if necessary, but they are rarely required.

Simple figures in n dimensions

15 The generalization of a polygon in two dimensions is a polytope. It is the figure bounded by a set of $(n-1)$-flats. These themselves intersect, adjacent $(n-1)$-flats meeting in an $(n-2)$-flat and so on. n faces meet in a point or vertex. The flat of r dimensions is called an r-boundary. Thus, in three dimensions the figure is bounded by planes; these meet in lines, the edges; and these meet in points, the corners or vertices.

16 The least number of $(n-1)$-flats which can enclose a space and form a polytope is $n+1$. Thus in two dimensions we have the triangle and in three dimensions the tetrahedron. Such a figure is called a *simplex*.

Example 3

The triangle has 3 sides, 3 vertices.

The tetrahedron has 4 faces, 6 sides, 4 vertices.

Consider the simplex in 4 dimensions. This has 5 S_3's as "faces". How many 2-flats, 1-flats and 0-flats has it?

The answer is 10, 10, 5. The 3-flats meet in pairs in $\binom{5}{2} = 10$ ways to form the 2-flats. They meet in triplets in $\binom{5}{3} = 10$ ways to form the 1-flats; and in sets of four in $\binom{5}{4} = 5$ ways to form the vertices.

The general law for n dimensions will now be clear. The number of flats of dimension $n-1, n-2, \ldots$ are the successive terms in the binomial $(1+1)^{n+1}$, omitting the first and last terms, which are each equal to unity.

17 A *parallelotope* is the n-dimensional analogue of the parallelogram and is bounded by pairs of parallel $(n-1)$-flats.

Anticipating a little our treatment of angle and distance, we may define an *orthotope* as the analogue of a rectangle, in which bounding $(n-1)$-flats are perpendicular, and a *hypercube* as an orthotope in which the parallel bounding $(n-1)$-flats are all the same distance apart.

Example 4

Consider a parallelogram in two dimensions. Take a third dimension and a line in it through one of the corners of the parallelogram. Consider the parallelogram as displaced with its corner moving along this line, remaining parallel to its original plane. It will then trace out a three-dimensional parallelogram or parallelopiped. Likewise, a parallelotope in n dimensions can be regarded as generated by an $(n-1)$-dimensional parallelotope moving parallel to itself along a straight line making an angle θ with the $(n-1)$ parallelotope. Let N_r be the number of r-boundaries of the n-dimensional parallelotope and N_r' the number of r-boundaries of the $(n-1)$-dimensional parallelotope. Consider the generating function

$$N_0 + N_1 t + \ldots + N_n t^n, \tag{4}$$

where $N_0 = 1$. The N_r boundaries are produced by the N_{r-1}' boundaries of the $(n-1)$-dimensional parallelotope together with its N_r' boundaries of r dimensions in their initial and final positions. Hence

$$N_r = N_{r-1}' + 2N_r'.$$

Thus

$$\sum_{j=0}^{n} N_j t^j = \left(\sum_{j=0}^{n} N_j' t^j\right)(2+t),$$

whence, by induction,

$$\Sigma N_j t^j = (2+t)^n. \tag{5}$$

This arrays the r-boundaries of the parallelotope and we have

$$N_r = 2^{n-r}\binom{n}{r}. \tag{6}$$

For example, a hypercube in four dimensions has

8 three-dimensional (cubic) "hyperfaces"
24 two-dimensional (square) "faces"
32 one-dimensional "edges"
16 vertices.

18 If Q is a quadratic form in the variables $x_1, \ldots, x_n$ (which may also include linear terms), the V_{n-1} defined by

$$Q = \text{constant}$$

is a *hyperquadric*. As we shall see later, a transformation to a new origin can be made to eliminate the linear terms, and with such an origin our most general hyperquadric may be written

$$\sum_{j,k=1}^{n} a_{jk}\, x_j\, x_k = c. \tag{7}$$

If $c = 0$ we have a *hypercone* with vertex at the origin. In the special case

$$\sum_{1}^{n} x_j^2 = c \tag{8}$$

we have a *hypersphere*.

Coordinate transformations

19 Consider a transformation

$$y_i = \sum_{j=1}^{n} l_{ij}\, x_j + a_i, \qquad i = 1, 2, \ldots, n. \tag{9}$$

This can be regarded as a displacement of the origin, represented by a_i, and a "rotation", represented by the coefficients l_{ij}. It is usually convenient to consider these separately.

Of particular importance is the so-called "orthogonal" transformation

$$y_i = \sum_{j=1}^{n} l_{ij}\, x_j \tag{10}$$

where

$$\sum_{j=1}^{n} l_{ij} l_{kj} = \delta_{ik}$$
$$= 1, \quad i = k$$
$$= 0, \quad i \neq k. \tag{11}$$

Coefficients l can always be chosen so as to obey (11). In fact, the conditions impose $p+\frac{1}{2}p(p-1) = \frac{1}{2}p(p+1)$ constraints on p^2 constants, leaving us $\frac{1}{2}p(p-1)$ further conditions at disposal. We may say that the transformation has $\frac{1}{2}p(p-1)$ degrees of freedom.

Writing $\mathbf{L}$ for the matrix (l_{ij}) and $\mathbf{L}'$ for its transpose we see that (11) is equivalent to

$$\mathbf{LL}' = \mathbf{I}.$$

Thus

$$|L||L'| = 1,$$

and since the determinants of the matrix and its transpose are equal, we have

$$|L|^2 = 1. \tag{12}$$

We shall usually take the determinant as $+1$. (The negative sign corresponds to a "left-handed" transposition and does not affect the properties with which we are concerned.)

Note also that

$$\mathbf{Y} = \mathbf{LX},$$

and hence

$$\mathbf{L'Y} = \mathbf{L'LX} = \mathbf{X}. \tag{13}$$

Thus the transformation is bi-orthogonal.*

Helmert's transformation

20 One particular transformation is of special interest in

* At this stage "orthogonal" may be regarded as a convenient term to describe this type of transformation. From the definition of angle in section **23** it follows that it has the usual properties of rectangularity.

statistics. We put

$$y_1 = \frac{1}{\surd 2}(x_1 - x_2)$$

$$y_2 = \frac{1}{\surd 6}(x_1 + x_2 - 2x_3)$$

$$y_3 = \frac{1}{\surd 12}(x_1 + x_2 + x_3 - 3x_4)$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$

$$y_{n-1} = \frac{1}{\surd\{n(n-1)\}}\{x_1 + x_2 + \ldots + x_{n-1} - (n-1)x_n\}$$

$$y_n = \frac{1}{\surd n}(x_1 + x_2 + \ldots + x_n). \tag{14}$$

The reader should verify that this transformation obeys the conditions (11).

To reverse the transformation we use the last two equations of (14) to obtain x_n. Then the last three to obtain x_{n-1} and so on. We may also use the property expressed in (13) and, reading the columns downward, write down at once

$$x_1 = \frac{1}{\surd 2}y_1 + \frac{1}{\surd 6}y_2 + \ldots + \frac{1}{\surd\{n(n-1)\}}y_{n-1} + \frac{1}{\surd n}y_n$$

$$x_2 = -\frac{1}{\surd 2}y_1 + \frac{1}{\surd 6}y_2 + \ldots + \frac{1}{\surd\{n(n-1)\}}y_{n-1} + \frac{1}{\surd n}y_n$$

$$x_3 = -\frac{2}{\surd 6}y_2 + \ldots + \frac{1}{\surd\{n(n-1)\}}y_{n-1} + \frac{1}{\surd n}y_n$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$

$$x_{n-1} = -\frac{n-2}{\surd\{(n-1)(n-2)\}}y_{n-2} + \frac{1}{\surd\{n(n-1)\}}y_{n-1} + \frac{1}{\surd n}y_n$$

$$x_n = -\frac{n-1}{\surd\{n(n-1)\}}y_{n-1} + \frac{1}{\surd n}y_n. \tag{15}$$

From the last equation in (14), or adding the columns in (15), we see that the mean of the x's is given by

$$\bar{x} = \frac{1}{\sqrt{n}} y_n. \tag{16}$$

Hence the variables

$$x_1 - \bar{x}, x_2 - \bar{x}, \ldots, x_{n-1} - \bar{x}$$

depend only on $y_1, y_2, \ldots, y_{n-1}$.

Example 5

Consider the quadratic form $\sum_{i=1}^{n} x_i^2$. We have the simple identity

$$\Sigma x_i^2 = \Sigma (x_i - \bar{x})^2 + n\bar{x}^2.$$

If we then transform to y's by (14), in virtue of the remark just made,

$$\Sigma x_i^2 = (\text{quadratic form in } y_1, y_2, \ldots, y_{n-1}) + y_n^2.$$

This is a well-known result in theoretical statistics. If n independent variables x are distributed in the normal form with zero mean and unit variance, their joint frequency is given by

$$dF = \frac{1}{(2\pi)^{\frac{1}{2}n}} \exp\{-\tfrac{1}{2}\Sigma x^2\}\, dx_1 \ldots dx_n. \tag{17}$$

When we transform to y's (the Jacobian being unity) we have

$$dF \propto \exp(\text{quadratic in } y_1, \ldots, y_{n-1})\, dy_1 \ldots dy_{n-1} \exp(-\tfrac{1}{2}y_n^2)\, dy_n,$$

and hence y_n, the mean of x's, is independent of $\Sigma(x_i - \bar{x})^2$, and therefore of the variance of x.

Furthermore, in virtue of the orthogonality of the transformation,

$$\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i^2$$

and thus

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n-1} y_i^2.$$

Hence the sum $\Sigma(x_i - \bar{x})^2$ is distributed as the sum of $n-1$ variables $y_1, \ldots, y_{n-1}$, which also are independent with zero mean and unit variance.

Example 6

Consider the bivariate form, similar to that of the previous example

$$\sum_{i=1}^{n} x_i^2 - 2\rho \sum_{i=1}^{n} x_i u_i + \sum_{i=1}^{n} u_i^2$$
$$= \Sigma(x-\bar{x})^2 - 2\rho\Sigma(x-\bar{x})(u-\bar{u}) + \Sigma(u-\bar{u})^2 + n\{\bar{x}^2 - 2\rho\bar{x}\bar{u} + \bar{u}^2\}. \tag{18}$$

Transforming by a Helmert transformation from x to y and from u to v, we see that this is equal to the sum of two components, a quadratic form in $y_1, \ldots, y_{n-1}, v_1, \ldots, v_{n-1}$ and a quadratic in y_n and v_n (or $\bar{x}$ and $\bar{u}$).

Thus, in samples of n from the bivariate normal form

$$dF \propto \exp\left\{-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xu + u^2)\right\} dx\, du$$

the mean statistics $\bar{x}, \bar{u}$ are distributed independently of a quadratic form in $x_i - \bar{x}$ and $u_i - \bar{u}$.

Polar coordinates in n dimensions

21 In two dimensions we have the familiar polar transformation

$$x_1 = r\cos\theta, \quad x_2 = r\sin\theta, \qquad 0 \leqslant r \leqslant \infty, \quad 0 \leqslant \theta \leqslant 2\pi \tag{19}$$

with

$$x_1^2 + x_2^2 = r^2 \tag{20}$$

and a Jacobian

$$J = \frac{\partial(x_1, x_2)}{\partial(r, \theta)} = \begin{vmatrix} \cos\theta & \sin\theta \\ -r\sin\theta & r\cos\theta \end{vmatrix} = r. \tag{21}$$

In three dimensions, in so-called spherical polars, we have

$$x_1 = r\cos\theta_1\cos\theta_2, \quad x_2 = r\cos\theta_1\sin\theta_2, \quad x_3 = r\sin\theta_1,$$

$$0 \leqslant r \leqslant \infty, \quad -\tfrac{1}{2}\pi \leqslant \theta_1 \leqslant \tfrac{1}{2}\pi, \quad 0 \leqslant \theta_2 \leqslant 2\pi, \tag{22}$$

with

$$\sum_{i=1}^{3} x_i^2 = r^2,$$

and (writing c for cos and s for sin)

$$J = \frac{\partial(x_1, x_2, x_3)}{\partial(r, \theta_1, \theta_2)} = \begin{vmatrix} c_1 c_2 & c_1 s_2 & s_1 \\ -r s_1 c_2 & -r s_1 s_2 & r c_1 \\ -r c_1 s_2 & r c_1 c_2 & 0 \end{vmatrix} = r^2 \cos\theta_1. \tag{23}$$

Now consider

$$\begin{aligned}
x_1 &= r c_1 c_2 \quad \cdots \quad c_{n-2} c_{n-1}, & 0 \leqslant \theta_{n-1} \leqslant 2\pi, \\
x_2 &= r c_1 c_2 \quad \cdots \quad c_{n-2} s_{n-1}, & -\tfrac{1}{2}\pi \leqslant \theta_{n-2} \leqslant \tfrac{1}{2}\pi, \\
x_3 &= r c_1 c_2 \quad \cdots \quad s_{n-2}, & -\tfrac{1}{2}\pi \leqslant \theta_{n-3} \leqslant \tfrac{1}{2}\pi, \\
\cdots & \quad \cdots \quad \cdots \quad \cdots & \\
x_j &= r c_1 \quad \cdots \quad c_{n-j} s_{n-j+1}, & \\
\cdots & \quad \cdots \quad \cdots \quad \cdots & \\
x_n &= r s_1, & -\tfrac{1}{2}\pi \leqslant \theta_1 \leqslant \tfrac{1}{2}\pi.
\end{aligned} \tag{24}$$

Note that all the angles go from $-\frac{1}{2}\pi$ to $\frac{1}{2}\pi$ except θ_{n-1}, which varies from 0 to 2π. It is easy to see that

$$\sum_{i=1}^{n} x_i^2 = r^2. \tag{25}$$

The Jacobian is

$$J = r^{n-1} \begin{vmatrix} c_1 & \dots & c_{n-1} & c_1 & \dots & c_{n-2}s_{n-1} & \dots & s_1 \\ -s_1c_2 & \dots & c_{n-1} & -s_1 & \dots & c_{n-2}s_{n-1} & \dots & c_1 \\ -c_1s_2 & \dots & c_{n-1} & -c_1s_2 & \dots & c_{n-2}s_{n-1} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ -c_1c_2 & \dots & s_{n-1} & c_1c_2 & \dots & c_{n-1} & \dots & 0 \end{vmatrix}$$

$$= r^{n-1} c_1^{n-1} c_2^{n-2} \dots c_{n-1} s_2 \dots s_{n-1} \begin{vmatrix} 1 & 1 & 1 & & 1 \\ -t_1 & -t_1 & -t_1 & \dots & 1/t_1 \\ -t_2 & -t_2 & -t_2 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ -t_{n-1} & 1/t_{n-1} & 0 & \dots & 0 \end{vmatrix}$$

where t stands for tangent. Subtracting each column from the preceding one, we find, with a little manipulation,

$$J = r^{n-1} c_1^{n-2} c_2^{n-3} \dots c_{n-2}. \tag{26}$$

It is readily verified that for $n = 2$ and $n = 3$ these results agree with those obtained directly, e.g. in (21) and (23).

Geometrically, we may picture the situation in this way: consider the space with an origin and hyperspheres of constant r centred on it like the layers of an onion. (The surface of constant r is then a V_{n-1}^2.) The flat x_n = constant corresponds to s_1 = constant or θ_1 = constant. The space for which this is true is given by

$$\sum_{i=1}^{n-1} x_i^2 = r^2 \cos^2 \theta_1$$

which is a V_{n-2}, another hypersphere, in the S_{n-1} determined by x_n = constant. Similarly in this S_{n-1} the flat x_{n-1} = constant corresponds to θ_2 = constant and determines a hypersphere V_{n-3} in an S_{n-2}; and so on with diminishing dimension number until we reach a circle in a plane.

Example 7

Consider again samples of n from a normal distribution with zero mean and unit variance. The distribution becomes

$$dF \propto \exp\left(-\tfrac{1}{2}\sum_{i=1}^{n} x_i^2\right) dx_1 \dots dx_n$$

$$\propto \exp(-\tfrac{1}{2}r^2)\, r^{n-1}\, dr\, c_1^{n-2}\, c_2^{n-3} \dots c_{n-2}\, d\theta_1 \dots d\theta_{n-1}. \quad (27)$$

Thus the distribution of r is independent of the distribution of the angles $\theta_1, \dots, \theta_{n-1}$ and is given by

$$dF \propto \exp(-\tfrac{1}{2}r^2)\, r^{n-1}\, dr. \quad (28)$$

This is the familiar distribution of χ^2.

The independence of r and the angles θ has one other far-reaching consequence. If we have two algebraic forms homogeneous and of the same degree in the x's, say f_1 and f_2, their ratio

$$t = f_1/f_2 \quad (29)$$

is of degree zero in x. Thus, in virtue of (27), the expectation of any power of t is independent of r. Consequently, if f_2 is a function of r, the moments of t (and hence the distribution of t) are independent of r. Thus, for normal variation, with t and f_2 independent,

$$E(f_1^k) = E(t^k f_2^k) = E(t^k)\, E(f_2^k)$$

and hence

$$E(t^k) = \frac{E(f_1^k)}{E(f_2^k)}, \quad (30)$$

a formula which enables us to calculate the moments of the ratio t from those of its numerator and denominator.

Equation of a flat through given points

22 In n dimensions the general equation of an $(n-1)$-flat is

$$a_1 x_1 + a_2 x_2 + \dots + a_n x_n = k. \quad (31)$$

If this passes through n points with coordinates $x_{1i}, x_{2i}, \dots, x_{ni}$, $i = 1, 2, \dots, n$, we have n equations typified by

$$a_1 x_{1i} + a_2 x_{2i} + \dots + a_n x_{ni} = k, \qquad i = 1, 2, \dots, n. \qquad (32)$$

Eliminating the $n+1$ constants $a_1, a_2, \dots, a_n, k$ from the $(n+1)$ equations (31) and (32), we have for the equation of the $(n-1)$-flat

$$\begin{vmatrix} x_1 & x_2 & \dots & x_n & 1 \\ x_{11} & x_{21} & \dots & x_{n1} & 1 \\ x_{12} & x_{22} & \dots & x_{n2} & 1 \\ \dots & \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nn} & 1 \end{vmatrix} = 0. \qquad (33)$$

If the matrix (x_{ij}) is not of rank n the points are not independent and the $(n-1)$-flat is not uniquely determined. The generalization to n dimensions of results which are familiar in two and three will offer no difficulty. For example, with $n = 3$, if the matrix (x_{ij}) is of rank 2 the three points lie on a line and a single infinity of planes will contain them.

Angles

23 Any two intersecting straight lines, in however many dimensions, determine a plane which contains them both. The angle between them in this plane is unique. Likewise, if two lines do not intersect, we can select a point on one and draw through it a parallel to the other. The angle between the two intersecting lines is independent of where we choose the point or which line we place it on, and hence determines the (unique) angle between the lines.

Consider now a line in n dimensions given by equation (1):

$$\frac{X_1 - x_1}{l_1} = \frac{X_2 - x_2}{l_2} = \dots = \frac{X_n - x_n}{l_n}.$$

This goes through the point $(x_1, x_2, \ldots, x_n)$. If we draw a parallel through the origin O we shall have the line

$$\frac{X_1}{l_1} = \frac{X_2}{l_2} = \ldots = \frac{X_n}{l_n}. \tag{34}$$

The constants l_i are indeterminate in the sense that the line remains the same if we multiply them all by the same constant. We may, therefore, without loss of generality, require them to obey the normalizing condition

$$\sum_{i=1}^{n} l_i^2 = 1. \tag{35}$$

Now consider a plane $X_1 = a_1$. This will meet the line in a point P, say, with coordinates $a_1, l_2 a_1/l_1, l_3 a_1/l_1$, etc. The distance of this point from the origin is given by

$$\begin{aligned} OP^2 &= a_1^2 + \left(\frac{l_2 a_1}{l_1}\right)^2 + \ldots + \left(\frac{l_n a_1}{l_1}\right)^2 \\ &= \frac{a_1^2}{l_1^2}, \quad \text{in virtue of (35).} \end{aligned}$$

Thus

$$l_1 = \frac{a_1}{OP}.$$

Now a_1/OP is the cosine of the angle between the line and the X_1-axis. Similarly l_i is the cosine of the angle between the line and the X_i-axis. The constants l_i are therefore known as *direction cosines*.

It follows that the orthogonal projection of a point $(x_1, \ldots, x_n)$ on to the line (34) is distant $\Sigma l_i x_i$ from the origin. Its distance from the line is then given by

$$\begin{aligned} &\Sigma x_i^2 - (\Sigma l_i x_i)^2 \\ &\quad = \Sigma l_i^2 \Sigma x_i^2 - (\Sigma l_i x_i)^2 \\ &\quad = \tfrac{1}{2} \sum_{i,j} (l_i x_j - l_j x_i)^2. \end{aligned}$$

24 Consider now a second line with direction cosines l'_i. The projection of OP on to this line is $OP\cos\phi$ where ϕ is the angle between the lines. The projection of the line represented by l on the X_i axis is OPl_i and the projection of OP on l' is then the sum of terms $OPl_i l'_i$. Thus

$$\cos\phi = \sum_{i=1}^{n} l_i l'_i. \tag{36}$$

This is a fundamental formula which is a direct extension of the results for two and three dimensions. The lines are orthogonal if $\Sigma ll' = 0$. We may also write

$$\begin{aligned}\sin^2\phi &= 1-(\Sigma ll')^2 = \Sigma l^2 \Sigma l'^2 - (\Sigma ll')^2\\ &= \tfrac{1}{2}\Sigma(l_i l'_j - l_j l'_i)^2, \qquad i,j = 1,2,\ldots,n.\end{aligned} \tag{37}$$

25 Consider now an $(n-1)$-flat which, without loss of generality, we may suppose to go through the origin. It has, say, the equation

$$\sum_{i=1}^{n} a_i X_i = 0. \tag{38}$$

A line, also through the origin, of the form

$$\frac{X_1}{l_1} = \ldots = \frac{X_n}{l_n} = \rho \tag{39}$$

will lie in the $(n-1)$-flat if and only if

$$\Sigma a_i \rho l_i = \rho \Sigma a_i l_i = 0. \tag{40}$$

Thus, the line with direction cosines proportional to the a's will be orthogonal to any line in the $(n-1)$-flat. This perpendicular is called the *normal* to the $(n-1)$-flat. If we write the equation of an arbitrary $(n-1)$-flat in the form

$$\Sigma l_i X_i = p \tag{41}$$

the l's are the direction cosines of the normal and p is the length of the perpendicular from the origin on to the $(n-1)$-flat.

26 The angle between two $(n-1)$-flats may be defined as the angle between their normals. Thus the angle between

$$\Sigma l_i X_i = p \tag{42}$$

and

$$\Sigma l'_i X_i = p' \tag{43}$$

is given by

$$\arccos(\Sigma l_i l'_i). \tag{44}$$

27 When we come to consider the angles between flats other than S_1 and S_{n-1} in S_n we encounter a new idea which is not visualizable in two or three dimensions. In fact, two flats may have more than one angle between them.

Consider, for example, two planes in S_4. They have one point in common. If we regard one plane as fixed and the common point as fixed on it, the other plane can vary about this point in a double infinity of ways; for two additional points are required to fix it. By fixing one angle with the first plane we do not determine the second plane completely. We need two conditions, equivalent to two angles, to do so. We may therefore expect to find two angles between the planes. Any two angles, in a sense, would do. But we shall adopt a criterion suggested by definition of "the" angle between a line and a plane in three dimensions. In fact, that angle is the minimum angle between the given line and an arbitrary line in the plane. We shall adopt this minimization principle.

28 Consider then an S_{n-p} defined by p equations of type

$$a_{i1} x_1 + a_{i2} x_2 + \dots a_{in} x_n = 0, \qquad i = 1, 2, \dots, p. \tag{45}$$

Writing $\mathbf{A}$ for the $p \times n$ matrix of coefficients and $\mathbf{X}$ for the column $(n \times 1)$ matrix of x's, we have

$$\mathbf{AX} = 0. \tag{46}$$

We lose no generality by supposing the S_{n-p} to go through the origin.

If, then, $\mathbf{L}$ is the column vector representing a line in this plane we have, as at (40),

$$\mathbf{AL} = 0 \tag{47}$$

subject to

$$\mathbf{L'L} = 1. \tag{48}$$

Likewise in a second space S_{n-q} determined by q equations with coefficients $\mathbf{B}$ we have

$$\mathbf{BM} = 0 \tag{49}$$

$$\mathbf{M'M} = 1, \tag{50}$$

where $\mathbf{M}$ represents an arbitrary line with direction cosines m_i.

The angle between a line in one space and a line in the other is the angle ϕ whose cosine is

$$\mathbf{L'M} = \mathbf{M'L} = \cos\phi = R, \text{ say.} \tag{51}$$

We require to find stationary values of R subject to the four conditions (47)–(50). Take Lagrange multipliers $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \alpha_1, \alpha_2$, where λ_1 is a $(1 \times p)$ vector and λ_2 is a $(1 \times q)$ vector. Then we require the unconditioned stationary values of

$$\mathbf{L'M} - \boldsymbol{\lambda}_1 \mathbf{AL} - \boldsymbol{\lambda}_2 \mathbf{BM} - \alpha_1(\mathbf{L'L} - 1) - \alpha_2(\mathbf{M'M} - 1) \tag{52}$$

for variations in $\mathbf{L}$ and $\mathbf{M}$. Differentiating by l_i, m_i in turn, we have

$$\mathbf{M'} - \boldsymbol{\lambda}_1 \mathbf{A} - 2\alpha_1 \mathbf{L'} = 0 \tag{53}$$

$$\mathbf{L'} - \boldsymbol{\lambda}_2 \mathbf{B} - 2\alpha_2 \mathbf{M'} = 0. \tag{54}$$

Postmultiplying (53) by $\mathbf{L}$, we have, in virtue of (47) and (48),

$$\mathbf{M'L} = 2\alpha_1 = R \tag{55}$$

and similarly

$$\mathbf{L'M} = 2\alpha_2 = R. \tag{56}$$

Hence

$$-\boldsymbol{\lambda}_1 \mathbf{A} = \mathbf{RL'} - \mathbf{M'} \tag{57}$$

$$-\boldsymbol{\lambda}_2 \mathbf{B} = \mathbf{RM'} - \mathbf{L'}. \tag{58}$$

Postmultiplying (57) by $\mathbf{A}'$ and by $\mathbf{B}'$, we have, since $(\mathbf{L}'\mathbf{A}') = (\mathbf{AL})' = 0$,

$$-\boldsymbol{\lambda}_1 \mathbf{AA}' = -\mathbf{M}'\mathbf{A}' \tag{59}$$

$$-\boldsymbol{\lambda}_1 \mathbf{AB}' = \mathbf{RL}'\mathbf{B}', \tag{60}$$

and similarly, postmultiplying (58) by $\mathbf{B}'$ and by $\mathbf{A}'$,

$$-\boldsymbol{\lambda}_2 \mathbf{BB}' = -\mathbf{L}'\mathbf{B}' \tag{61}$$

$$-\boldsymbol{\lambda}_2 \mathbf{BA}' = \mathbf{RMA}. \tag{62}$$

From (59) and (62) we then have

$$R\boldsymbol{\lambda}_1 \mathbf{AA}' + \boldsymbol{\lambda}_2 \mathbf{BA}' = 0, \tag{63}$$

and from (60) and (61) similarly

$$\boldsymbol{\lambda}_1 \mathbf{AB}' + R\boldsymbol{\lambda}_2 \mathbf{BB}' = 0. \tag{64}$$

Writing now

$$\mathbf{U} = \mathbf{AA}' \text{ (a symmetrical } p \times p \text{ matrix)}, \tag{65}$$

$$\mathbf{V} = \mathbf{BA}' \text{ (} q \times p \text{ matrix)}, \tag{66}$$

$$\mathbf{V}' = (\mathbf{BA}')' = \mathbf{AB}' \text{ (} p \times q \text{ matrix)}, \tag{67}$$

$$\mathbf{W} = \mathbf{BB}' \text{ (a symmetrical } q \times q \text{ matrix)}, \tag{68}$$

we have

$$\left.\begin{aligned} R\boldsymbol{\lambda}_1 \mathbf{U} + \boldsymbol{\lambda}_2 \mathbf{V} &= 0 \\ \boldsymbol{\lambda}_1 \mathbf{V}' + R\boldsymbol{\lambda}_2 \mathbf{W} &= 0 \end{aligned}\right\} \tag{69}$$

and eliminating $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2$, we have

$$\begin{vmatrix} RU' & V \\ V' & RW' \end{vmatrix} = \begin{vmatrix} RU & V' \\ V & RW \end{vmatrix} = 0 \tag{70}$$

or, equivalently,

$$(R^2)^{q-p} \begin{vmatrix} R^2 U & V' \\ V & W \end{vmatrix} = 0. \tag{71}$$

This is a $(p+q)^2$ determinant of degree p in R^2. It can be solved for R^2 and hence there are p stationary (minimal) angles between the spaces. We may reduce (70) to a p-way determinant as follows:

In (69) multiply the first equation by R and the second by $\mathbf{W}^{-1}\mathbf{V}$, and subtract. We find

$$(R^2\mathbf{U}-\mathbf{V}'\mathbf{W}^{-1}\mathbf{V})\lambda_1 = 0$$

and hence

$$|R^2\mathbf{U}-\mathbf{V}'\mathbf{W}^{-1}\mathbf{V}| = 0. \tag{72}$$

29 Before considering an example let us pick up a few loose ends. First, we are assuming that **U** and **W** have inverses, that is to say, that **A, B** are not degenerate. If they are, some of the relations determining S_{n-p} and S_{n-q} are redundant and those spaces are really of greater dimensions. We may suppose that this point has been examined before the investigation begins.

Second, we have, in effect, assumed that $p \leqslant q$. This point cropped up in proceeding from (70) to (71). If $p > q$ we merely invert the roles of the two spaces. It is the smaller of the numbers p, q which determines the number of non-zero minimal angles between them; in (71) $q-p$ of the values of R^2 are zero.

Third, the lines corresponding to the minimal angles are orthogonal in their respective spaces.

In fact, $\mathbf{V}'\mathbf{W}^{-1}\mathbf{V}$ and **U** are symmetric matrices and hence so is $\mathbf{U}^{-1}\mathbf{V}'\mathbf{W}^{-1}\mathbf{V}$. Call it **K**. We have

$$\mathbf{K}\boldsymbol{\lambda}_1 = R^2\boldsymbol{\lambda}_1. \tag{73}$$

If R, S are two different roots of (72), with corresponding vectors $\boldsymbol{\lambda}_1, \mathbf{p}_1$, we also have

$$\mathbf{K}\mathbf{p}_1 = S^2\mathbf{p}_1.$$

Thus

$$\mathbf{p}_1'\mathbf{K}\boldsymbol{\lambda}_1 = R^2\mathbf{p}_1'\boldsymbol{\lambda}_1,$$
$$\boldsymbol{\lambda}_1'\mathbf{K}\mathbf{p}_1 = S^2\boldsymbol{\lambda}_1'\mathbf{p}_1.$$

In virtue of the symmetry of **K**, $\mathbf{p}_1'\mathbf{K}\boldsymbol{\lambda}_1 = \boldsymbol{\lambda}_1'\mathbf{K}\mathbf{p}_1$ and hence

$$(R^2-S^2)\mathbf{p}_1'\boldsymbol{\lambda}_1 = 0.$$

Thus, R and S being different, the vectors $\mathbf{p}_1$ and $\boldsymbol{\lambda}_1$ are orthogonal. It then follows from (59) that the corresponding **M**'s are orthogonal.

Example 8

Suppose we have two planes in S_4 determined by

$$\left.\begin{aligned} x_1+7x_2+x_3 &= 0 \\ x_4 &= 0 \end{aligned}\right\}, \tag{74}$$

$$\left.\begin{aligned} x_1+x_2 &= 0 \\ x_3+x_4 &= 0 \end{aligned}\right\}. \tag{75}$$

We have $\mathbf{A} = \begin{pmatrix} 1 & 7 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ $\quad \mathbf{B} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}$.

Thus $\mathbf{AA}' = \begin{pmatrix} 51 & 0 \\ 0 & 1 \end{pmatrix}$ $\quad \mathbf{BB}' = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$

$\mathbf{AB}' = \begin{pmatrix} 8 & 1 \\ 0 & 1 \end{pmatrix}$ $\quad \mathbf{BA}' = \begin{pmatrix} 8 & 0 \\ 1 & 1 \end{pmatrix}$.

Equation (71) is then

$$\begin{vmatrix} 51R^2 & 0 & 8 & 1 \\ 0 & R^2 & 0 & 1 \\ 8 & 0 & 2 & 0 \\ 1 & 1 & 0 & 2 \end{vmatrix} = 0, \tag{76}$$

which reduces to

$$51R^4-58R^2+16 = 0$$

$$(17R^2-8)(3R^2-2) = 0.$$

$$R = \pm\sqrt{\tfrac{8}{17}} \quad \text{or} \quad \pm\sqrt{\tfrac{2}{3}}.$$

These are the cosines of the two angles between the planes.

Equivalently, from (72) we find, since

$$V'W^{-1}V = \begin{pmatrix} \frac{65}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix},$$

$$\begin{vmatrix} 51R^2 - \frac{65}{2} & -\frac{1}{2} \\ -\frac{1}{2} & R^2 - \frac{1}{2} \end{vmatrix} = 0 \tag{77}$$

leading to the same result.

Example 9

It follows from our treatment that two p-flats in S_{2p} have p angles between them. In S_{2p+1} two p-flats also have p angles. Whereas in S_{2p} they meet in a point, in S_{2p+1} they do not. One root in R, as at (71), then vanishes. Thus there is a line, S_1, perpendicular to both p-flats, and the distance between the points where it intersects them is the "distance" between the flats. (Cf. the case of three dimensions.)

Reciprocity

30 In classical geometry there is a very useful reciprocal relation between points and lines (in a plane) or between points and planes (in three dimensions). In the plane, for example, two points determine a common line; two lines determine a common point. It will be found that the correspondence exists in the other axioms or postulates of geometry, with the result that we can translate propositions into a dual form. For example, if a hexagon is inscribed in a conic, the meets of opposite pairs of sides are collinear (Pascal's theorem). Since (we assume this for the purpose of exemplification) the conic traced by a point is dual to a conic enveloped by lines, we have immediately the dual proposition: if a hexagon is circumscribed to a conic the joins of opposite pairs of vertices are concurrent (Brianchon's theorem). In n dimensions a point corresponds to a flat $(n-1)$, an S_p to an S_{n-p+1} and so on.

Analytically we may consider the $(n-1)$-flat

$$l_1 X_1 + \ldots + l_n X_n = \text{constant} \tag{78}$$

from two points of view: given the X's, the l's obey a linear relation, given the l's the X's do so. The symmetrical bilinear form (78) determines a (1, 1) relation between a family of points determined by the X's and a family of flats determined by the l's. This type of duality has not been much used in a statistical context.

31 There is, however, a second and distinct kind of duality which should be noticed. If n points in p dimensions are given by the array

$$\begin{array}{c} x_{11} \ldots x_{1n} \\ \cdot \quad \cdot \quad \cdot \\ x_{p1} \ldots x_{pn} \end{array} \tag{79}$$

we may, so to speak, read the matrix downwards (transpose it) so as to get

$$\begin{array}{c} x_{11} \ldots x_{p1} \\ \cdot \quad \cdot \quad \cdot \\ x_{1n} \ldots x_{pn}. \end{array} \tag{80}$$

We may thus regard the array as corresponding to a set of p points in n dimensions, rather than n points in p dimensions. This kind of duality between spaces of n and p dimensions will be exemplified later.

Solid angles

32 In three dimensions we find the notion of "solid angle" of a cone. This is the area which the cone cuts off on the surface of a unit sphere centred at the apex of the cone. Likewise, in n dimensions, we can determine a solid angle of a hypercone by considering the content (volume) of a region cut off on the hypersphere centred at the apex of the cone. This kind of measure, though termed "angle", is really related to "content", and we shall discuss below the interpretation to be placed upon that quantity, which corresponds to area in the plane and volume in three dimensions.

Tangent $(n-1)$-flats

33 An equation such as

$$f(X_1, \dots, X_n) = 0 \tag{81}$$

determines a V_{n-1} in S_n. Let $x_1, \dots, x_n$ be a point on the V_{n-1}. By Taylor's theorem we have

$$f(X_1, \dots, X_n) = f(x_1, \dots, x_n) + \sum_i \left(\frac{\partial f}{\partial X_i}\right)_{x_i} (X_i - x_i)$$
$$+ \text{terms of higher order in } (X_i - x_i). \tag{82}$$

Thus at the point x the $(n-1)$-flat

$$\sum_{i=1}^{n} (X_i - x_i) \left(\frac{\partial f}{\partial X_i}\right)_{x_i} = 0 \tag{83}$$

has first-order contact with the V_{n-1}. It is called the tangent $(n-1)$-flat.

It follows that the line

$$\frac{X_1 - x_1}{\left(\frac{\partial f}{\partial X_1}\right)_{x_1}} = \frac{X_2 - x_2}{\left(\frac{\partial f}{\partial X_2}\right)_{x_2}} = \dots = \frac{X_n - x_n}{\left(\frac{\partial f}{\partial X_n}\right)_{x_n}} \tag{84}$$

passes through the point $x_1, \dots, x_n$ and is perpendicular to the tangent $(n-1)$-flat there. It is called the *normal* to the V_{n-1} at that point.

Reduction of a quadric to canonical form

34 Consider the quadric V_{n-1}

$$\Sigma a_{ij} X_i X_j + \Sigma b_{ij} X_i + c = 0. \tag{85}$$

Let us make this homogeneous by introducing a new dummy variable X_{n+1} equal to unity and writing $b_{i,n+1} = a_{i,n+1}$,

$c = a_{n+1,n+1}$. Then (85) may be written

$$\sum_{i,j=1}^{n+1} a_{ij} X_i X_j = 0 \tag{86}$$

or, in matrix form,

$$\mathbf{X}'\mathbf{A}\mathbf{X} = 0, \tag{87}$$

where $\mathbf{A}$ is the matrix (a_{ij}) and $\mathbf{X}$ is the column vector $(x_1, x_2, \ldots, x_{n+1})$.

Let us make an orthogonal transformation to new variables $\mathbf{Y}$ given by

$$\mathbf{X} = \mathbf{L}\mathbf{Y}. \tag{88}$$

The form (87) then becomes

$$\mathbf{Y}'\mathbf{L}'\mathbf{A}\mathbf{L}\mathbf{Y} = 0. \tag{89}$$

If we now can find a diagonal matrix

$$\mathbf{\Lambda} = \begin{pmatrix} \lambda_1 & & & & 0 \\ & \lambda_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ 0 & & & & \lambda_{n+1} \end{pmatrix}$$

such that

$$\mathbf{L}'\mathbf{A}\mathbf{L} = \mathbf{\Lambda}, \tag{90}$$

our form (89) reduces to

$$\mathbf{Y}'\mathbf{\Lambda}\mathbf{Y} = \lambda_1 Y_1^2 + \lambda_2 Y_2^2 + \ldots \lambda_{n+1} Y_{n+1}^2 = 0. \tag{91}$$

The form is then said to be canonical.

For (90) to be true we must have

$$\mathbf{L}\mathbf{L}'\mathbf{A}\mathbf{L} = \mathbf{L}\mathbf{\Lambda}$$

and since, by hypothesis, $\mathbf{L}\mathbf{L}' = 1$, we have

$$\mathbf{A}\mathbf{L} = \mathbf{L}\mathbf{\Lambda}. \tag{92}$$

Now let $\mathbf{l}_i$ be the ith column vector in L. We shall have, in virtue of (92) and the diagonal character of Λ,

$$\mathbf{A}\mathbf{l}_i - \lambda_i \mathbf{l}_i = 0 \tag{93}$$

and hence

$$|A-\lambda_i I| = 0. \tag{94}$$

The equation $|A-\lambda I| = 0$ is of degree $n+1$ in λ. It is clear that its roots are the $n+1$ values of λ in Λ, for (93) is true of any i. Thus, if we can solve (94) we can find l_i from (93) together with the orthogonality condition $l'l = 1$; hence we can find L, which gives the required transformation to the canonical form (91). It will, in general, be unique.

35 The following points are to be noted.

(a) In practice it is normally more convenient to remove the linear terms from (85) by a transformation of the origin before reducing to canonical form, which then becomes

$$\sum_{i=1}^{n} \lambda_i Y_i^2 = \text{constant}. \tag{95}$$

The constant on the right is almost always positive in statistical applications.

(b) We quote without proof a result from matrix theory to the effect that if $\mathbf{A}$ is a symmetric real matrix (which is so in our case since we may take $a_{rs} = a_{sr}$ without loss of generality) then all the roots in λ are real. Accordingly the transformation is real.

(c) If the form $\mathbf{X'AX}$ is positive definite then all the roots λ are positive; for otherwise the canonical form (91) could be negative or zero. This case is of particular interest in statistics.

(d) In degenerate cases some λ's may be zero. The form (91) then contains fewer than n variables, and $\mathbf{X'AX}$ is a variety of lower dimension.

(e) If certain λ's are equal the transformation is not unique. If, for example, λ_j and λ_i are equal, l_i and l_j become indeterminate unless a further condition is imposed. The transformation then has a single infinity of freedom.

36 From the purely geometrical viewpoint the transformation we are here considering is the analogue of transformation to the principal axes of a conic in two dimensions or a quadric in three dimensions. Equation (95) may be regarded as a hyper-ellipsoid if all the λ's are positive and a hyper-hyperboloid if some are negative. The case when certain λ's are equal corresponds to circular symmetry.

37 The transformation we have so far considered is a "rotation" of the coordinate axes. If we care to make a further transformation of scale on each axis of the type

$$Z_i = Y_i\sqrt{\lambda_i} \tag{96}$$

equation (95) becomes simply

$$\sum_{i=1}^{n} Z_i^2 = \text{constant},$$

corresponding to a hypersphere. This is the analogue of the theorem of two-dimensional geometry that a conic can be projected into a circle.

It follows, since (96) is invariant under any further rotation of the axes, that if we have two quadratic forms

$$\left.\begin{aligned} \sum_{i,j=1}^{n} a_{ij}X_iX_j &= Q_1 \\ \sum_{a,j=1}^{n} b_{ij}X_iX_j &= Q_2 \end{aligned}\right\} \tag{97}$$

there exists a linear transformation of the coordinates which will reduce them simultaneously to canonical form. Suppose that the transformation is

$$\mathbf{X} = \mathbf{CY}.$$

The two forms become

$$\left.\begin{aligned} \mathbf{Y'C'ACY} &= Q_1 \\ \mathbf{Y'C'BCY} &= Q_2 \end{aligned}\right\}. \tag{98}$$

If

$$\left.\begin{aligned}\mathbf{C'AC} &= \boldsymbol{\lambda}, \quad \text{a diagonal matrix}\\ \mathbf{C'BC} &= \boldsymbol{\mu}, \quad \text{a diagonal matrix}\end{aligned}\right\}, \tag{99}$$

the forms become

$$\left.\begin{aligned}\mathbf{Y'\boldsymbol{\lambda} Y} &= Q_1\\ \mathbf{Y'\boldsymbol{\mu} Y} &= Q_2\end{aligned}\right\}. \tag{100}$$

Now $\boldsymbol{\mu}^{-1}\boldsymbol{\lambda}$ is also diagonal, say ν, and is equal, from (99), to

$$\begin{aligned}(\mathbf{C'BC})^{-1}(\mathbf{C'AC}) &= \mathbf{C}^{-1}\mathbf{B}^{-1}(\mathbf{C'})^{-1}\mathbf{C'AC}\\ &= \mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{AC}.\end{aligned} \tag{101}$$

Writing $\mathbf{D}$ for $\mathbf{B}^{-1}\mathbf{A}$ we then have

$$\mathbf{C}^{-1}\mathbf{DC} = \boldsymbol{\nu} \tag{102}$$

$$\mathbf{DC} = \mathbf{C}\boldsymbol{\nu}$$

$$(\mathbf{D}-\boldsymbol{\nu})\,\mathbf{C} = 0, \tag{103}$$

which is of the same form as (92). Thus, given $\mathbf{A}, \mathbf{B}$ (and hence $\mathbf{D}$) we find the roots in ν of

$$|\mathbf{B}^{-1}\mathbf{A} - \nu\mathbf{I}| = 0 \tag{104}$$

and this gives us the required transformation. Equation (104) is equivalent to

$$|A - \nu B| = 0. \tag{105}$$

We may, in fact, regard (94) as a particular case of (105) with $B = I$.

38 The manual solution of equations like (92) and (104) for more than two dimensions is a somewhat tedious process which is best carried out by an iterative procedure, for the details of which see Kendall (1957). Most electronic computers may be programmed to solve them and at the same time give the corresponding transformation.

The projection of an angle

39 Consider a line in S_n through the origin O and a point P a unit distance from O. If the direction cosines of the line are $l_1, \ldots, l_n$, the coordinates of P are also $l_1, \ldots, l_n$. If we project the line on to, say, P', in the plane $X_n = 0$, the coordinates of P' are $l_1, \ldots, l_{n-1}$ and the line OP' has direction cosines proportional to $l_1, \ldots, l_{n-1}$. Similarly for a line with direction cosines typified by l'_i.

The angle between the original lines is $\Sigma l_i l'_i$. That between the projected lines is

$$\frac{\sum_{i=1}^{n-1} l_i l'_i}{\{(1-l_n^2)(1-l_n'^2)\}^{\frac{1}{2}}}, \tag{106}$$

the denominator arising from the fact that the sum of squares of a set of direction cosines is unity. We may write (106) as

$$\frac{\sum_{i=1}^{n} l_i l'_i - l_n l'_n}{\{(1-l_n^2)(1-l_n'^2)\}^{\frac{1}{2}}}. \tag{107}$$

Now angles are invariant under rotation of axes. Thus, if two lines make an angle θ and we project them on to an arbitrary plane whose normal makes angles α, β with them, the angle between the projections, say ϕ, is given by

$$\cos\phi = \frac{\cos\theta - \cos\alpha\cos\beta}{\{(1-\cos^2\alpha)(1-\cos^2\beta)\}^{\frac{1}{2}}}. \tag{108}$$

This formula, in another guise, is familiar in statistics as the relation between a partial correlation and total correlations.

Content

40 By definition the content (n-dimensional volume) of a closed V_{n-1} is the n-fold integral

$$\int \ldots \int dx_1 \ldots dx_n$$

taken over the "inside" of the hypersurface. It will be evident that the content of a hypercube of side length a is a^n. A rectangular parallelotope of sides $a_1, \dots, a_n$ has content $a_1 a_2 \dots a_n$. We proceed to find the content of some of the simpler V_{n-1}'s.

Content of a hypersphere

41 Let the hypersphere of radius a be determined by

$$\sum_{i=1}^{n} X_i^2 = a^2.$$

Make a polar transformation of the type considered in section **21**. The integral giving the content becomes

$$\int_0^a dr \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} d\theta_1 \int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} d\theta_2 \dots \int_0^{\pi} d\theta_{n-1}\, r^{n-1} c_1^{n-2} c_2^{n-3} \dots c_{n-2}. \quad (109)$$

All the variables are independent. We have, putting $\cos^2\theta = u$,

$$\int_{-\frac{1}{2}\pi}^{\frac{1}{2}\pi} \cos^j\theta\, d\theta = -2\int_0^{\frac{1}{2}\pi} \cos^j\theta \frac{d(\cos^2\theta)}{2\cos\theta\sin\theta}$$
$$= \int_0^1 (1-u)^{-\frac{1}{2}} u^{\frac{1}{2}(j-1)} = B\{\tfrac{1}{2}, \tfrac{1}{2}(j+1)\}$$
$$= \frac{\Gamma(\frac{1}{2})\Gamma\{\frac{1}{2}(j+1)\}}{\Gamma\{\frac{1}{2}(j+2)\}}.$$

Substituting such results in (109) we find for the content C

$$C = \frac{a^n}{n} \frac{\Gamma(\frac{1}{2})\Gamma\{\frac{1}{2}(n-1)\}}{\Gamma(\frac{1}{2}n)} \Gamma(\tfrac{1}{2}) \frac{\Gamma\{\frac{1}{2}(n-2)\}}{\Gamma\{\frac{1}{2}(n-1)\}} \cdots \frac{\Gamma(\frac{1}{2})\Gamma(1)}{\Gamma(\frac{3}{2})}.2\pi$$
$$= \frac{2a^n}{n} \frac{\pi^{\frac{1}{2}n}}{\Gamma(\frac{1}{2}n)}. \quad (110)$$

For example, with $n = 2$ we have πa^2, the area of a circle; with $n = 3$ we have $4\pi a^3/3$, the volume of a sphere; and so on.

42 The analogue of the "surface area" of a sphere is derivable immediately. If we differentiate the content of a hypersphere with respect to a we obtain the surface content. Thus the surface content is given by

$$\frac{2a^{n-1}\pi^{\frac{1}{2}n}}{\Gamma(\frac{1}{2}n)}. \tag{111}$$

Thus, for $n = 2$ we have $2\pi a$, the circumference of a circle; for $n = 3$ we have $4\pi a^2$, the surface of a sphere; and so on.

Content of a hyperellipsoid

43 Content being independent of coordinate axes, we may, without loss of generality, suppose the hyperellipsoid put in the form

$$\frac{X_1^2}{a_1^2}+\frac{X_2^2}{a_2^2}+\ldots+\frac{X_n^2}{a_n^2} = 1. \tag{112}$$

The lengths $a_1, \ldots, a_n$ are the semi-axes of the hyperellipsoid. A volume integral over a region determined by (112) is seen, by the transformation $X_i = a_i Y_i$, to be equal to $a_1, \ldots, a_n$ times the content of a unit sphere. Thus our required content is

$$\frac{2\,a_1 \ldots a_n}{n}\,\frac{\pi^{\frac{1}{2}n}}{\Gamma(\frac{1}{2}n)}. \tag{113}$$

Content of a hyperprism

44 A hyperprism is the figure generated by a flat region of content C in S_{n-1} moving parallel to itself along a straight line which makes an angle θ, say, with the S_{n-1}. It scarcely needs proof that if the distance moved is h the content of the hyperprism is $Ch\sin\theta$.

In particular if the flat region is a hypersphere and the line is perpendicular to it we have the n-dimensional analogue of a cylinder. The content is then $2ha^{n-1}\pi^{\frac{1}{2}(n-1)}/\Gamma\{\frac{1}{2}(n-1)\}$.

Content of a parallelotope

45 Let $x_{ij}, i, j = 1, 2, \ldots, n$ be n points $P_1, \ldots, P_n$ in an S_n. Consider them together with the origin O. Join O to each P

and complete the parallelotope of which O is one corner, by drawing lines through P_1 parallel to $OP_2, \ldots, OP_n$, etc. Let us find the content of this parallelotope.

Make a variate transformation (linear but not orthogonal)

$$\mathbf{X} = \mathbf{AY}. \tag{114}$$

The Jacobian of the transformation is simply $|A|$. Now let the new variables $\mathbf{Y}$ be orthogonal and the points P transform to points with unit distance from the origin. The content of the parallelotope is then simply $|A|$ and we have only to express this in terms of the coordinates of the P's. But from (114) itself we have, for unit values of the Y's, Y then becoming the identity matrix, $X = A$. Hence our content is simply $|X|$.

If now our parallelotope is defined by $(n+1)$ points, excluding the origin, say x_{ij}, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, n+1$, we may translate the origin to one of them, say $x_{i,n+1}$, and the coordinates of the others are then $x_{ij} - x_{i,n+1}$. The content is then

$$|x_{ij} - x_{i,n+1}|$$

which we may write in the symmetrical form

$$C = \begin{vmatrix} x_{11} & x_{12} & \cdots & x_{1n} & x_{1,n+1} \\ x_{21} & x_{22} & \cdots & x_{2n} & x_{2,n+1} \\ x_{n1} & x_{n2} & \cdots & x_{nn} & x_{n,n+1} \\ 1 & 1 & & 1 & 1 \end{vmatrix}. \tag{115}$$

Content of a hyperpyramid

46 A hyperpyramid is the figure generated by joining the vertices of a polytope in S_{n-1} to a point not in that space (called the apex).

Take a set of coordinates in S_{n-1} and let the remaining coordinate be measured along a line through the apex perpendicular to the S_{n-1}. A hyperplane parallel to the S_{n-1}

cuts the pyramid in a polytope which is "similar" to the base polytope. Let the content of the base be C, and the height of the pyramid (the non-zero coordinate of the vertex) be h. A hyperplane at x_n parallel to the base then cuts the pyramid in a parallelotope of content Cx_n^{n-1}/h^{n-1}. The content of the pyramid is then

$$\int_0^h \frac{x_n^{n-1}}{h^{n-1}} C\, dx_n = \frac{Ch}{n}. \tag{116}$$

Content of a simplex

47 Consider the simplex as a pyramid with one point, say P_n, as vertex. If C_{n-1} is the content of the base, we have from the foregoing result for C_n, the content of the simplex.

Let the vertices of the simplex be $x_{ij}, i, j = 1, 2, \ldots, n$ together with the origin. Make a transformation as in section **45**. The content of the parallelotope defined by these points is $|X|$, the vertices of the simplex transforming into the origin and the points at unit distance from it along the coordinate axes. Consider the content of this transformed simplex, say D_n. Regarding it as a pyramid with one vertex at P_n we have, from (116),

$$D_n = \frac{1}{n} D_{n-1},$$

and by repetition of the process

$$D_n = \frac{1}{n!}. \tag{117}$$

Thus the content of the original simplex is $1/n!$ times the corresponding parallelotope. If it is defined by $(n+1)$ points, excluding the origin, the content is

$$\frac{1}{n!} C \tag{118}$$

where C is given by (115).

48 Formula (118) gives the content of the simplex in terms of the coordinates of its vertices. We may also derive an expression in terms of the lengths of its S_1-edges. Writing D for the content we have

$$n!\,D = \begin{vmatrix} x_{11} & x_{12} & \cdots & x_{1,n+1} \\ x_{21} & x_{22} & \cdots & x_{2,n+1} \\ x_{n1} & x_{n2} & \cdots & x_{n,n+1} \\ 1 & 1 & & 1 \end{vmatrix}$$

which we may write as

$$(-1)^n 2^n n!\,D = \begin{vmatrix} 1 & \sum_{i=1}^{n} x_{i1}^2 & \sum_{i=1}^{n} x_{i2}^2 & \cdots & \sum_{i=1}^{n} x_{i,n+1}^2 \\ 0 & -2x_{11} & -2x_{12} & \cdots & -2x_{1,n+1} \\ 0 & -2x_{21} & -2x_{22} & \cdots & -2x_{2,n+1} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & -2x_{n1} & -2x_{n2} & \cdots & -2x_{n,n+1} \\ 0 & 1 & 1 & \cdots & 1 \end{vmatrix} \qquad (119)$$

or equivalently

$$(-1)^{2n+1} n!\,D = \begin{vmatrix} 0 & 0 & 0 & \cdots & 0 & 1 \\ 1 & x_{11} & x_{21} & \cdots & x_{n1} & \Sigma x_{i1}^2 \\ 1 & x_{12} & x_{22} & \cdots & x_{n2} & \Sigma x_{i2}^2 \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ 1 & x_{1,n+1} & x_{2,n+1} & \cdots & x_{n,n+1} & \Sigma x_{i,n+1}^2 \end{vmatrix}. \qquad (120)$$

Postmultiply (120) by (119). In virtue of relations such as

$$\left.\begin{aligned} \Sigma x_{ij}^2 - 2\Sigma x_{ij}\,x_{ik} + \Sigma x_{ik}^2 = \Sigma\,(x_{ij}-x_{ik})^2 &= a_{jk}^2, \quad j \neq k \\ &= 0, \quad j = k \end{aligned}\right\} \qquad (121)$$

where a_{jk} is the length of the edge determined by x_j and x_k, we find

$$(-1)^{n+1} 2^n (n!)^2 D^2 = \begin{vmatrix} 0 & 1 & 1 & \dots & 1 \\ 1 & 0 & a_{12}^2 & \dots & a_{1,n+1}^2 \\ 1 & a_{21}^2 & 0 & \dots & a_{2,n+1}^2 \\ \cdot & \cdot & \cdot & \dots & \cdot \\ 1 & a_{n+1,1}^2 & a_{n+1,2}^2 & \dots & 0 \end{vmatrix}. \tag{122}$$

For the regular simplex all the a_{ij} are equal, say to a. Substituting in (122) and subtracting the last column from the others except the first we find

$$(-1)^{n+1} 2^n (n!)^2 D^2 = \begin{vmatrix} 0 & 0 & 0 & \dots & 1 \\ 1 & -a^2 & 0 & \dots & a^2 \\ 1 & 0 & -a^2 & \dots & a^2 \\ \cdot & \cdot & \cdot & \dots & \cdot \\ 1 & a^2 & a^2 & \dots & 0 \end{vmatrix}$$

$$= (-1)^{n+1} a^{2n} \begin{vmatrix} -1 & 0 & \dots & 1 \\ 0 & -1 & \dots & 1 \\ \cdot & \cdot & \dots & 1 \\ 0 & 0 & \dots & 1 \\ 1 & 1 & \dots & 1 \end{vmatrix}$$

$$= (-1)^{n+1} (n+1) a^{2n}$$

giving

$$D = \frac{a^n}{n!} \left(\frac{n+1}{2^n} \right)^{\frac{1}{2}}. \tag{123}$$

For example with $n = 3$ we have for the volume of a regular tetrahedron $a^3/6\sqrt{2}$, a result which is easy to verify directly.

49 We shall not be concerned with the differential geometry of n dimensions to any considerable extent, but two points of interest may be remarked upon before we proceed to statistical applications.

(a) In transforming from one coordinate system to another, say from x's to y's, the content of an elementary volume $dx_1 \dots dx_n$ transforms to $J\,dy_1 \dots dy_n$ where J is the Jacobian $\partial(x_1, x_2, \dots, x_n)/\partial(y_1, y_2, \dots, y_n)$. We have used this result repeatedly in the foregoing and it needed no *ad hoc* geometrical proof, being a result in analysis which is derived in most textbooks on the real variable. It may, however, be regarded from a geometrical viewpoint. If we consider a small displacement from x_i to $x_i + dx_i$, the corresponding point in the y-domain has coordinates

$$y_i + \sum_{j=1}^{n} \frac{\partial x_i}{\partial y_j} \partial y_j.$$

The content of the parallelotope corresponding to $dx_1 \dots dx_n$ is then, after the manner of section 45, equal to

$$\left| \frac{\partial x_i}{\partial y_j} dy_j \right| = \left| \frac{\partial x_i}{\partial y_j} \right| dy_1 \dots dy_n = J\,dy_1 \dots dy_n.$$

(b) If we make a coordinate transformation from x to y, the lines $y_i =$ constant trace out a coordinate mesh in the x domain and vice versa. The normal to the hypersurface (V_{n-1}) given by $y_i =$ constant has direction cosines (cf. section **33**) proportional to

$$\frac{\partial x_j}{\partial y_i}, \quad j = 1, 2, \dots, n.$$

In particular, consider the polar transformation of section **21**. We have

$$\frac{\partial x_1}{\partial r} = c_1 c_2 \dots c_{n-2} c_{n-1} \qquad \frac{\partial x_1}{\partial \theta_1} = -r s_1 c_2 \dots c_{n-2} c_{n-1}$$

$$\frac{\partial x_2}{\partial r} = c_1 c_2 \dots c_{n-2} s_{n-1} \qquad \frac{\partial x_2}{\partial \theta_1} = -r s_1 c_2 \dots c_{n-2} s_{n-1}$$

$$\cdot \quad \cdot \quad \cdot \qquad\qquad \cdot \quad \cdot \quad \cdot$$

$$\frac{\partial x_n}{\partial r} = s_1 \qquad \frac{\partial x_n}{\partial \theta} = r c_1.$$

Hence

$$\sum_{i=1}^{n} \frac{\partial x_i}{\partial r}\frac{\partial x_i}{\partial \theta_1} = -rc_1 s_1 \{c_2^2 \dots c_{n-1}^2 + c_2^2 \dots s_{n-1}^2 + \text{etc.}\} + rc_1 s_1 = 0$$

and hence the hypersurfaces $r =$ constant and $\theta_1 =$ constant are orthogonal. Likewise it will be found that $r =$ constant is orthogonal to every $\theta =$ constant and that the latter are orthogonal among themselves. It is easy to verify this situation in two and three dimensions.

PART 2

STATISTICAL APPLICATIONS

Introduction

50 The normal distribution with mean μ and variance σ^2 is given by the density (frequency) function

$$dF = \frac{1}{\sigma\sqrt{(2\pi)}}\exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right\}dx. \tag{124}$$

The distribution of a sample of n independent values from such a population is then given by the joint frequency element

$$dF = \frac{1}{\sigma^n(2\pi)^{\frac{1}{2}n}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}dx_1\ldots dx_n. \tag{125}$$

Given a statistic $t = t(x_1, x_2, \ldots, x_n)$ we can find its distribution function by integrating (125) over a region for which $t \leqslant t_0$, a given value of t. The boundary of this region is the V_{n-1} in the S_n space of $(x_1, x_2, \ldots, x_n)$ given by $t = t_0$. If we regard

$$\frac{1}{\sigma^n(2\pi)^{\frac{1}{2}n}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\} \tag{126}$$

as a "density" in the S_n, our problem is then to find the "weight" in the region bounded by $t = t_0$. Equivalently, we may seek the element of weight in the range dt, which gives us the frequency, as distinct from the distribution, function of t. It is to this problem that we proceed to apply our n-dimensional geometry.

First of all, by a simple transformation $x_i - \mu = y_i$ we may write the density without the term in μ. This will modify the boundary t_0. Our density function is then

$$\frac{1}{\sigma^n (2\pi)^{\frac{1}{2}n}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n} y_i^2\right) = \frac{1}{\sigma^n (2\pi)^{\frac{1}{2}n}} \exp\left(-\frac{r^2}{2\sigma^2}\right) \quad (127)$$

where

$$r^2 = \sum_{i=1}^{n} y_i^2.$$

Our surfaces of constant density are then hyperspheres. It is to this spherical symmetry that the relative simplicity of our results is due.

We have already observed in Example 7 that when $\sigma = 1$ a polar transformation gives us for the distribution of r^2

$$dF \propto \exp(-\tfrac{1}{2}r^2)\, r^{n-1}\, dr. \quad (128)$$

It follows at once that for non-unit σ the distribution is

$$dF \propto \frac{r^{n-1}}{\sigma^n} \exp\left(-\frac{1}{2}\frac{r^2}{\sigma^2}\right) dr. \quad (129)$$

We also remarked in Examples 5 and 6 that a variate transformation could be made under which one new variable was $\bar{x}$ and this was independent of the others. It follows that the sum $\Sigma(x-\bar{x})^2$ is distributed in the form (129) with one lower dimension number, i.e. with $n-2$ as the power of r instead of $n-1$.

It also follows that if the x's are subject to a homogeneous linear restriction

$$a_1 x_1 + \ldots + a_n x_n = 0 \quad (130)$$

the variation lies in the S_{n-1} determined by (130) and the surfaces of constant density are hyperspheres V_{n-2}^2. Thus the distribution of

$$\sum_{i=1}^{n} x_i^2$$

is of the form (129) except, again, that $n-1$ is to be replaced by $n-2$.

Consider now a positive definite quadratic form in the x's

$$Q = \sum_{i,j=1}^{n} a_{ij} x_i x_j. \tag{131}$$

We can transform to new variables y such that

$$Q = \sum_{i=1}^{n} \lambda_i y_i^2. \tag{132}$$

At the same time the distribution is transformed into

$$dF = \frac{1}{\sigma^n (2\pi)^{\frac{1}{2}n}} \exp\left(-\tfrac{1}{2}\sigma^2 \sum_1^n y_i^2\right) dy_1 \dots dy_n. \tag{133}$$

The surfaces $Q =$ constant and the hyperspheres of constant density no longer intersect in flats or hyperspheres and consequently no simple result is derivable for the distribution of Q. We may, however, use (132) and (133) to determine the moments of Q very simply. For example,

$$E(Q) = \int \Sigma\lambda_i y_i^2 \, dF = \sigma^2 \Sigma\lambda_i,$$

$$E(Q^2) = \int (\Sigma\lambda_i y_i^2)^2 \, dF = 3\sigma^4 \Sigma\lambda_i^2 + \sigma^4 \sum_{i \neq j} \lambda_i \lambda_j$$

$$= \sigma^4 \{(\Sigma\lambda_i)^2 + 2\Sigma\lambda_i^2\},$$

and so on.

51 Consider now the coefficient r_k defined by

$$r_k = \frac{\sum_{i=1}^{n-k} x_i x_{i+k}}{\sum_{i=1}^{n} x_i^2} \cdot \frac{n}{n-k}. \tag{134}$$

By the argument of Example 5 we see that the numerator and denominator of (134) are independent if the x's are normal and independently distributed.

We can now transform to new variables $y_1, \ldots, y_n$ such that the distribution remains as at (133) and

$$r_k = \frac{\sum_{i=1}^{n} \lambda_i y_i^2}{\sum_{i=1}^{n} y_i^2} \cdot \frac{n}{n-k} \tag{135}$$

where the λ's are the roots of the equation (derived from (94))

$$\begin{vmatrix} -\lambda & 0 & 0 & . & \frac{1}{2} & 0 & 0 & . & 0 \\ 0 & -\lambda & 0 & . & 0 & \frac{1}{2} & 0 & . & 0 \\ 0 & 0 & -\lambda & . & 0 & 0 & \frac{1}{2} & . & 0 \\ . & . & . & . & . & . & . & . & . \\ \frac{1}{2} & 0 & 0 & . & -\lambda & 0 & 0 & . & \frac{1}{2} \\ 0 & \frac{1}{2} & 0 & . & 0 & -\lambda & 0 & . & 0 \\ . & . & . & . & . & . & . & . & . \\ 0 & 0 & . & . & \frac{1}{2} & 0 & 0 & . & -\lambda \end{vmatrix} = 0. \tag{136}$$

Again we cannot find a simple expression for the distribution of r. However, the moments of the denominator in (135) are ascertainable from the distribution (129) and those of the numerator in the manner of section **49**. Use of Example 5 then gives us the moments of r_k.

This coefficient is known as the "serial correlation coefficient", in this case with lag k and no parental correlation between successive values of the x's. Slight variations of the form of definition (134) are encountered in time-series analysis. For example, if we write $x_1 = x_{n+1}, x_2 = x_{n+2}, \ldots, x_k = x_{n+k}$, the so-called "circular" coefficient r_c of order k may be defined by

$$r_{ck} = \frac{\sum_{i=1}^{n} x_i x_{i+k}}{\sum_{i=1}^{n} x_i^2}. \tag{137}$$

The usefulness of this result in ascertaining the distribution of r_c is that the determinant (136) then becomes a circulant and the roots in λ can be obtained explicitly.

"Student's" *t*

52 The ratio

$$t = \frac{(\bar{x}-\mu)\sqrt{\{n(n-1)\}}}{\{\Sigma(x-\bar{x})^2\}^{\frac{1}{2}}} \tag{138}$$

is known as "Student's" t. By the foregoing arguments we see that we can transform to the mean $(\mu, \ldots, \mu)$ without loss of generality; and that the denominator is independent of the numerator in normal variation.

If O is the origin, P the sample point and Q the orthogonal projection of P on to the line

$$X_1 = X_2 = \ldots = X_n \tag{139}$$

the length OQ is $\Sigma x_i/\sqrt{n} = \bar{x}\sqrt{n}$ and OP is $(\Sigma x_i^2)^{\frac{1}{2}}$. Thus

$$PQ^2 = \Sigma x_i^2 - n\bar{x}^2 = \Sigma(x_i - \bar{x})^2. \tag{140}$$

It follows that t is constant over the surface OQ/PQ = constant, i.e. over the "circular" cone subtending a fixed half-angle at the origin. Let this angle be ϕ and consider the weight between two cones given by ϕ and $\phi + d\phi$. This will be proportional to the annulus cut off on a fixed hypersphere (say, the unit sphere) with width $d\phi$ and radius $\sin^{n-2}\phi$. We thus have for the distribution of the angle ϕ

$$dF \propto \sin^{n-2}\phi\, d\phi, \quad 0 \leqslant \phi \leqslant \pi.$$

The distribution of

$$t = \frac{\bar{x}\sqrt{\{n(n-1)\}}}{\sqrt{\Sigma(x-\bar{x})^2}} = \frac{OQ\sqrt{(n-1)}}{PQ} = (n-1)^{\frac{1}{2}}\cot\phi$$

is then given by

$$dF \propto \left(1 + \frac{t^2}{n-1}\right)^{-\frac{1}{2}n} dt, \qquad -\infty \leqslant t \leqslant \infty$$

and the constant is easily evaluated to give

$$dF = \frac{1}{B\{\frac{1}{2}(n-1), \frac{1}{2}\}}\left(1+\frac{t^2}{n-1}\right)^{-\frac{1}{2}n}\frac{dt}{(n-1)^{\frac{1}{2}}}. \tag{141}$$

The mean in rectangular variation

53 We consider now a variable distributed over an interval with uniform frequency. Taking the interval to be unity, without loss of generality we may write the distribution as

$$dF = dx, \qquad 0 \leqslant x \leqslant 1. \tag{142}$$

In a sample of n values the density will be unity over the interior of a hypercube $0 \leqslant x_i \leqslant 1$ and zero outside it. This discontinuity in density at the faces of the hypercube gives rise to some difficulty. The frequency function of the mean $\bar{x}$ = constant is given by the weight between the hyperplanes $\bar{x}$ and $\bar{x}+d\bar{x}$. The general hyperplane $\bar{x}$ = constant meets the boundary of the hypercube in a region bounded by flats but changing its shape as $\bar{x}$ increases.

We deal with this problem by introducing "marker" variables. The apexes (corners) of the hypercube may have 0 or 1 or 2 or ... n coordinates equal to unity and the rest zero. The bounding hyperplanes of the cube, if extending indefinitely over positive values of the x's, define orthants (analogous to the quadrant of two dimensions). Consider the orthants defined by $x_j \geqslant r_j$ where r_j may be 0 or 1. These in total contain all the apexes of the cube and each apex is the corner of an orthant. Divide them into $(n+1)$ sets according as the corner contains 0 or 1 or 2 or ... n coordinates equal to unity, i.e. as

$$\sum_{j=1}^{n} r_j = r = 0, 1, \ldots, n.$$

Let Q_i be the number in the ith set.

For example, in three dimensions

$Q_0 = 1, \quad Q_1 = 3, \quad Q_2 = 3, \quad Q_3 = 1,$ totalling 8 corners,

and in general Q_p is the coefficient of t^p in the binomial $(1+t)^n$. (Cf. sections **16** and **17**.) There are 2^n corners and correspondingly 2^n orthants.

We now assign a density to each orthant over its whole domain from 0 or 1 to infinity. To each Q_p we assign the density $(-1)^p$. That this may be negative need not worry us. Now let P be a point (with non-negative x's) such that s of its coordinates are greater than unity. It will then belong to $\binom{s}{p}$ of the Q_p's. Now for $s \geqslant 1$

$$0 = (1-1)^s = \sum_{p=0}^{s} (-1)^p \binom{s}{p}. \tag{143}$$

Hence the total density is zero everywhere so long as $s \geqslant 1$, i.e. the point lies outside the cube. If it lies on or inside the cube the density is unity.

Now let the segment of the hyperplane $\Sigma x = z$ lying in Q_0 have content $C_n(z)$. If this hyperplane also lies in some other quadrant, say a Q_r, with r unit coordinates, the content in Q_r is $C_n(z-r)$. (If $z \leqslant r$ this is zero.) Hence the total weight we require is

$$\sum_{r=0}^{k} (-1)^r \binom{n}{r} C_n(z-r), \tag{144}$$

where k is the greatest integer less than z. It remains to find $C_n(z)$.

Now $C_n(z)$ lies in the hyperplane $\Sigma x =$ constant which meets the coordinate axes in a point with coordinate z. The content of the pyramid determined by the coordinate planes and $\Sigma x = z$ is (cf. (118)) $z^n/n!$. The content $C_n(z)$ of the face of the pyramid not lying in the coordinate planes is obtained by differentiating with respect to z, i.e. is $z^{n-1}/(n-1)!$. Thus we have for the frequency of z in the class of orthants Q_r

$$\frac{1}{(n-1)!}\sum_{r=0}^{k}(-1)^r\binom{n}{r}(z-r)^{n-1}, \qquad k \leqslant z \leqslant k+1. \quad (145)$$

For the mean $\bar{x} = z/n$ we have by a simple substitution

$$dF = \frac{n^n}{(n-1)!}\sum_{r=0}^{k}(-1)^r\binom{n}{r}\left(\bar{x}-\frac{r}{n}\right)^{n-1}d\bar{x}, \qquad \frac{k}{n} \leqslant \bar{x} \leqslant \frac{k+1}{n}. \quad (146)$$

The correlation coefficient in bivariate normal variation

54 The bivariate normal distribution with zero means and unit variances is given by

$$dF = \frac{1}{2\pi(1-\rho^2)^{\frac{1}{2}}}\exp\left\{-\frac{1}{2(1-\rho^2)}(x^2-2\rho xy+y^2)\right\}dx\,dy. \quad (147)$$

We are interested in the distribution of the correlation coefficient, in a sample of n values, defined by

$$r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\{\Sigma(x_i-\bar{x})^2\Sigma(y_i-\bar{y})^2\}^{\frac{1}{2}}}. \quad (148)$$

This statistic is independent of the origin and scale of x and y, so we have lost no generality in using the form (147).

In Example 6 we saw that by performing a Helmert transformation on each of x and y we could reduce the joint frequency of a sample of n to the form

$$k\exp\left\{-\frac{1}{2(1-\rho^2)}\left(\sum_{i=1}^{n-1}u_i^2-2\rho\sum_{i=1}^{n-1}u_i v_i+\sum_{i=1}^{n-1}v_i^2\right)\right\}du_1\ldots du_{n-1}$$
$$\times\, dv_1\ldots dv_{n-1}\exp\left\{-\frac{n}{2(1-\rho^2)}(u_n^2-2\rho u_n v_n+v_n^2)\right\}du_n\,dv_n, \quad (149)$$

where $u_n = \bar{x}$, $v_n = \bar{y}$ and the other u's are independent of u_n and v_n.

In this coordinate system we have

$$r = \frac{\sum_{1}^{n-1} u_i v_i}{\left\{\sum_{1}^{n-1} u_i^2 \sum_{1}^{n-1} v_i^2\right\}^{\frac{1}{2}}}. \qquad (150)$$

We can accordingly consider the distribution of r given by (150) in the distribution given by the first factor on the right in (149). Effectively we have removed the variation of means $\bar{x}$ and $\bar{y}$ from the picture at the expense of going down into one lower dimension.

We will find the joint distribution of r and of s_1 and s_2 given by

$$s_1^2 = \frac{1}{n-1} \sum_{1}^{n-1} u_i^2, \quad s_2^2 = \frac{1}{n-1} \sum_{1}^{n-1} v_i^2. \qquad (151)$$

The frequency function, from (149), is immediately written down as

$$k \exp\left\{-\frac{n-1}{2(1-\rho^2)}(s_1^2 - 2\rho r\, s_1 s_2 + s_2^2)\right\}. \qquad (152)$$

Our only problem is to find how the differential element $du_1 \ldots du_{n-1} dv_1 \ldots dv_{n-1}$ is transformed in terms of s_1, s_2 and r.

Consider two S_{n-1}, one for u and one for v. We can picture these as superposed one on the other. $(n-1)s_1^2$ then represents the distance of the sample point P (corresponding to u) from the origin O, $(n-1)s_2^2$ that of Q (corresponding to v) from O. The coefficient r of (150) is then seen to be equal to the cosine of the angle between OP and OQ.

For an increase ds_1 the content of the hyperspherical shell between s_1 and $s_1 + ds_1$ is, since our surface is in $n-1$ dimensions, proportional to $s_1^{n-2}\, ds_1$. Now the angle $\arccos r$, say θ, is independent of s_1 and s_2. Moreover (section **49**) the variation of θ is orthogonal to that of s_1 and s_2; the content corresponding to increases $ds_1\,dr\,ds_2$ is then the product of the increases corresponding to each separately, i.e. to $s_1^{n-2}s_2^{n-2}\,ds_1\,ds_2$ times the increase due to variation dr. For fixed OP, the vector OQ which makes an angle θ with it varies in the v-space over an annulus of radius $s_2\sqrt{(n-1)}\sin\theta$, and the contribution is therefore proportional to

$$\sin^{n-3}\theta\,d\theta = \sin^{n-4}\theta\,d\cos\theta = (1-r^2)^{\frac{1}{2}(n-4)}\,dr.$$

Hence our total differential element is proportional to

$$s_1^{n-2}s_2^{n-2}(1-r^2)^{\frac{1}{2}(n-4)}. \tag{153}$$

This, multiplied by the frequency of (152), gives us the joint distribution of s_1, s_2 and r. To find the distribution of r alone we have to integrate out s_1 and s_2. For details the reader is referred to Kendall and Stuart, 1957, Vol. 1, p. 384.

Wishart's distribution

55 The procedure we have just followed for two-way variation, namely of finding the joint distribution of sample variances and covariances, may be extended to p-way variation.

We consider a sample of n of a p-variate complex, arrayed by

$$\begin{array}{ccc} x_{11} & \cdots & x_{1p} \\ x_{21} & \cdots & x_{2p} \\ \cdot & \cdots & \cdot \\ x_{n1} & \cdots & x_{np}. \end{array} \tag{154}$$

The multivariate normal distribution of p variates $x_1, \ldots, x_p$ is given by

$$dF = \frac{|A|^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}p}}\exp\left\{-\tfrac{1}{2}\left(\sum_{i,j=1}^{p} A^{ij}x_i x_j\right)\right\}dx_1\ldots dx_p \tag{155}$$

$$= \frac{|A|^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}p}}\exp\left(-\tfrac{1}{2}\mathbf{X}'\mathbf{A}\mathbf{X}\right)dx_1\ldots dx_p \tag{156}$$

where (A^{ij}) is the matrix which is inverse to the so-called correlation matrix $(\rho_{ij}), i,j = 1,2,\ldots,p$. Here we have taken the x's to have unit variances and zero means. It is readily verified that (155) reduces to the univariate and bivariate results of earlier examples when $p = 1, 2$.

For a sample of n (denoting for convenience now summation over the sample by S) we have for the frequency function

$$\frac{|A|^{\frac{1}{2}n}}{(2\pi)^{\frac{1}{2}np}} \exp\left\{-\tfrac{1}{2} \mathop{S}_{k=1}^{n} \sum_{i,j=1}^{p} A^{ij} x_{ki} x_{kj}\right\}$$
$$= \frac{|A|^{\frac{1}{2}n}}{(2\pi)^{\frac{1}{2}np}} \exp\left\{-\tfrac{1}{2} S\Sigma A^{ij}(x_{ki}-\bar{x}_i)(x_{kj}-\bar{x}_j) - \frac{n}{2} S\Sigma A^{ij} \bar{x}_i \bar{x}_j\right\}. \tag{157}$$

Exactly as before we may, by a Helmert transformation, remove the means $\bar{x}$ from the picture and concentrate on the frequency

$$\frac{|A|^{\frac{1}{2}(n-1)}}{n^{\frac{1}{2}p}(2\pi)^{\frac{1}{2}(n-1)p}} \exp\left\{-\tfrac{1}{2} \mathop{S}_{k=1}^{n-1} \sum_{i,j=1}^{p} A^{ij} u_{ki} u_{kj}\right\}. \tag{158}$$

(In the ordinary way, we do not bother about preliminary constants in this class of work, being content to evaluate at the final stage by use of the fact that the total frequency must be unity. Here, however, a final integration would be formidable. We therefore attach

$$\frac{n^{\frac{1}{2}p}|A|^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}p}}$$

to the second factor in (157) to make the distribution of means have unit frequency, and remove it from the constant in (157), giving us the constant in (158).)

Put

$$\mathop{S}_{k=1}^{n-1} u_{ki} u_{kj} = \xi_{ij} = n a_{ij}. \tag{159}$$

The quantities a are then our sample variances and covariances. We have for the frequency (158)

$$\frac{|A|^{\frac{1}{2}(n-1)}}{n^{\frac{1}{2}p}(2\pi)^{\frac{1}{2}(n-1)p}} \exp\left(-\frac{n}{2}\Sigma A^{ij} a_{ij}\right). \tag{160}$$

As in the bivariate case, our main problem arises in determining the differential element to be adjoined to this quantity.

Let us take p flat spaces of $n-1$ dimensions, one for each u, and let the sample points be represented by $P_1, \ldots, P_p$. We consider the variation of P_1, then the variation of P_2 given P_1, then that of P_3 given P_1 and P_2, and so on. We then multiply these together to get the variation of $P_1, \ldots, P_p$. The point O is the origin.

Consider, in fact, P_m given $P_1, P_2, \ldots, P_{m-1}$. For fixed OP_m and angles $P_m OP_1, P_m OP_2, \ldots, P_m OP_{m-1}$, P_m lies on a hypersphere in $n-m$ dimensions. Regarding the spaces as superposed, let the length of the perpendicular from P_m on to the $(m-1)$-flat determined by $O, P_1, \ldots, P_{m-1}$ be t_m. The content of variation of P_m is then, from (111),

$$\frac{2\pi^{\frac{1}{2}(n-m)}\, t_m^{n-m-1}}{\Gamma\{\frac{1}{2}(n-m)\}}. \tag{161}$$

We require the content due to variation perpendicular to this hypersphere. Consider the transformation, based on (159)

$$\xi_{mj} = \sum_{k=1}^{n-1} u_{km} u_{ki}, \qquad i, = 1, 2, \ldots, m. \tag{162}$$

The Jacobian is

$$\begin{vmatrix} u_{11} & u_{12} & \ldots & u_{1m} \\ u_{21} & u_{22} & \ldots & u_{2m} \\ \cdot & \cdot & \ldots & \cdot \\ 2u_{1m} & 2u_{2m} & \ldots & 2u_{m,m} \end{vmatrix} = 2v_m, \tag{163}$$

where v_m is the content of the parallelotope determined by $O, P_1, \ldots, P_m$. Moreover

$$|\xi_{ij}| = v_m^2 \tag{164}$$

and

$$t_m = v_m / v_{m-1}. \tag{165}$$

Thus the differential element is

$$\frac{1}{2v_m}\prod_{i=1}^{m} d\xi_{mi}$$

and, adjoining this to (161), we have for the total element of variation in P_m, given $O, P_1, \ldots, P_{m-1}$,

$$\frac{\pi^{\frac{1}{2}(n-m)}\, v_m^{n-m-2}}{\Gamma\{\frac{1}{2}(n-m)\}\, v_{m-1}^{n-m-1}} \prod_{i=1}^{m} d\xi_{mi}. \qquad (166)$$

We now multiply expressions like this for $m = 1, 2, \ldots, p$, remembering that $v_0 = 1$. We find

$$\frac{\pi^{\frac{1}{4}p(2n-p-1)}}{\prod_{k=1}^{p}\Gamma\{\frac{1}{2}(n-k)\}} v_p^{n-p-2} \prod_{j=1}^{p}\prod_{k=1}^{j} d\xi_{jk}.$$

Finally, since from (164) $v_p^2 = n^p\,|a|$, we have for the required distribution

$$dF = \frac{(\frac{1}{2}n)^{\frac{1}{2}p(n-1)}\,|A|^{\frac{1}{2}(n-1)}\,|a|^{\frac{1}{2}(n-p-2)}}{\pi^{\frac{1}{4}p(p-1)}\prod_{k=1}^{p}\Gamma\{\frac{1}{2}(n-k)\}}$$

$$\times \exp\left(-\frac{n}{2}\Sigma A^{ij} a_{ij}\right) \prod_{j<k=1}^{p} da_{ij}. \qquad (167)$$

Correlations as angles

56 In the case of a single variable we represent a set of n values $x_1, x_2, \ldots, x_n$ as a point P_1 in n dimensions, or equivalently as a vector OP_1 where O is the origin. Similarly we can represent a set of n values of a p-variate, specified by

$$(x_{ij}), i = 1, 2, \ldots, n, j = 1, 2, \ldots, p,$$

as p points $P_1, \ldots, P_p$ or as p vectors $OP_1, \ldots, OP_p$. In effect this is what we did in regarding spaces as "superposed" in deriving the Wishart distribution. Just as in the one-dimensional case our distribution is represented by a density of points, so in the

p-dimensional case we may picture a density of sets of p points or vectors.

The length of the vector OP_j is given by

$$OP_j^2 = \sum_{i=1}^{n} x_{ij}^2.$$

If we transfer O to the mean of the n values, that is to say, to the point whose jth coordinate is $\sum_{i=1}^{n} x_{ij}/n = \bar{x}_j$, the length becomes

$$\begin{aligned} OP_j^2 &= \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2 \\ &= n \operatorname{var} x_j, \end{aligned} \tag{168}$$

where $\operatorname{var} x_j$ is the variance of the variate x_j in the set of n values.

Furthermore

$$\begin{aligned} \cos P_j OP_k &= \frac{\sum_{i=1}^{n} (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)}{\{\Sigma (x_{ij} - \bar{x}_j)^2 \Sigma (x_{ik} - \bar{x}_k)^2\}^{\frac{1}{2}}} \\ &= \text{correlation between } x_j \text{ and } x_k, \text{ say } r_{jk}. \end{aligned} \tag{169}$$

Hence the variance and correlation structure of the p variables (in this set of n) is represented by lengths and angles.

57 If $r_{ij} = 1$ then the angle between the vectors is zero and one variate is a linear function of the other. If $r_{ij} = 0$ the variates are uncorrelated (in this set of n). It does not follow that the random variables from which they emanate are independent, for two reasons: first, two variables can be uncorrelated without being independent; second, the observed set of n need not, and in general will not, exactly reproduce the parent value of the correlation. We may nevertheless conceive of the $(n-1)$-flat orthogonal to OP_j as containing variation which is uncorrelated with x_j. In the particular case when the variation is multivariate-normal zero correlation implies independence.

Thus, a knowledge of x_i imposes no constraint on the variation of the other variables, which can vary in the $(n-1)$-flat just as if there were no x_i.

58 Consider three variables corresponding to OP_1, OP_2, OP_3 and project OP_1 and OP_2 on the $(n-1)$-flat perpendicular to OP_3. If the projected points are Q_1 and Q_2 the correlation between x_1 and x_2 in that flat is measured by the cosine of the angle Q_1OQ_2. This is written $r_{12.3}$ and is interpreted as a measure of the relationship between x_1 and x_2 independently of their dependence on x_1. Its magnitude, from section 39, is given by

$$r_{12.3} = \frac{r_{12}-r_{13}r_{23}}{\{(1-r_{13}^2)(1-r_{23}^2)\}^{\frac{1}{2}}}. \tag{170}$$

It is called a partial correlation.

Likewise, if we project on to a flat orthogonal to a further vector, say x_4, we derive a partial correlation of second order, $r_{12.34}$, expressing the relationship between x_1 and x_2 "when their dependence on x_3 and x_4 has been eliminated". And so on.

59 For normal variation the surfaces of constant density for x_i are hyperspheres centred at O. Parental correlations between the variables are represented by constraints on angles between the vectors. The whole representation, however, is invariant under a rotation of axes. Let us make such a rotation such that OP_3 is one of them. If we project the whole complex on to the S_{n-1} which is orthogonal, we see that hyperspheres of constant density remain so and total correlations are replaced by partial correlations. Thus the sampling distribution of $r_{12.3}$ will be exactly the same as that of r_{12} except that we are in one lower dimension. The reader may care to try to prove this proposition without geometrical reasoning; there is no more convincing demonstration of the power of the geometrical approach.

Regression and multiple correlation

60 Suppose now that each sample-member, in addition to bearing values of $x_1, \ldots, x_p$, also bears the values of a variate y; and that we are interested in the dependence of y on the x's.

The points P_1 to P_p, together with the origin, define a containing space of p dimensions. This makes a certain angle (unique) with the vector OY. The cosine of this angle is called the coefficient of multiple correlation of y on $x_1, x_2, \ldots, x_p$. It is usually denoted by R.

If y is a linear function of the x's, say

$$y = \beta_1 x_1 + \ldots + \beta_p x_p, \tag{171}$$

then OY lies entirely in the space of x's and R is unity. If y does not depend on the x's OY is orthogonal and $R = 0$. R may assume any value in the range $0 \leqslant R \leqslant 1$. That it is a correlation may be seen from the fact that it is the cosine of an angle between OY and the orthogonal projection of OY on the space of x's. It measures the closeness with which the linear representation of y in terms of the x's is realized. It is evident intuitively (and readily checked) that R is the cosine of the least angle of the family between OY and any vector OX in the x-space, i.e. R is a maximum. This consideration allows us to determine the β's of equation (171). In fact we have to find them so as to maximize R between y and a linear function $\sum_{j=1}^{p} \beta_j x_j$. This is equivalent to minimizing the square of the distance from Y to its projection on the x-space, i.e. to minimizing the length of the vector $y - \Sigma\beta_j x_j$, i.e. to minimizing

$$\sum_{\text{sample}} \left(y - \sum_{j=1}^{p} \beta_j x_j\right)^2 \tag{172}$$

which leads us to the familiar equations of least-squares regression theory.

61 R is unaffected by the length of the vectors. Let us take them as unity. Consider the content of the parallelotope determined by $OYP_1, \ldots, P_p$. Although we have defined these vectors in n dimensions they effectively define a space of $p+1$ dimensions. Take a coordinate system in this space and let the coordinates of P_j be $\xi_{ij}, i = 1, \ldots, p$ and those of Y be η_1. The content of the parallelotope is then

$$C = \begin{vmatrix} \eta_{11} & \xi_{11} & \cdots & \xi_{p1} \\ \eta_2 & \xi_{12} & \cdots & \xi_{p2} \\ \eta_3 & \xi_{13} & \cdots & \xi_{p3} \\ . & . & \cdots & . \\ \eta_{1p} & \xi_{1p} & \cdots & \xi_{pp} \end{vmatrix}. \tag{173}$$

Multiplying this by its transpose and writing r_{ij} for the correlation between ξ_i and ξ_j and r_{yi} for that between Y_i and x_i, we have

$$C^2 = \begin{vmatrix} 1 & r_{y1} & r_{y2} & \cdots & r_{yp} \\ r_{y1} & 1 & r_{12} & \cdots & r_{1p} \\ . & r_{12} & 1 & \cdots & r_{2p} \\ . & . & . & \cdots & . \\ r_{yp} & r_{p1} & . & \cdots & 1 \end{vmatrix}. \tag{174}$$

The r's are invariant under coordinate transformation and can, therefore, equally well be calculated from the original x's.

Now if θ is the angle between OY and the space of x's, the content C is $\sin\theta$ times the content of the parallelotope determined by $O, P_1, \ldots, P_p$, which is the minor of the top left-hand element in (174). Thus we have

$$1 - R^2 = \sin^2\theta = \frac{W}{W_{11}} \tag{175}$$

where W is the matrix whose determinant is given in (174) and W_{11} is the minor of the element in the top left-hand corner.

62 We may also consider the content as obtained in a different way: by constructing the plane area OYP_1; then multiplying it by the length of the perpendicular from P_2 on to that space; then by the length of the perpendicular from P_3 on to the space OYP_1P_2, and so on. These perpendiculars are sines of angles each of which is a partial correlation. Thus we find

$$1-R^2 = (1-r_{y1}^2)(1-r_{y2.1}^2)(1-r_{y3.12}^2)\dots(1-r_{yp.12\dots(p-1)}^2), \tag{176}$$

a decomposition of $1-R^2$ which can be carried out in a number of different ways according to the order in which we select the x-vectors.

63 Consider now the distribution of R for samples of n from a multivariate normal distribution in which the variate represented by OY is independent of the other variates. This means that in repeated sampling the vector OY will be randomly orientated with respect to the S_p of x's. We therefore lose no generality by supposing the x's fixed, namely the space S_p fixed, and considering the distribution of the angle made with it by a randomly orientated vector OY; or equivalently of the angle determined by a point Y moving on a hypersphere of unit radius. The angle θ is the one between OY and, say, OZ in the S_p, this being the minimum possible angle for varying positions of Z.

If OZ and θ are fixed, Y varies on the surface of a hypersphere in $n-p-1$ dimensions with content proportional to $(\sin\theta)^{n-p-2}$. O may vary independently in an S_p on a hypersphere with surface content $(\cos\theta)^{p-1}$. For θ, therefore, we have

$$dF \propto (\sin\theta)^{n-p-2}(\cos\theta)^{p-1}\,d\theta.$$

Putting $R = \cos\theta$ we find

$$dF \propto R^{p-1}(1-R^2)^{\frac{1}{2}(n-p-3)}\,dR$$

or, expressing the distribution in terms of R^2 and evaluating the constant,

$$dF = \frac{1}{B\{\frac{1}{2}p, \frac{1}{2}(n-p+1)\}} (R^2)^{\frac{1}{2}(p-2)} (1-R^2)^{\frac{1}{2}(n-p-3)} \, d(R^2). \tag{177}$$

64 If the parent value of R is not zero a more complicated argument is necessary, but it still hinges essentially upon geometrical considerations. For the details of the derivation see Kendall and Stuart, Vol. 1, p. 339.

Canonical correlations

65 In generalization of the regression of one variable y on a set of others $x_1, \ldots, x_p$, we may consider the relationship between a set of q variables $y_1, y_2, \ldots, y_q$ on $x_1, x_2, \ldots, x_p$. We shall not develop the subject here, but one result of some importance is worth mentioning in order to relate it to the discussion of section **28** concerning angles between flats.

We may, in fact, regard the y's as corresponding to q vectors in an S_q and the x's corresponding to p vectors in an S_p. The correlation relationships between the two spaces depend on angles between vectors in them. We may state, without formal proof, the following propositions, which bear an obvious analogy to the results of section **28**.

It is possible to find linear transformations of x's to ξ's and y's to η's such that (1) all the ξ's are independent (i.e. the axes in the p-space are orthogonal); (2) all the η's are independent (all the axes in the q-space are orthogonal); (3) all the ξ's are uncorrelated with all the η's except for $p < q$ correlations between $(\xi_1, \eta_1)(\xi_2, \eta_2), \ldots, (\xi_p, \eta_p)$; and these are stationary values of the possible correlations between a vector in one space and a vector in the other. These correspond to the canonical angles between the spaces. They express, in the simplest possible form, the relationship between the two sets of variables and are known as "canonical correlations".

Component analysis

66 Let us revert to the case of a set of p vectors in S_n. They determine, as we have remarked, an S_p which may be considered as immersed in the S_n. In this S_p we can find linear transformations to new variables $y_1, \ldots, y_p$ which are orthogonal. In fact, we can do so in more ways than one. Let us choose as y_1 the axis for which the corresponding variable $\Sigma l_{ik} x_k$ has the greatest variance, i.e. such that the sum of projections of the vectors on to it is a maximum. Measuring, as usual, from an origin at the means and taking the vector to have unit length, we have to minimize

$$(\Sigma l_i x_i)' (\Sigma l_i x_i)$$

subject to $\Sigma l_i^2 = 1$. Now this is the same problem that we considered in section **34**. We find the unconditioned maximum of

$$\Sigma' l_i l_j r_{ij} - \lambda \Sigma l_i^2$$

leading to

$$|r - \lambda I| = 0, \tag{178}$$

where r is the correlation matrix $\mathbf{r}$.

Thus the process of finding the major axes of the quadratic form $\Sigma r_{ij} X_i X_j$ is the same as finding the orthogonal vectors, linear in the x's, with stationary variances. These vectors are called principal components. For an extended account of them see Kendall's *Multivariate Analysis*, in this Series.

67 The above examples by no means exhaust the applications of n-dimensional geometry in statistics. Equally effective use of it can be made in other branches of multivariate analysis. In statistical estimation, however, it is differential geometry which can make the greatest contribution. Notwithstanding some applications of geometry in the large to problems of linear estimation (cf., for example, Durbin and Kendall, 1951), estimation in general has not yet been explored from the geometrical viewpoint. The reader who is interested may refer to papers by Huzurbazar (1949) and Rao (1960), who develop some suggestive results concerning the curvature of the likelihood surface from the point of view of Riemannian geometry.

REFERENCES

DURBIN, J. and KENDALL, M. G. (1951). The geometry of estimation. *Biometrika*, **38**, 150.

HUZURBAZAR, V. S. (1949). On a property of distributions admitting sufficient statistics. *Biometrika*, **36**, 71.

KENDALL, M. G. (1957). *A Course in Multivariate Analysis*. C. Griffin and Co., London.

KENDALL, M. G. and STUART, A. (1958). *The Advanced Theory of Statistics*, Vol. 1. C. Griffin and Co., London.

RAO, B. RAJA (1960). A formula for the curvature of the likelihood surface of a sample drawn from a distribution admitting sufficient statistics. *Biometrika*, **47**, 203.

SOMMERVILLE, D. M. Y. (1958, republication of edition of 1929). *An Introduction to the Geometry of N Dimensions*. Dover Publications, New York; Constable and Co., London.