AN INTRODUCTION TO INFINITELY MANY VARIATES

ENDERS A. ROBINSON, PH.D.

Assistant Professor of Mathematics
in the University of Wisconsin

BEING NUMBER SIX OF

GRIFFIN'S STATISTICAL MONOGRAPHS & COURSES

EDITED BY

M. G. KENDALL, SC.D.

QUIS SEPARABIT NOS?

1820

CG&C^{O}LTD

CHARLES GRIFFIN & COMPANY LIMITED
LONDON

CHARLES GRIFFIN & COMPANY LIMITED
42 DRURY LANE, LONDON, WC 2

First published in 1959

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

PREFACE

During the past twenty years abstract mathematical concepts have come into general use in statistical research. Advanced methods make it possible conveniently to treat cases that involve an infinite number of random variables. Undergraduate courses in mathematical statistics and probability theory, however, are usually confined to finite cases. Consequently, those students who go on to graduate work and research are faced with the problem of reorientating their undergraduate work in terms of the more general mathematical setting.

There is a growing feeling that an undergraduate should be exposed to advanced techniques early in his statistical training—that is, in his first or second course in mathematical statistics. In this way the student would be able right away to co-ordinate the more concrete methods of the finite case with the more abstract methods of the infinite case.

This book gives a concise presentation of probability theory, limit theorems, and stationary stochastic processes. The underlying mathematical theme of the book is presented in Chapter I: the metric space and the measure space. In Chapter II, probability theory is presented as measure theory, which makes possible the rigorous treatment of an unlimited number of random trials. Chapter III deals with sums of an indefinite number of independent random variables; and the resulting limit theorems, such as the Central Limit Theorem, are corner-stones of statistical theory. The concept of convergence in metric space plays a predominant role in Chapter III. The rest of the book presents a generalization of the regression analysis of a finite number of random variables to the infinite number of the stationary stochastic process. Again the concept of convergence in metric space plays a predominant role.

This book can be used in undergraduate statistics courses as a supplement to standard texts that require only calculus. Some preliminary knowledge of measure and integration such as is found in Cramér's *Mathematical Methods of Statistics* would be helpful, but is not necessary if some class time is spent on this subject. Learning mathematics has points in common with learning a new language. A student can plunge into this book and think in terms of such concepts as measure, the Lebesgue–Stieltjes integral, metric space, and Hilbert space, before he has had formal courses on these subjects.

In the two-semester first course in statistics at the University of Wisconsin, intended for students who have had one or two years of calculus, a standard text is used. During the first part of the year 1958–9, in conjunction with the text-book, the first few chapters of this volume were presented to a class which was especially receptive and interested. Most of the students are majoring in the physical, social, and biological sciences, mathematics, and engineering. These fields, in common, are feeling the impact of modern statistics. Because this book is concerned with the mathematics of statistics, we used during the latter part of the year Quenouille's *The Fundamentals of Statistical Reasoning* (this Series, No. 3) to introduce the principles of the science of statistics.

So that overall unity would not be lost in details and lengthy description, many of the proofs and applications are either outlined or left as exercises. Throughout the book the student is encouraged to go to the references, which is the key toward making him research-minded.

E. A. Robinson

Madison, Wisconsin, U.S.A.
May, 1959

CONTENTS

VII STATIONARY PROCESSES

CHAPTER I

METRIC AND MEASURE

1.1 Metric space

An *abstract mathematical space* is any collection of things, such as functions, or sequences, or geometrical points. Each thing is called an element, or point, of the abstract space.

A metric space is an abstract space for which a metric, or distance between two points, is defined.

Definition 1.10 A *metric space* is a set $\mathbf{K}$ of points for which the distance between any two points x, y is given by a single-valued, non-negative, real function $d(x,y)$, which satisfies the metric conditions:

(1) $d(x,y) = 0$ if and only if $x = y$ (equivalence axiom);

(2) $d(x,y) = d(y,x)$ (the axiom of symmetry);

(3) $d(x,y)+d(y,z) \geqslant d(x,z)$ (the triangle axiom).

Example 1.10 Let $\mathbf{E}_1$ be the space with points x the real numbers $x = u$ where u is real. If $x = u$ and $y = v$ where u and v are real numbers, let the distance function be $d(x,y) = |v-u|$. Then $\mathbf{E}_1$ is a metric space and is called *Euclidean one-tuple space.* Let $\mathbf{E}_n$ be the space with points x the ordered n-tuples $x = (u_1, \ldots, u_n)$ of real numbers $u_1, \ldots, u_n$. If $x = (u_1, \ldots, u_n)$ and $y = (v_1, \ldots, v_n)$, let the distance function be

$$d(x,y) = \left[\sum_{i=1}^{n} (v_i - u_i)^2\right]^{\frac{1}{2}}$$

Then $\mathbf{E}_n$ is a metric space and is called *Euclidean n-tuple space.*

Example 1.11 Let $\mathbf{E}_\infty$ be the space with points x the sequences $x = (u_1, u_2, \ldots)$ of real numbers $u_1, u_2, \ldots$, where $\sum_{i=1}^{\infty} u_i^2$ is finite (i.e. each point must be a finite distance from the origin $(0, 0, \ldots)$). If $x = (u_1, u_2, \ldots)$ and $y = (v_1, v_2, \ldots)$, let the distance function be

$$d(x,y) = \left[\sum_{i=1}^{\infty} (v_i - u_i)^2\right]^{\frac{1}{2}}$$

Then $\mathbf{E}_\infty$ is a metric space and is called a *Euclidean sequence space.*

Example 1.12 Let $\mathbf{C}^2$ be the space with points x the real continuous functions $x = u(t)$ of the real variable t. If $x = u(t)$ and $y = v(t)$, let the distance function be

$$d(x,y) = \left[\int_a^b \{v(t) - u(t)\}^2\, dt\right]^{\frac{1}{2}}$$

Then $\mathbf{C}^2$ is a metric space and is called the *space of continuous functions with quadratic metric.*

An ϵ-neighbourhood of a point x in a metric space $\mathbf{K}$ is the set of all points y of $\mathbf{K}$ that satisfy the condition $d(y,x) < \epsilon$. Let $\mathbf{S}$ be a subset of the metric space $\mathbf{K}$. A point x is called a *contact point* of the set $\mathbf{S}$ if every neighbourhood of x contains at least one point of $\mathbf{S}$. The set of all contact points of $\mathbf{S}$ is called the *closure* $[\mathbf{S}]$ of $\mathbf{S}$. It follows that every set is contained in its closure: $\mathbf{S} \subset [\mathbf{S}]$. The point x is called a *limit point* of the set $\mathbf{S}$ if every neighbourhood of x contains an infinite number of points of $\mathbf{S}$. The point x of the set $\mathbf{S}$ is called an *isolated point* of $\mathbf{S}$ if x has a neighbourhood that does not contain any points of $\mathbf{S}$ different from x. It can be shown that every contact point of $\mathbf{S}$ is either a limit point of $\mathbf{S}$ or an isolated point of $\mathbf{S}$. Hence the closure $[\mathbf{S}]$ in general consists of (1) isolated points of $\mathbf{S}$, (2) limit points of $\mathbf{S}$ that belong to $\mathbf{S}$, and (3) limit points of $\mathbf{S}$ that do not belong to $\mathbf{S}$. We may obtain the closure $[\mathbf{S}]$ by taking the union of the set $\mathbf{S}$ and the set of all the limit points of $\mathbf{S}$.

Let $x_1, x_2, \ldots$ be a sequence of points in the metric space $\mathbf{K}$. We say that the sequence $\{x_n\}$ *converges to the element* x if the numbers $d(x_1, x), d(x_2, x), \ldots$ converge to zero, and shall denote such convergence in metric space by : $x_n \Rightarrow x$ as $n \to \infty$. The element x is called the *limit* of the sequence $\{x_n\}$.

The concept of contact point and limit point and the concept of limit are related in the following way. The point x is a contact point of $\mathbf{S}$ if and only if x is the limit of a sequence $x_1, x_2, \ldots$ of points of $\mathbf{S}$. The point x is a limit point of $\mathbf{S}$ if and only if x is the limit of a sequence $x_1, x_2, \ldots$ of distinct points of $\mathbf{S}$.

A set $\mathbf{S}$ is said to be *closed* if it coincides with its closure $[\mathbf{S}]$; i.e. if $\mathbf{S}$ contains all its limit points. The closure $[\mathbf{S}]$ of a set $\mathbf{S}$ is closed, and $[\mathbf{S}]$ is the smallest closed set that contains $\mathbf{S}$.

The set $\mathbf{S}$ is called *everywhere dense* in the metric space $\mathbf{K}$ if its closure $[\mathbf{S}]$ coincides with the entire space $\mathbf{K}$. A metric space $\mathbf{K}$ is called *separable* if there exists a countable set $\mathbf{S}$ (i.e. a sequence $x_1, x_2, \ldots$) of points in $\mathbf{K}$ that is everywhere dense in $\mathbf{K}$. The metric space $\mathbf{E}_1$ (Example 1.10) is separable because the rational numbers form an everywhere dense countable set in $\mathbf{E}_1$. The metric space $\mathbf{E}_n$ (Example 1.10) is separable because the set of all n-tuples $(u_1, u_2, \ldots, u_n)$ with rational co-ordinates $u_1, u_2, \ldots, u_n$ is an everywhere dense countable set. The metric space $\mathbf{E}_\infty$ (Example 1.11) is separable because the set of all sequences in each of which all terms are rational, and only a finite (but arbitrary) number of terms is distinct from zero, forms an everywhere dense countable set in $\mathbf{E}_\infty$. The metric space $\mathbf{C}^2$ (Example 1.12) is separable because the set of all polynomials with rational coefficients is countable and everywhere dense.

A sequence $\{x_n\}$ of points of a metric space $\mathbf{K}$ is called a *fundamental sequence* if it satisfies the Cauchy criterion, i.e. if for arbitrary $\epsilon > 0$ there exists an N such that $d(x_n, x_m) < \epsilon$ for all $m, n > N$. It follows from the triangle axiom that every convergent sequence is fundamental. For if $\{x_n\}$ converges to x, then for $\epsilon > 0$, there is an N such that $d(x_n, x) < \epsilon/2$ for all $n > N$. Then

$$d(x_n, x_m) \leqslant d(x_n, x) + d(x_m, x) < \epsilon$$

for arbitrary $n, m > N$.

If every fundamental sequence in the metric space $\mathbf{K}$ converges to an element in $\mathbf{K}$, then $\mathbf{K}$ is said to be *complete*. It is a theorem in analysis that the space of real numbers (Euclidean one-tuple space) is complete, from which it follows that Euclidean n-tuple space $\mathbf{E}_n$ is complete (Example 1.10). It can also be shown that the metric space $\mathbf{E}_\infty$ (Example 1.11) is complete. Nevertheless the metric space $\mathbf{C}^2$ (Example 1.12) is not complete. For example, the sequence of continuous functions $x_n(t) = \arctan nt$ is fundamental, but does not converge to a continuous function.

1.2 Measure space

Let Ω be an abstract space of points ω. A point-set, denoted by A, B, etc., is a set of points ω.

Definition 1.20 A class $\mathbf{F}$ of point-sets is called a *field* if

(1) the whole space Ω is an element of $\mathbf{F}$;
(2) the complement, $\Omega - A$, of any element A of $\mathbf{F}$ is also an element of $\mathbf{F}$;
(3) finite unions and intersections of elements of $\mathbf{F}$ are also elements of $\mathbf{F}$.

A field is called a *Borel field* if, in addition, countable unions and intersections of elements of $\mathbf{F}$ are also elements of $\mathbf{F}$.
A measure may be defined on a Borel field as follows.

Definition 1.21 A *measure* is a real, non-negative set function $\mu(A)$ for which

(1) the domain of definition is a Borel field of sets;
(2) for any finite or countable number of disjoint sets A_n, the measure of their union is the sum of their measures; that is, $\mu(\cup A_n) = \Sigma\mu(A_n)$.

The sets A belonging to the domain of definition are called μ-measurable sets, or simply, *measurable sets.* A *measure space,* denoted by $(\Omega, \mathbf{F}, \mu)$, is formed by the space Ω, the Borel field $\mathbf{F}$ of measurable sets in this space, and the measure μ defined on this Borel field.

If a class $\mathbf{F}_0$ of point-sets of Ω do not form a Borel field, then a measure cannot be defined for $\mathbf{F}_0$. In this case, we consider all Borel fields that contain the sets of $\mathbf{F}_0$. It can be shown that the intersection of all these Borel fields is itself a Borel field, and thus is the minimal Borel field containing $\mathbf{F}_0$. This minimal Borel field, denoted by $\mathbf{B}(\mathbf{F}_0)$, is called the *Borel closure* of $\mathbf{F}_0$ with respect to Ω.

Example 1.20 Let the space Ω be *one-dimensional Cartesian space*, i.e. the real line $\mathbf{E}_1$, so that the points ω are the real numbers u. Let $\mathbf{F}_0$ be the class of finite unions of right semi-closed intervals $\{a < u \leqslant b\}$. Then the sets of $\mathbf{B}(\mathbf{F}_0)$ are called the *linear Borel sets*. Let F be a distribution function on the real line; that is, F is a monotone, non-decreasing function of the real variable u, continuous on the right, and with $F(-\infty) = 0$ and $F(\infty) = 1$. If A is the finite union of non-overlapping right semi-closed intervals $\{a_i < u \leqslant b_i\}$, $i = 1, 2, \ldots, n$, define

$$\mu(A) = \sum_{i=1}^{n} [F(b_i) - F(a_i)]$$

The domain of definition of μ can be extended to $\mathbf{B}(\mathbf{F}_0)$ so that μ is a measure on the linear Borel sets. The measure μ is called the *Lebesgue–Stieltjes measure* for the distribution function F. Because $\mu(\mathbf{E}_1) = 1$, μ is a bounded measure.

Example 1.21 Let the space Ω be 2-*dimensional Cartesian space*, i.e. the plane $\mathbf{E}_2$, so that the points ω are the ordered pairs (u_1, u_2) of real numbers. Let $\mathbf{F}_0$ be the class of finite unions of right semi-closed rectangles. Then the sets of $\mathbf{B}(\mathbf{F}_0)$ are the *two-dimensional Borel sets*. Let F be a distribution function on the plane; that is, F is monotone, non-decreasing, and continuous on the right for each variable u_1, u_2, and furthermore

$$F(-\infty, u_2) = F(u_1, -\infty) = 0, \quad F(u_1, \infty) = F(\infty, u_2) = 1$$

and for $a_1 \leqslant b_1, a_2 \leqslant b_2$,

$$\Delta F = F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2) \geqslant 0$$

If A is a right semi-closed rectangle,

$$A = \{a_1 < u_1 \leqslant b_1, a_2 < u_2 \leqslant b_2\}$$

define $\mu(A) = \Delta F$. Also if A is a finite union of disjoint right semi-closed rectangles, define $\mu(A)$ as the corresponding sum of the measures of the individual rectangles. The domain of definition of μ can then be extended to $\mathbf{B}(\mathbf{F}_0)$ so that μ is a measure on the two-dimensional Borel sets, and is called the *Lebesgue–Stieltjes measure* for the distribution function F.

Example 1.22 In a way similar to Examples 1.20 and 1.21, Lebesgue–Stieltjes measure can be defined for a distribution function in *n-dimensional Cartesian space* $\mathbf{E}_n$.

A *transformation* or *mapping* is a single-valued function T that maps points ω of one set to points ω' of another set; that is, $\omega' = T(\omega)$. The *image* of the set A, denoted by $T(A)$, is the set of points ω' that are of the form $T(\omega)$ for some ω in A. The *inverse image* of the set A', denoted by $T^{-1}(A')$, is the set of all points ω that are mapped onto A' by $T(\omega)$. The function T is called a *measurable mapping* of the measure space $(\Omega, \mathbf{F}, \mu)$ into the measure space $(\Omega', \mathbf{F}', \mu')$ if the inverse image of every measurable set of $\mathbf{F}'$ is a measurable set of $\mathbf{F}$. The measure μ', of course, can be defined on $\mathbf{F}'$ in any way. If we define the measure μ' of a set A' of $\mathbf{F}'$ to be

$$\mu'(A') = \mu[T^{-1}(A')]$$

then μ' is called the measure on $\mathbf{F}'$ induced by the mapping from the measure μ on $\mathbf{F}$; that is, μ' is the *induced measure.*

A *measurable function* $x(\omega)$ is a measurable mapping of the measure space $(\Omega, \mathbf{F}, \mu)$ into the real line. The *Lebesgue integral* of a measurable function x over the measurable set A, which is an element of the domain of definition $\mathbf{F}$ of the measure μ, is denoted by

$$\int_A x(\omega)\, \mu(d\omega)$$

This integral is finite if and only if

$$\int_A |x(\omega)|\, \mu(d\omega)$$

is finite, in which case we say that $x(\omega)$ is *integrable* on A.

If two different measures μ and ν are assigned to the same Borel field $\mathbf{F}$, then ν is said to be *absolutely continuous* with respect to μ if $\mu(A) = 0$ implies $\nu(A) = 0$ for every set A of $\mathbf{F}$. Then the following theorem is true.

Theorem 1.20 (*the Radon–Nikodym Theorem*) If ν is absolutely continuous with respect to μ, then there exists a measurable function f such that

$$\nu(A) = \int_A f(\omega)\, \mu(d\omega)$$

for every measurable set A of $\mathbf{F}$. Moreover, the function f is unique, except on a set of μ-measure zero.

CHAPTER II

PROBABILITY

2.1 Probability space

André Kolmogorov (1933) established the mathematical foundations of probability theory on a postulational basis with the

Normed Measure Postulate: Probability $P(A)$ is a measure satisfying the additional normalization condition $P(\Omega) = 1$.

There is general agreement on the correctness of Kolmogorov's postulate in its characterization of the mathematical properties of the concept that we call "probability". As with all other scientific developments, it is expected that refinements ultimately can and will be made. In fact, Kolmogorov himself has recently introduced the postulate of normed "perfect" measure as an improved characterization of probability.

Whereas the mathematical properties of probability are accepted by most people, the more philosophical questions concerning the interpretation of the mathematical concepts are very much open and highly controversial. The most popular view on the interpretation of probability is the relative frequency theory associated chiefly with the name of Richard von Mises. Gaining in popularity, however, are the classical theories, so-called because of the early work of Laplace.

Under the Kolmogorov postulate, the measure theoretic concepts of Chapter I become probability concepts. Thus measure space $(\Omega, \mathbf{F}, \mu)$ becomes, in probability terminology, *probability space* $(\Omega, \mathbf{F}, P)$. A point ω represents an *elementary event*, and the aggregate of all these points is the *sample space* Ω. A measurable set A is called a *random event*, and its measure $P(A)$ is the *probability* of the occurrence of A. A property that holds for all points of Ω, except for points in a set of probability zero, is said to hold with *probability* 1, or *almost everywhere*.

A *random variable* $x(\omega)$ is a measurable function defined over the sample space Ω; that is, a random variable is an assignment of a real number to each sample point ω. A complex random variable is an ordered pair of two (real) random variables. A synonym for random variable is *variate*.

The *expectation*, or *mean value*, $E(x)$ of a variate x is given by the integral

$$E(x) = \int_{\Omega} x(\omega) P(d\omega)$$

The *variance* $D^2(x)$ of a variate x is given by $E[(x-E(x))^2]$. The *standard deviation* $D(x)$ is the positive square root of the variance.

Example 2.10 A fair coin is tossed once. The simplest suitable space Ω consists of 2 points, represented by the number ξ, where ξ is either 1 (for head) or 0 (for tail). The measurable sets are the null set, $\{0\}$, $\{1\}$, and $\{0, 1\}$, to which we assign the measures 0, 0·5, 0·5, and 1, respectively. That is, the measure of a set is 1/2 times the number of points which lie in the set.

Example 2.11 A fair coin is tossed twice. The simplest suitable space Ω consists of 4 points ω which are the ordered pairs (ξ_1, ξ_2), where each ξ_1, ξ_2 can be either 1 or 0. All Ω point-sets are measurable, and the measure assigned to each set is equal to 1/4 times the number of points which lie in the set.

Example 2.12 A fair coin is tossed n times. An appropriate space Ω consists of 2^n points ω which are the ordered finite sequences $(\xi_1, \xi_2, \ldots, \xi_n)$, where each ξ_j can be either 1 or 0. All Ω point-sets are measurable, and the measure assigned is equal to $1/2^n$ times the number of points ω which fall in the set.

Example 2.13 A fair coin is tossed an unlimited number of times. A suitable space Ω consists of all points ω which are infinite sequences $(\xi_1, \xi_2, \xi_3, \ldots)$, where each ξ_j is either 1 or 0. Call the set of all sequences $(1, \xi_2, \xi_3, \ldots)$ the random event that "head" comes up on the first throw, and assign probability 1/2 to this set. Similarly, assign measure 1/2 to the set of sequences $(0, \xi_2, \xi_3, \ldots)$. Call the set of all sequences $(1, 0, \xi_3, \xi_4, \ldots)$ the random event that "head" comes up on the first throw and "tail" on the second, and assign measure 1/4 to the set. More generally, the set of all sequences $(a_1, a_2, \ldots, a_n, \xi_{n+1}, \xi_{n+2}, \ldots)$ is called the event that the faces $a_1, a_2, \ldots, a_n$ (each $a_j = 1$ or 0) come up in the first n tosses, to which event is assigned a probability $1/2^n$. Thus we have defined a measure of a sample space with an infinite number of points ω, where each point ω is a sample infinite sequence.

2.2 Distribution functions

A variate $x(\omega)$, by definition, maps the sample space into the real line $\mathbf{E}_1$. The induced measure P_x on the real line $\mathbf{E}_1$ resulting from this mapping is called the (one-dimensional) *probability distribution* of the variate $x(\omega)$. The domain of definition of the probability distribution P_x contains all (linear) Borel sets of $\mathbf{E}_1$.

Define the function $F(u)$ of the real variable u by

$$F(u) = P\{x(\omega) \leqslant u\} = P_x((-\infty, u])$$

where $(-\infty, u]$ is the set consisting of all points on the real line lying to the left of and including u. It follows that $F(u)$ is monotone, continuous on the right, non-decreasing, and satisfies $F(-\infty)=0$ and $F(\infty)=1$. Hence F is a distribution function on the real line, as defined in Example 1.20, and so is called the *distribution function* of the variate $x(\omega)$.

As we have seen in Example 1.20, the distribution function $F(u)$ defines a Lebesgue–Stieltjes measure on the real line. This Lebesgue–Stieltjes measure is the (one-dimensional) probability distribution P_x. The (one-dimensional) probability distribution P_x is uniquely determined by its distribution function F.

Since the variate $x(\omega)$ is a mapping of points ω of the sample space Ω to real numbers u of the real line $\mathbf{E}_1$, i.e. $x(\omega)=u$, and since under this mapping the measure P on the sample space Ω induces a measure P_x on the real line $\mathbf{E}_1$, it can be shown that

$$E(x)=\int_\Omega x(\omega)P(d\omega)=\int_{\mathbf{E}_1} uP_x(du)$$

If $F(u)$ is the distribution function of the probability distribution P_x, and if $(a,b]$ is the half-open interval $a<u\leqslant b$ on the real line $\mathbf{E}_1$, then

$$\int_{(a,b]} uP_x(du)=\int_a^b udF(u)$$

More generally, if the variate $y=h(x)$ is a Borel measurable function of the variate x, then

$$E(y)=\int_\Omega h[x(\omega)]P(d\omega)=\int_{\mathbf{E}_1} h(u)P_x(du)$$

and

$$J=\int_{(a,b]} h(u)P_x(du)=\int_a^b h(u)\,dF(u)$$

If the function $h(u)$ is continuous in the closed interval $a\leqslant u\leqslant b$, then the integral J can be approximated by the *Stieltjes sum*

$$S_n=\sum_{i=1}^n h(u_i)[F(a_i)-F(a_{i-1})],$$
$$a=a_0\leqslant u_1\leqslant a_1\leqslant u_2\leqslant\ldots\leqslant u_n\leqslant a_n=b$$

As $n\to\infty$ and $\max(a_i-a_{i-1})\to 0$, the sums S_n converge to the integral J. We then have

$$E(y)=\int_{\mathbf{E}_1} h(u)P_x(du)=\lim_{\substack{a\to-\infty\\ b\to\infty}}\int_a^b h(u)\,dF(u)=\int_{-\infty}^{\infty} h(u)\,dF(u)$$

provided the limit exists.

Any n variates $x_1,\ldots,x_n$ can be considered as an n-tuple $x=(x_1,\ldots,x_n)$ that maps the sample space Ω into n-dimensional Cartesian space $\mathbf{E}_n$. The induced measure $P_x=P_{x_1,x_2,\ldots,x_n}$ on $\mathbf{E}_n$ resulting from this mapping is called the *(n-dimensional) probability* *distribution* P_x of the n-tuple x, or the *joint or multivariate distribution* $P_{x_1, x_2, \ldots, x_n}$ of the variates $x_1, x_2, \ldots, x_n$. The domain of definition of the probability distribution P_x contains all (n-dimensional) Borel sets of $\mathbf{E}_n$.

Define the function $F(u_1, \ldots, u_n)$, of n real variables $u_1, \ldots, u_n$ by

$$F(u_1, \ldots, u_n) = P\{x_1(\omega) \leqslant u_1, \ldots, x_n(\omega) \leqslant u_n\}$$

The function F so defined is monotone, continuous on the right, and non-decreasing for each variable, with

$$F(u_1, \ldots, u_{j-1}, -\infty, u_{j+1}, \ldots, u_n) = 0$$

$$F(\infty, \ldots, \infty) = 1$$

$$\Delta F = \Delta_{u_1}(a_1, b_1) \ldots \Delta_{u_n}(a_n, b_n) F(u_1, \ldots, u_n) \geqslant 0$$

for $a_i \leqslant b_i$, $i = 1, 2, \ldots, n$. The operator $\Delta_{u_j}(a, b)$ is defined by

$$\Delta_{u_j}(a, b) = F(u_1, \ldots, u_{j-1}, b, u_{j+1}, \ldots, u_n) - F(u_1, \ldots, u_{j-1}, a, u_{j+1}, \ldots, u_n)$$

Hence F satisfies the conditions of a distribution function in n-dimensional Cartesian space $\mathbf{E}_n$ (Example 1.22), and so F is called the *distribution function* of the variates $x_1(\omega), x_2(\omega), \ldots, x_n(\omega)$. It is seen that $\Delta F = P\{a_1 < x_1(\omega) \leqslant b_1, \ldots, a_n < x_n(\omega) \leqslant b_n\}$.

As we have seen in Example 1.22, the distribution function F defines a Lebesgue–Stieltjes measure in $\mathbf{E}_n$. This Lebesgue–Stieltjes measure is the (n-dimensional) probability distribution P_x. The (n-dimensional) probability distribution P_x is uniquely determined by its distribution function F.

Mathematical expectations may be obtained from multivariate probability distributions in a way similar to the single variate case.

2.3 Discrete distributions

The two most useful special cases of distribution functions are the discrete and continuous distributions. A distribution function is called discrete if it can be represented as

$$F(u_1, \ldots, u_n) = \sum_{u^{(k)} \leqslant u} p_k$$

where $u = (u_1, \ldots, u_n)$ and $u^{(k)} = (u_1^{(k)}, \ldots, u_n^{(k)})$. The points $u^{(k)}$ must be finite or countable in number, and the non-negative p_k must add up to one.

In particular, the one-dimensional discrete distributions

$$F(u) = \sum_{k=-\infty}^{n} p_k \qquad (n \leqslant u < n+1)$$

are well known in elementary statistics.

Example 2.30 Suppose a biased coin, with probability p for heads (denoted by 1), and probability $q = 1-p$ for tails (denoted by 0), is tossed n times. A suitable sample space Ω consists of 2^n

sequences $\omega = (\xi_1, \xi_2, \ldots, \xi_n)$, where each $\xi_j\,(j = 1, \ldots, n)$ is either 1 or 0. Assign probability $p^k q^{n-k}$ to each point ω that has exactly k ones in its representation $\omega = (\xi_1, \ldots, \xi_n)$. Let the variate $x(\omega)$ be the function that assigns the integer k to any point ω that contains exactly k ones. That is, by definition $x(\omega) = k$ if and only if $\omega = (\xi_1, \ldots, \xi_n)$ has exactly k of the ξ_i's equal to one. Since there are $\binom{n}{k}$ such points ω, we have

$$P\{x(\omega) = k\} = \binom{n}{k} p^k q^{n-k}, \quad (k = 0, 1, \ldots, n)$$

The distribution of x is known as the *binomial distribution*. The distribution function is

$$F(k) = \sum_{j=0}^{k} \binom{n}{k} p^k q^{n-k}$$

Let $N(A)$ be the number of the integers $0, 1, \ldots, n$ in the Borel set A on the real line. It can be readily checked that $N(A)$ is a measure. By the Radon–Nikodym theorem there must exist a function $f(u)$ for which

$$P(A) = \int_A f(u)\, N(du)$$

which, from the distribution function, is

$$P(A) = \int_A \binom{n}{u} p^u q^{n-u} N(du)$$

The *mean* of a binomial distribution is np, and *variance* is npq.

Example 2.31 Suppose that a lopsided, N-sided die is tossed n times. On each toss, the probability of faces $1, 2, \ldots, N$ turning up is $p_1, p_2, \ldots, p_N$ respectively, where $p_1 + p_2 + \ldots + p_N = 1$. The sample space Ω can be the N^n points $\omega = (\xi_1, \xi_2, \ldots, \xi_n)$ where each ξ_j is either $1, 2, \ldots, N-1$, or N. Assign probability $p_1^{k_1} p_2^{k_2} \ldots p_N^{k_N}$ to each point ω that has exactly k_1 ones, k_2 twos, $\ldots, k_N = n - k_1 \ldots - k_{N-1}$ N's in its representation $(\xi_1, \ldots, \xi_n)$. Let the variate $x_1(\omega)$ be the function of ω that assigns the number k_1 to any point ω with exactly k_1 ones. Let $x_2(\omega)$ be the function that assigns the number k_2 to each ω with k_2 twos, and similarly for $x_3, \ldots, x_N$. Since there are

$$\frac{n!}{k_1!\, k_2! \ldots k_N!} \quad (k_1 + k_2 + \ldots + k_N = n)$$

points ω that contain k_1 ones, k_2 twos, $\ldots, k_N$ N's, we have

$$P\{x_1(\omega) = k_1, x_2(\omega) = k_2, \ldots, x_N(\omega) = k_N\}$$
$$= \frac{n!}{k_1!\, k_2! \ldots k_N!} p_1^{k_1} p_2^{k_2} \ldots p_N^{k_N}$$

This probability measure defines the *multinomial distribution*, a multivariate distribution.

Example 2.32 A biased coin, with probability p for heads (1) and $q = 1-p$ for tails (0), is tossed an indefinite number of times. The points ω of the sample space Ω are chosen to be the sequences $\omega = (\xi_1, \xi_2, \ldots)$ where each ξ_j is equal to 1 or 0. Assign probability p to the point-set of $\omega = (1, \xi_2, \xi_3, \ldots)$, and q to the point-set of $\omega = (0, \xi_2, \xi_3, \ldots)$. For any n, assign probability $p^r q^k$ to the point-set of $\omega = (a_1, a_2, \ldots, a_n, \xi_{n+1}, \ldots)$ where $a_1, a_2, \ldots, a_n$ represents a sequence of r 1's and k 0's in a specific order $(r+k = n)$. Define the variate $x(\omega)$ to be the number of tails preceding the first head. That is, $x(\omega) = 0$ when and only when $\omega = (1, \xi_2, \xi_3, \ldots)$, and $x(\omega) = k$ when and only when $\omega = (0, 0, \ldots, 0, 1, \xi_{k+2}, \ldots)$. Therefore we have

$$P\{x(\omega) = k\} = q^k p \quad (k = 0, 1, 2, \ldots)$$

The distribution of $x(\omega)$ is known as the *geometric distribution.* The mean of x is q/p and variance is q/p^2.

Example 2.33 Continuing Example 2.32, let the variate $y(\omega)$ be the number of tails preceding the rth head. That is, $y(\omega) = 0$ when and only when $\omega = (1, 1, \ldots, 1, \xi_{r+1}, \ldots)$; and $y(\omega) = k$ when and only when $\omega = (a_1, a_2, \ldots, a_{r+k-1}, 1, \xi_{r+k+1}, \ldots)$ with exactly $r-1$ of the a_i's equal to 1, and k of the a_i's equal to 0. There are $\binom{r+k-1}{k}$ such points ω, and the probability of each of them is $p^r q^k$. Therefore we have

$$P\{y(\omega) = k\} = \binom{r+k-1}{k} p^r q^k \quad (k = 0, 1, 2, \ldots)$$

This is known as the *Pascal distribution.* The mean of y is $r\,q/p$, and the variance is $r\,q/p^2$.

Example 2.34 Let $x(\omega)$ be a variate, which for n a non-negative integer,

$$P\{x(\omega) = n\} = \frac{a^n \mathrm{e}^{-a}}{n!} \quad (n = 0, 1, 2, \ldots, a > 0)$$

This probability defines the *Poisson distribution.* The mean and variance of x are each equal to a. Let $N(A)$ be counting measure on the real line; that is, $N(A)$ is the number of non-negative integers in the Borel set A on the real line. Then we have

$$P(A) = \int_A \frac{a^u \mathrm{e}^{-a}}{u!} N(du)$$

2.4 Continuous distributions

A distribution function is called *continuous* if it can be represented as

$$F(u_1, \ldots, u_n) = \int_{-\infty}^{u_1} \cdots \int_{-\infty}^{u_n} F'(v_1, \ldots, v_n)\, dv_1, \ldots, dv_n$$

where $dv_1 \dots dv_n$ represents ordinary n-dimensional Lebesgue measure. The function F' is called the *density function* which by the Radon–Nikodym theorem exists when and only when F is absolutely continuous with respect to Lebesgue measure.

Example 2.40 A variate $x(\omega)$ that is degenerate at a has a *degenerate distribution* $F(u) = 0$ for $u < a$, and $F(u) = 1$ for $u \geqslant a$. The mean of x is a, and the variance of x is zero.

Example 2.41 A variate $x(\omega)$ has a *rectangular* or *uniform distribution* if its density function $F'(u)$ is equal to the constant $1/2b$ on the interval $a-b \leqslant u \leqslant a+b$, and equal to zero outside this interval. The mean of x is a and the variance is $(2b)/3$.

Example 2.42 A variate $x(\omega)$ has a *normal distribution* if

$$F(u) = \frac{1}{\sigma\sqrt{(2\pi)}} \int_{-\infty}^{u} \exp\left\{-\frac{(v-a)^2}{2\sigma^2}\right\} dv$$

The *normal density function* is then

$$F'(u) = \frac{1}{\sigma\sqrt{(2\pi)}} \exp\left\{-\frac{(u-a)^2}{2\sigma^2}\right\}$$

The mean of x is a and the variance is σ^2.

Example 2.43 A variate $x(\omega)$ has an *exponential distribution* if

$$F(u) = 0 \quad \text{for} \quad u < 0$$

$$F(u) = 1 - e^{-au} = \int_0^u a\, e^{-av}\, dv \quad \text{for} \quad u \geqslant 0$$

The exponential density function is then $F'(u) = a\,e^{-au}$ for $u > 0$, and $F'(u) = 0$ for $u \leqslant 0$. The mean of an exponential variate is $1/a$, and the variance is $1/a^2$.

Example 2.44 A variate $x(\omega)$ has a *gamma* (or *Pearson Type III*) *distribution* if it has the density function

$$F'(u) = 0 \quad \text{for} \quad u \leqslant 0$$

$$F'(u) = \frac{a^k}{\Gamma(k)} u^{k-1} e^{-au} \quad \text{for} \quad u > 0$$

The mean of a gamma variate is k/a, and the variance is k/a^2.

Example 2.45 A variate $x(\omega)$ with a χ^2 *distribution* with $2k$ *degrees of freedom* is obtained from the gamma distribution if $a = 1/2$ and $2k$ is an integer. The mean of x is $2k$ and the variance of x is $4k$.

Example 2.46 A variate $x(\omega)$ has a *Cauchy distribution* if it has the density function

$$F'(u) = \frac{1}{\pi}\frac{b}{b^2+(u-a)^2}, \quad b>0$$

The mean and variance of x are not finite.

Example 2.47 A variate $x(\omega)$ has a *Laplace distribution* if it has density

$$F'(u) = \frac{1}{2b}\exp\left\{-\frac{|u-a|}{b}\right\}$$

The mean is a and the variance is $2b^2$.

2.5 Independence

Let $(\Omega, \mathbf{F}, P)$ be a probability space. Random events A_i are called *statistically independent* if

$$P\left\{\bigcap_{i=1}^{n} A_i\right\} = P\{A_1\}P\{A_2\}\ldots P\{A_n\}$$

for every finite subset $A_1, A_2, \ldots, A_n$. Classes of random events are called statistically independent if events selected arbitrarily one from each class are statistically independent.

Suppose we have a finite or infinite set $\mathbf{S}$ of variates x_t. Associate with each random variable x_t a field of Borel sets M_{x_t} on the real line. Then the inverse image $\{x_t^{-1}(M_{x_t})\}$ is a Borel field, say $\mathbf{B}_{x_t}$, of random events in the probability space $(\Omega, \mathbf{F}, P)$, where $\mathbf{B}_{x_t} \subset \mathbf{F}$. In this way we obtain a Borel sub-field $\mathbf{B}_{x_t}$ for each variate x_t. If the sub-fields $\mathbf{B}_{x_t}$ are statistically independent, then the variates x_t are called statistically independent. Equivalently, variates x_t of a set $\mathbf{S}$ are *statistically independent* if and only if

$$P\{x_1(\omega)\in M_1, \ldots, x_n(\omega)\in M_n\} = P\{x_1(\omega)\in M_1\}\ldots P\{x_n(\omega)\in M_n\}$$

for every finite subset $x_1, \ldots, x_n$ of $\mathbf{S}$ and every finite class $M_1, \ldots, M_n$ of Borel sets on the real line. It follows then that the variates x_t are statistically independent if and only if the multivariate probability distribution $P_{x_1 \ldots x_n}$ (or distribution function F) of every finite subset $x_1, \ldots, x_n$ is the product of the individual probability distributions $P_{x_1}, \ldots, P_{x_n}$ (or distribution functions $F_1, \ldots, F_n$); that is,

$$P_{x_1 x_2 \ldots x_n} = P_{x_1}P_{x_2}\ldots P_{x_n}, \quad F(u_1, u_2, \ldots, u_n) = F(u_1)F(u_2)\ldots F(u_n)$$

Let h be a real-valued, Borel-measurable function of a real variable. Then $h(x_t)$ is a function of the variate x_t. Since the inverse images h^{-1} are contained in $\mathbf{B}_{x_t}$, and since the $\mathbf{B}_{x_t}$ are statistically independent if and only if the x_t are statistically independent, it follows that functions of statistically independent variates are statistically independent.

Useful in the study of statistically independent variates is the concept of Cartesian product. If $A_1, A_2, \ldots$ is any sequence of sets, not necessarily subsets of the same space, the *Cartesian product* $A_1 \times A_2 \times \ldots$ is defined to be the set of all sequences $(\omega_1, \omega_2, \ldots)$ where $\omega_1 \in A_1, \omega_2 \in A_2, \ldots$. If $(\Omega_1, \mathbf{F}_1, \mu_1), (\Omega_2, \mathbf{F}_2, \mu_2), \ldots$ is a sequence of measure spaces, then their Cartesian product is a space $\Omega_1 \times \Omega_2 \times \ldots$ with the Borel field $\mathbf{F}_1 \times \mathbf{F}_2 \times \ldots$, on which a bona fide measure μ can be defined. In the special case when we take

$$\mu(A_1 \times A_2 \times \ldots) = \mu_1(A_1)\,\mu_2(A_2)\ldots$$

with $A_1 \in \mathbf{F}_1, A_2 \in \mathbf{F}_2, \ldots$, we call μ the *product measure*, which is sometimes denoted by $\mu = \mu_1 \times \mu_2 \times \ldots$. The measure space

$$(\Omega_1 \times \Omega_2 \times \ldots, \mathbf{F}_1 \times \mathbf{F}_2 \times \ldots, \mu_1 \times \mu_2 \times \ldots)$$

is then called the *product space* of the component measure spaces $(\Omega_i, \mathbf{F}_i, \mu_i)$.

Let $x_1, x_2, \ldots$ be a sequence of variates on a probability space $(\Omega, \mathbf{F}, P)$. Since each x_i assigns values on the real line $\mathbf{E}_1$, let $\Omega' = \mathbf{E}_1 \times \mathbf{E}_1 \times \ldots$ be the Cartesian product of a sequence of real lines $\mathbf{E}_1$, one line for each variate, and let the measure on the Borel sets of each line be the (single-variate) probability distribution of the variate associated with that line. The points ω' of the Cartesian product space Ω' are the n-tuple or sequence

$$\omega' = [x_1(\omega), x_2(\omega), \ldots]$$

If we define the transformation $T(\omega)$ by $\omega' = T(\omega)$, then it follows that $T(\omega)$ is measurable transformation from $(\Omega, \mathbf{F}, P)$ to the product space $(\Omega', \mathbf{F}', P')$. The measure P' induced by $T(\omega)$ on the set $A' = A_1' \times A_2' \times \ldots$ of the product space is the (multivariate) probability distribution $P_{x_1 x_2 \ldots}$ and is given by $P'(A') = P\{T^{-1}(A')\}$. On the other hand, the product measure of the set A' is, by definition,

$$P\{x_1^{-1}(A_1')\}\,P\{x_2^{-1}(A_2')\}\ldots$$

That is, the product measure is the product of the (single-variate) probability distributions P_{x_i} of the variates $x_1, x_2, \ldots$. It is readily seen that this product measure is equal to the induced measure $P\{T^{-1}(A')\}$ (i.e. the multivariate probability distribution) if and only if the variates $x_1, x_2, \ldots$ are statistically independent. In other words, the probability space $(\Omega', \mathbf{F}', \mu')$ induced by a sequence of *statistically independent* random variables is the *product space* of the component probability spaces induced by each one of the random variables separately. That is, as a measure in the product space

$$\Omega' = \mathbf{E}_1 \times \mathbf{E}_1 \times \ldots$$

the multivariate probability distribution $P_{x_1 x_2 \ldots}$ of statistically independent variates is the product of the single-variate probability distributions P_{x_i} on each line $\mathbf{E}_1$:

$$P_{x_1 x_2 \ldots} = P_{x_1} \times P_{x_2} \times \ldots$$

Let $g_n(\omega')$ be the *co-ordinate function* on the space Ω'; that is, if $\omega' = (\omega'_1, \omega'_2, \ldots)$, then $g_n(\omega') = \omega'_n$ for $n = 1, 2, \ldots$. Since

$$\omega' = T(\omega) = \{x_1(\omega), x_2(\omega), \ldots\}$$

we see that the variate $x_n(\omega)$ is given by

$$x_n(\omega) = g_n(\omega') = \omega'_n$$

Hence, the statistically independent variate x_n is the nth *co-ordinate* ω'_n of the point $\omega' = (\omega'_1, \omega'_2, \ldots)$ in the product space Ω'.

The use of product space makes it possible conveniently to represent a finite or indefinite number of statistically independent repetitions or observations of a statistical experiment. If the description of an experiment is given by a probability space $(\Omega, \mathbf{F}, P)$, then the description of a finite, or infinite, sequence of statistically independent repetitions is the finite, or infinite, product space

$$(\Omega \times \Omega \times \ldots, \mathbf{F} \times \mathbf{F} \times \ldots, P \times P \times \ldots)$$

Example 2.50 Suppose a biased coin, with probability p for head (1) and $q = 1-p$ for tail (0), is tossed twice. A single toss has probability space $(\Omega, \mathbf{F}, P)$ with points $\omega = 0$ and 1. The sets are $A = \{1, 0\}, \{1\}, \{0\}$, and the null set, with probabilities 1, p, q, and 0 respectively. (Note that the symbol $\{1, 0\}$ represents the set with elements 1 and 0, and does not represent an ordered pair or two-tuple, which would be denoted by $(1, 0)$.) Thus two statistically independent tosses would be described by the product space $(\Omega \times \Omega, \mathbf{F} \times \mathbf{F}, P \times P)$ with points ω' the ordered pairs $(1, 1)$, $(1, 0)$, $(0, 1)$, $(0, 0)$, where, for example, $(1, 0)$ would mean heads on the first toss and tails on the second toss. The sets of the product space are

$\{(1,1), (1,0), (0,1), (0,0)\}$
$\{(1,1), (1,0), (0,1)\}$, $\{(1,1), (0,1), (0,0)\}$,
$\{(1,1), (1,0), (0,0)\}$, $\{(1,0), (0,1), (0,0)\}$,
$\{(1,1), (1,0)\}$, $\{(1,1), (0,1)\}$, $\{(1,1), (0,0)\}$,
$\{(1,0), (0,1)\}$, $\{(1,0), (0,0)\}$, $\{(0,1), (0,0)\}$,
$\{(1,1)\}$, $\{(1,0)\}$, $\{(0,1)\}$, $\{(0,0)\}$, null set,

with probabilities $1, p^2+2pq, p^2+pq+q^2$,
$p^2+pq+q^2, 2pq+q^2, p^2+pq, p^2+pq$,
$p^2+q^2, 2pq, pq+q^2, pq+q^2, p^2, pq, pq, q^2, 0$,
respectively.

Basic to statistical theory are the relationships between integrals (or expectations) of independent random variables. Fubini's Theorem, which relates integrals on a product space and integrals on the component spaces, is the key theorem in this context. First, however, let us state some definitions. Let $(\Omega_1, \mathbf{F}_1, \mu_1)$ and $(\Omega_2, \mathbf{F}_2, \mu_2)$ be measure spaces, and let $(\Omega, \mathbf{F}, \mu) = (\Omega_1 \times \Omega_2, \mathbf{F}_1 \times \mathbf{F}_2, \mu_1 \times \mu_2)$ be their product space, where $\omega = (\omega_1, \omega_2)$. Let ω_1 be any point of Ω_1, and let $A = A_1 \times A_2$ be a subset of Ω. Define A_{ω_1} to be the set of points ω_2 for which $\omega = (\omega_1, \omega_2)$ is an element of A. Then A_{ω_1}, which is a subset of the component space Ω_2, is called the Ω_1-*section of* A. Now, let $f(\omega)$ be a function defined on A, and ω_1 be any point of Ω_1. Define the function f_{ω_1} on the Ω_1-section A_{ω_1} by

$$f_{\omega_1}(\omega_2) = f(\omega_1, \omega_2)$$

Then f_{ω_1} is called the Ω_1-*section of* f. In a similar manner, A_{ω_2}, the Ω_2-*section of* A, and f_{ω_2}, the Ω_2-*section of* f, are defined.

The integral

$$\int f(\omega)\,\mu(d\omega) \equiv \int f(\omega_1, \omega_2)\,\mu(d\omega_1\,d\omega_2) \equiv \int f(\omega_1, \omega_2)\,\mu_1 \times \mu_2(d\omega_1\,d\omega_2)$$

on the product space Ω is called the *double integral of* f. Denote the integral of the Ω_1-section of f by g; that is,

$$g(\omega_1) = \int f_{\omega_1}(\omega_2)\,\mu_2(d\omega_2) \tag{2.50}$$

and denote the integral of $g(\omega_1)$ by any of the equivalent expressions:

$$\begin{aligned}\int g(\omega_1)\,\mu_1(d\omega_1) &\equiv \iint f(\omega_1, \omega_2)\,\mu_2(d\omega_2)\,\mu_1(d\omega_1)\\ &\equiv \int \mu_1(d\omega_1) \int f(\omega_1, \omega_2)\,\mu_2(d\omega_2)\end{aligned}$$

which is called the *iterated integral of* f. Similarly, let h denote the integral

$$h(\omega_2) = \int f_{\omega_2}(\omega_1)\,\mu_1(d\omega_1) \tag{2.51}$$

so that the other *iterated integral of* f is denoted by any of the equivalent expressions:

$$\begin{aligned}\int h(\omega_2)\,\mu_2(d\omega_2) &\equiv \iint f(\omega_1, \omega_2)\,\mu_1(d\omega_1)\,\mu_2(d\omega_2)\\ &\equiv \int \mu_2(d\omega_2) \int f(\omega_1, \omega_2)\,\mu_1(d\omega_1)\end{aligned}$$

We then have

Theorem 2.50 (*Fubini's Theorem*) If f is an integrable function on $\Omega = \Omega_1 \times \Omega_2$, then every section of f is integrable, except for a set of points of measure zero. Moreover, both g and h as defined by equations (2.50) and (2.51) are integrable, and the double integral of f and the two iterated integrals are all equal; that is,

$$\int f(\omega_1, \omega_2)\,\mu_1 \times \mu_2(d\omega_1\, d\omega_2) = \int \mu_1(d\omega_1) \int f(\omega_1, \omega_2)\,\mu_2(d\omega_2)$$

$$= \int \mu_2\, d(\omega_2) \int f(\omega_1, \omega_2)\,\mu_1(d\omega_1)$$

Fubini's theorem can now be used to derive theorems for statistically independent variates $x_1(\omega)$ and $x_2(\omega)$ on a probability space $(\Omega, \mathbf{F}, P)$. Let

$$(\Omega', \mathbf{F}', P') = (\Omega \times \Omega, \mathbf{F} \times \mathbf{F}, P \times P)$$

be the product space. The points $\omega' = (\omega_1', \omega_2')$ of the product space Ω' have the variates x_1, x_2 as co-ordinates ω_1', ω_2'; i.e. $x_1(\omega) = \omega_1'$, $x_2(\omega) = \omega_2'$. Using Fubini's theorem, we have

$$\int x_1(\omega)\, x_2(\omega)\, P(d\omega) = \int \omega_1' \omega_2' P'(d\omega')$$

$$= \int \omega_1' \omega_2' P \times P(d\omega_1'\, d\omega_2') = \int P(d\omega_1') \int \omega_1' \omega_2' P(d\omega_2')$$

$$= \int \omega_1' P(d\omega_1') \int \omega_2' P(d\omega_2') = \int x_1(\omega) P(d\omega) \int x_2(\omega) P(d\omega)$$

so
$$E(x_1 x_2) = E(x_1) E(x_2)$$

where x_1 and x_2 are statistically independent. Also, since (Borel) functions of statistically independent random variables are statistically independent, we have, in particular,

$$E\{e^{il(x_1+x_2)}\} = E(e^{ilx_1})\, E(e^{ilx_2})$$

These theorems can be readily extended to an arbitrary number of statistically independent variates.

2.6 Conditional probability

The key concepts involved in the definition of conditional probability are those of measurable transformation and the Radon–Nikodym theorem. Let $\omega' = T(\omega)$ be a measurable transformation of the probability space $(\Omega, \mathbf{F}, P)$ into the probability space $(\Omega', \mathbf{F}', P')$ so that $P'(A') = P(A)$ where $A' = T(A)$. Let $y(\omega)$ be a variate integrable on Ω, and define $\lambda(A')$ to be

$$\lambda(A') = \int_A y(\omega) P(d\omega), \quad A' = T(A)$$

Now, if $P'(A') = 0$, then $P(A) = 0$ also, which, in view of the above equation, implies $\lambda(A') = 0$. Hence λ is absolutely continuous with respect to P'. Therefore, by the Radon–Nikodym theorem, there exists an integrable function $E'(y|\omega')$ of ω' defined uniquely almost everywhere on Ω' by

$$\lambda(A') = \int_{A'} E'(y|\omega')P'(d\omega')$$

Since $\omega' = T(\omega)$ is a measurable mapping of points ω into points ω', and $E'(y|\omega')$ is a measurable mapping of points ω' into points u on the real line, it follows that $E(y|\omega)$ defined by

$$E(y|\omega) = E'(y|\omega'), \quad \omega = T^{-1}(\omega')$$

is the equivalent measurable mapping of points ω directly into points u on the real line. Hence $E(y|\omega)$ is a measurable function, or variate, which satisfies

$$\lambda(A') = \int_{A'} E'(y|\omega')P'(d\omega') = \int_{A} E(y|\omega)P(d\omega)$$

Combining the two expressions for $\lambda(A')$, we have

$$\int_{A} y(\omega)P(d\omega) = \int_{A} E(y|\omega)P(d\omega)$$

for every measurable set A of $\mathbf{F}$. The variate $E(y|\omega) = E'(y|\omega')$ is called the *conditional expectation of y given $\omega' = T(\omega)$*.

Let B be a set belonging to $\mathbf{F}$, and let y be the variate

$$\begin{aligned} y(\omega) &= 1, \quad \omega \in B \\ &= 0, \quad \omega \in \Omega - B \text{ (i.e. complement of } B) \end{aligned}$$

The variate y, so defined, is called the *characteristic function of the set B*. Then the conditional expectation of y, given $\omega' = T(\omega)$, for the characteristic function $y(\omega)$ is defined to be the *conditional probability of B given $\omega' = T(\omega)$*. The conditional probability is denoted by $P'(B|\omega') = P(B|\omega)$. Because

$$\int_{A} y(\omega)P(d\omega) = P(A \cap B)$$

the conditional probability satisfies the relation

$$P(A \cap B) = \int_{A} P(B|\omega)P(d\omega)$$

for every measurable set A of $\mathbf{F}$.

Example 2.60 In a probability space $(\Omega, \mathbf{F}, P)$, consider a set A_1 with non-zero probability, and its complement A_2. The random events A_1 and A_2 are then *mutually exclusive* ($A_1 \cap A_2$ = null set) and *exhaustive* ($A_1 \cup A_2 = \Omega$). Let T be the measurable transformation of $(\Omega, \mathbf{F}, P)$ into a probability space $(\Omega', \mathbf{F}', P')$, where Ω' consists of exactly two points a_1 and a_2, and where

$$T(\omega) = a_1 \quad \text{if} \quad \omega \in A_1$$
$$T(\omega) = a_2 \quad \text{if} \quad \omega \in A_2$$

The class $\mathbf{F}'$ of random events consists of the sets

$$\{a_1\},\ \{a_2\},\ \Omega' = \{a_1, a_2\}$$

and the null set, with probabilities $P(A_1)$, $P(A_2)$, 1, and 0, respectively. Let B be any measurable set (random event) of the original space $(\Omega, \mathbf{F}, P)$. Then

$$P(A_1 \cap B) = \int_{A_1} P(B|\omega)\, P(d\omega) = P(B|A_1)\, P(A_1)$$

so
$$P(B|A_1) = \frac{P(A_1 \cap B)}{P(A_1)}$$

which is the usual *elementary definition of conditional probability.*

Example 2.61 Continuing Example 2.60, suppose that B_1, B_2, B_3 are mutually exclusive ($B_i \cap B_j =$ null set, $i \neq j$) and exhaustive ($B_1 \cup B_2 \cup B_3 = \Omega$) random events of $\mathbf{F}$. Define the variate $y(\omega)$ to be

$$y(\omega) = b_1 \quad \text{if} \quad \omega \in B_1$$
$$y(\omega) = b_2 \quad \text{if} \quad \omega \in B_2$$
$$y(\omega) = b_3 \quad \text{if} \quad \omega \in B_3$$

Then, since $a_1 = T(A_1)$, we have

$$\begin{aligned} \lambda(a_1) &= \int_{A_1} y(\omega)\, P(d\omega) \\ &= b_1 P(B_1 \cap A_1) + b_2 P(B_2 \cap A_1) + b_3 P(B_3 \cap A_1) \end{aligned}$$

But, we also have

$$\lambda(a_1) = \int_{A_1} E(y|\omega)\, P(d\omega) = E(y|A_1)\, P(A_1)$$

Combining the two expressions for $\lambda(a_1)$, we have

$$P(A_1) E(y|A_1) = \sum_{i=1}^{3} b_i P(B_i \cap A_1)$$

which is
$$E(y|A_1) = \sum_{i=1}^{3} b_i P(B_i|A_1)$$

an elementary definition of conditional expectation.

Example 2.62 Consider a probability space $(\Omega, \mathbf{F}, P)$ which can be partitioned into random events $A_1, A_2, \ldots, A_n$, each with a positive probability, and which are mutually exclusive ($A_i \cap A_j =$ null set, $i \neq j$) and exhaustive $\left(\bigcup_{i=1}^{n} A_i = \Omega\right)$. Let $T(\omega)$ be a measurable transformation of $(\Omega, \mathbf{F}, P)$ into $(\Omega', \mathbf{F}', P')$ where the space Ω' consists of exactly n points $a_1, a_2, \ldots, a_n$, and where $T(\omega) = a_i$ if $\omega \in A_i$. Let B be any random event of $\mathbf{F}$, and let A be the random event made up of any collection of the A_i's, say

$$A = A_{i_1} \cup A_{i_2} \cup \ldots \cup A_{i_k}$$

Then
$$P(A \cap B) = \int_A P(B \mid \omega) P(d\omega)$$

which is

$$P(A \cap B) = P(B \mid A_{i_1}) P(A_{i_1}) + \ldots + P(B \mid A_{i_k}) P(A_{i_k})$$

the so-called *theorem of total probability.*

CHAPTER III

THE GENERAL LIMIT THEOREM

3.1 Characteristic functions

Definition 3.10 The *characteristic function* $f(t)$ of the variate (i.e. random variable) $x(\omega)$ on the probability space $(\Omega, \mathbf{F}, P)$ is defined to be

$$f(t) = E(e^{itx}) = \int_{\Omega} e^{itx(\omega)} P(d\omega) \tag{3.10}$$

If $F(u)$ is the distribution function of $x(\omega)$, then the characteristic function may be written

$$f(t) = \int_{-\infty}^{\infty} e^{itu} dF(u) \tag{3.11}$$

In this chapter the capital letters F, G, H will denote distribution functions, and the lower-case letters f, g, h the corresponding characteristic functions. Throughout this chapter we assume that all variates are real. As usual, x, y, z, w will denote variates, and t, u, v real variables. Since for real t, $|e^{itu}| = 1$, we have

$$|f(t)| \leqslant \int_{-\infty}^{\infty} |e^{itu}| dF(u) = 1$$

so that the characteristic function is finite for all real t. Moreover, a characteristic function is continuous (in fact, uniformly continuous) for real t.

Let x_1 and x_2 be variates with distribution functions F_1 and F_2, and characteristic functions f_1 and f_2, respectively. If $x_2 = ax_1 + b$ where a and b are real constants, we have

$$f_2(t) = E(e^{itx_2}) = E(e^{it(ax_1+b)}) = e^{itb} E(e^{itax_1})$$

which is

$$f_2(t) = e^{itb} f_1(at) \tag{3.12}$$

Provided $a > 0$, we also have

$$F_2(u) = P\{ax_1 + b \leqslant u\} = P\left\{x_1 \leqslant \frac{u-b}{a}\right\} = F_1\left(\frac{u-b}{a}\right) \tag{3.13}$$

Suppose that x_1 and x_2 are statistically independent variates. Then we have seen in Section 2.5 that

$$E(e^{it(x_1+x_2)}) = E(e^{itx_1}) E(e^{itx_2}) = f_1(t) f_2(t)$$

which says that the characteristic function of the sum, x_1+x_2, of two statistically independent variates is the product $f_1(t)f_2(t)$ of the characteristic functions of the summands, x_1 and x_2. More generally the characteristic function of the sum, $x_1+x_2+\ldots+x_n$ of n statistically independent variates is equal to the product of the characteristic functions of the summands, i.e.

$$E(\mathrm{e}^{it(x_1+x_2\ldots+x_n)}) = f_1(t)f_2(t)\ldots f_n(t) \tag{3.14}$$

Given the characteristic function $f(t)$ of the variate x, we may recover the distribution function $F(u)$ by the inversion formula

$$F(u_2)-F(u_1) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{\mathrm{e}^{-itu_1}-\mathrm{e}^{-itu_2}}{it}f(t)\,dt \tag{3.15}$$

where u_1 and u_2 are continuity points of the function $F(u)$. If the variate x has a density function $F'(u)$, then the inversion formula takes the form

$$F'(u) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{e}^{-itu}f(t)\,dt$$

The inversion formula establishes a one-to-one correspondence between distribution functions $F(u)$ and characteristic functions $f(t)$. That is, a distribution function is uniquely determined by its characteristic function, and conversely. Consequently we may use $F(u)$ or $f(t)$ interchangeably to characterize a variate x.

Example 3.10 As we have seen in Example 2.40, the family of degenerate distribution functions are those with one and only one point of increase, a. The corresponding family of degenerate characteristic functions are those of the form

$$f(t) = \mathrm{e}^{iat} = \cos at + i\sin at$$

so that their moduli are equal to 1. In particular, $f(t) = 1$ is the characteristic function of a random variable degenerate at 0.

Example 3.11 Let x be a normal variate with mean a and variance σ^2 (Example 2.42). The characteristic function is

$$f(t) = \int_{-\infty}^{\infty}\mathrm{e}^{itu}\frac{1}{\sigma\sqrt{(2\pi)}}\exp\left\{-\frac{(u-a)^2}{2\sigma^2}\right\}du = \exp\{ita-\tfrac{1}{2}t^2\sigma^2\} \tag{3.16}$$

Let the statistically independent variates x_1 and x_2 be each normally distributed with means a_1 and a_2, and variances σ_1^2 and σ_2^2, respectively. The characteristic function of their sum, x_1+x_2, is

$$E[\exp\{it(x_1+x_2)\}] = f_1(t)f_2(t) = \exp\{it(a_1+a_2)-\tfrac{1}{2}(\sigma_1^2+\sigma_2^2)t^2\}$$

which is the characteristic function of a normal variate with mean a_1+a_2 and variance $\sigma_1^2+\sigma_2^2$. Hence if two statistically independent variates are normally distributed their sum is also normally distributed.

Example 3.12 Suppose the variate x has the Poisson distribution with mean a and variance a (Example 2.34). The characteristic function of x is then

$$f(t) = \sum_{n=0}^{\infty} e^{itn} \frac{a^n e^{-a}}{n!} = e^{-a} \sum_{n=0}^{\infty} \frac{(a e^{it})^n}{n!} = e^{a(e^{it}-1)}$$

Let x_1 and x_2 be statistically independent Poisson variates with means a_1 and a_2 respectively. Then the characteristic function of their sum is

$$E[\exp\{it(x_1+x_2)\}] = f_1(t)f_2(t) = \exp[\{a_1+a_2\}\{\exp(it)-1\}]$$

so their sum x_1+x_2 is also a Poisson variate with mean a_1+a_2. A more general form of the Poisson variate may be obtained by the transformation $y = b+cx$. Then we have

$$P\{y = b+cn\} = \frac{a^n e^{-a}}{n!}, \qquad a > 0, n = 0, 1, 2, \ldots$$

with characteristic function, by equation (3.12),

$$f_y(t) = e^{itb} e^{a(e^{ict}-1)} \tag{3.17}$$

3.2 Convergence properties

Let $F_1(u), F_2(u), \ldots$ be a sequence of one-dimensional distribution functions. Then we say that the $F_n(u)$ converge weakly to a distribution function $F(u)$ as $n \to \infty$ when and only when

$$\lim_{n\to\infty} F_n(u) = F(u)$$

at every point of continuity of $F(u)$. Weak convergence of F_n to F is denoted by $F_n \Rightarrow F$. Paul Lévy has given the following definition of distance $d(F, G)$ between two distribution functions, F and G. The graphs of F and G are drawn with vertical line segments at points of discontinuity. The distance, or metric, $d(F, G)$, is defined as the maximum distance between points on the two graphs along lines of slope -1.

With $d(F, G)$ as the metric, the metric space of one-dimensional distribution functions can be shown to be complete. Weak convergence of distribution functions then becomes convergence in this metric space; i.e. for the weak convergence $F_n \Rightarrow F$, it is necessary and sufficient that the numbers $d(F_n, F) \to 0$.

Under the one-to-one correspondence between distribution functions and characteristic functions, the totality of characteristic functions becomes a metric space if we define the distance function

$\delta(f, g)$ to be $\delta(f, g) = d(F, G)$ where f, g are the characteristic functions corresponding to the distribution functions F, G. The convergence $f_n \Rightarrow f$, defined by $\delta(f_n, f) \to 0$, is then equivalent to $F_n \Rightarrow F$ for the corresponding distribution functions.

The following theorem, and its converse, relate the convergence $f_n \Rightarrow f$ to the properties of the functions f_n and f.

Theorem 3.20 If $f_n(t)$ and $f(t)$ are characteristic functions of the distribution functions $F_n(u)$ and $F(u)$, and if $F_n(u) \Rightarrow F(u)$, then $f_n(t) \to f(t)$ as $n \to \infty$ uniformly in every bounded interval $|t| < T$.

Theorem 3.21 *The Lévy Continuity Theorem.* Conversely, if $f_n(t)$ is the characteristic function of the distribution function $F_n(u)$ and $f_n(t)$ converges as $n \to \infty$ for all t to a continuous function $f(t)$, then $F_n(u) \Rightarrow F(u)$, where $F(u)$ is a distribution function with the characteristic function $f(t)$.

These theorems show that the convergence $f_n(t) \Rightarrow f(t)$ for characteristic functions is equivalent to each of the following:

(1) uniform convergence of $f_n(t)$ on every finite interval;
(2) convergence at every point t without any requirement of uniformity. (Note: We recall that characteristic functions, unlike distribution functions, must be continuous.)

3.3. Infinitely divisible variates

Definition 3.30 A variate x, its distribution function $F(u)$, and characteristic function $f(t)$, are called *infinitely divisible* if for every positive integer n, the variate x can be represented as the sum

$$x = x_{n1} + x_{n2} + \dots + x_{nn}$$

where $x_{n1}, x_{n2}, \dots, x_{nn}$ are n statistically independent and identically distributed variates.

Denote the common distribution function of each x_{nj} $(j = 1, 2, \dots, n)$ by $F_{n\cdot}(u)$ and the common characteristic function by $f_{n\cdot}(t)$. Then, clearly, a variate x is infinitely divisible if and only if its characteristic function $f(t)$ may be written for every n as

$$f(t) = [f_{n\cdot}(t)]^n \tag{3.30}$$

where $f_{n\cdot}(t)$ is a characteristic function. By requiring $f_{n\cdot}(0) = 1$ and $f_{n\cdot}(t)$ to be continuous, the equation

$$f_{n\cdot}(t) = [f(t)]^{1/n}$$

makes it possible to determine uniquely $f_{n\cdot}(t)$ from $f(t)$.

Example 3.30 The degenerate distribution is infinitely divisible, since, as we have seen in Example 3.10, $f(t) = e^{ita}$ and

$$f_{n\cdot}(t) = e^{ita/n}$$

is a characteristic function of a variate degenerate at a/n, for every $n > 0$.

Example 3.31 Let x be a normal variate with mean a and variance σ^2, and with characteristic function $f(t)$ given by equation (3.16) (Example 3.11). Since, for every $n > 0$,

$$f_{n\cdot}(t) = [f(t)]^{1/n} = \exp\left(i\frac{a}{n}t - \frac{\sigma^2}{2n}t^2\right)$$

is the characteristic function of a normal distribution, it follows that the normal variate x is infinitely divisible.

Example 3.32 Let x be a Poisson variate with possible values $b+ck$ ($k = 0, 1, 2, \ldots$), with characteristic function given by equation (3.17) (Example 3.12). Since for every $n > 0$,

$$f_{n\cdot}(t) = \exp\left(it\frac{b}{n}\right)\exp\left[\frac{a}{n}\{\exp(itc)-1\}\right]$$

is the characteristic function of a Poisson variate, it follows that the Poisson variate x is infinitely divisible.

Theorem 3.30 The distribution function of the sum of a finite number of independent infinitely divisible variates is infinitely divisible.

Proof. Let $f(t)$ and $g(t)$ be the characteristic functions of the infinitely divisible variates x and y, respectively. Hence, for every positive integer n, we have

$$f(t) = [f_{n\cdot}(t)]^n, \qquad g(t) = [g_{n\cdot}(t)]^n$$

where $f_{n\cdot}(t)$ and $g_{n\cdot}(t)$ are characteristic functions. The characteristic function $h(t)$ of the sum $x+y$ is then

$$h(t) = f(t)g(t) = [f_{n\cdot}(t)g_{n\cdot}(t)]^n$$

and hence is infinitely divisible. Thus the distribution function of the sum $x+y$ of infinitely divisible variates is also infinitely divisible, and the theorem follows by induction. Q.E.D.

Theorem 3.31 A distribution function $F(u)$ that is the limit (weak convergence) of a sequence $F^{(1)}(u), F^{(2)}(u), \ldots$ of infinitely divisible distributions is infinitely divisible.

Proof. Suppose that the sequence $f^{(1)}(t), f^{(2)}(t), \ldots$ of infinitely divisible characteristic functions converges to the characteristic function $f(t)$; i.e.

$$f^{(k)}(t) \Rightarrow f(t) \quad \text{as} \quad k \to \infty \tag{3.31}$$

Since the $f^{(k)}(t)$ are infinitely divisible, the function

$$f_{n\cdot}^{(k)}(t) = [f^{(k)}(t)]^{1/n} \tag{3.32}$$

is a characteristic function for every n and never vanishes for any t (see Exercise 3.9). From (3.31) and (3.32) it follows that

$$f_{n\cdot}^{(k)}(t) \Rightarrow [f(t)]^{1/n}$$

From the Lévy Continuity Theorem 3.21, the function $[f(t)]^{1/n}$, which we call $f_{n\cdot}(t)$, is a characteristic function. Hence, for every n, equation (3.30) is true, and so $f(t)$, and its distribution function $F(u)$, are infinitely divisible. Q.E.D.

Theorem 3.32 A characteristic function is infinitely divisible if and only if it is the product of a finite number of Poisson characteristic functions, or is the limit ($\Rightarrow$) of a sequence of characteristic functions of such products. A distribution function is infinitely divisible if and only if it is the distribution function of the sum of a finite number of statistically independent Poisson variates, or is the limit ($\Rightarrow$) of a sequence of distribution functions of such sums.

Proof. Suppose that $F(u)$ is the distribution function of a sum of statistically independent Poisson variates or that $F(u)$ is the limit ($\Rightarrow$) of a sequence of distribution functions of such sums. Because a Poisson variate is infinitely divisible (Example 3.32), application of Theorems 3.30 and 3.31 shows immediately that $F(u)$ is infinitely divisible.

Conversely, suppose that $F(u)$ is an infinitely divisible distribution function, and so its characteristic function $f(t)$ satisfies

$$f_{n\cdot}(t) = [f(t)]^{1/n} = \int_{-\infty}^{\infty} e^{itu}\, dF_{n\cdot}(u) \tag{3.33}$$

where $f_{n\cdot}(t)$ is a characteristic function, and $F_{n\cdot}(u)$ the corresponding distribution function. Since

$$n(\sqrt[n]{a}-1) = n(e^{(1/n)\log a}-1) = n\left(1+\frac{1}{n}\log a + o(1/n) - 1\right) \to \log a$$

it follows that

$$n\{f_{n\cdot}(t)-1\} \Rightarrow \log f(t) \quad \text{as} \quad n\to\infty \tag{3.34}$$

and so

$$\exp[n\{f_{n\cdot}(t)-1\}] \Rightarrow f(t) \quad \text{as} \quad n\to\infty \tag{3.35}$$

The integral (3.33) for $f_{n\cdot}(t)$ can be represented as the limit of Stieltjes sums (as $m\to\infty$):

$$\sum_{k=1}^{m} e^{itu_k}[F_{n\cdot}(u_k)-F_{n\cdot}(u_{k-1})] \Rightarrow f_{n\cdot}(t) \tag{3.36}$$

Put $a_k = n[F_{n\cdot}(u_k)-F_{n\cdot}(u_{k-1})]$, and substitute (3.36) into (3.35), which gives

$$\exp\left\{\sum_{k=1}^{m} a_k(e^{itu_k}-1)\right\} \Rightarrow f(t)$$

since $\Sigma a_k \to n$ as $m \to \infty$. Hence $f(t)$ is the limit ($\Rightarrow$) of a sequence of characteristic functions that are the products of a finite number of Poisson characteristic functions (Example 3.12). Q.E.D.

3.4 The canonical representation

André Kolmogorov (1932) established the following canonical representation for infinitely divisible distribution functions with finite variances. This section will be devoted to its proof.

Theorem 3.40 *The canonical representation.* The function $f(t)$ is the characteristic function of an infinitely divisible distribution function $F(u)$ with mean a and (finite) variance σ^2 if and only if

$$\log f(t) = iat + \int_{-\infty}^{\infty} (\mathrm{e}^{itu} - 1 - itu)\frac{1}{u^2}\, dK(u) \tag{3.40}$$

where $K(u)$ is a non-decreasing function with

$$K(\infty) - K(-\infty) = \sigma^2$$

The representation of $\log f(t)$ by formula (3.40), called Kolmogorov's formula, is unique.

Note. The integrand at $u = 0$ in formula (3.40) is defined by continuity; that is,

$$\left[(\mathrm{e}^{itu} - 1 - itu)\frac{1}{u^2}\right]_{u=0} = \frac{-t^2}{2} \tag{3.41}$$

Also with loss of generality we may and shall take $K(-\infty) = 0$.

Proof ("only if" statement). By hypothesis $F(u)$ is infinitely divisible, so its characteristic function can be represented by equation (3.30) for every $n > 0$. By equations (3.34) and (3.33), we have

$$n\,[f_{n\cdot}(t) - 1] = n \int_{-\infty}^{\infty} (\mathrm{e}^{itu} - 1)\, dF_{n\cdot}(u) \Rightarrow \log f(t) \tag{3.42}$$

where $f_{n\cdot}(t)$ and $F_{n\cdot}(u)$ are the common characteristic function and common distribution function of each individual summand x_{nj} in the sum $x_n = x_{n1} + x_{n2} + \ldots + x_{nn}$. Because the summands x_{nj} are statistically independent, the variance σ^2 of x_n is

$$\sigma^2 \equiv D^2(x_n) = D^2(x_{n1}) + \ldots + D^2(x_{nn}) = n \int_{-\infty}^{\infty} u^2\, dF_{n\cdot}(u)$$

With this motivation, we define the "partial" variance $K_n(u)$ by

$$K_n(u) = n \int_{-\infty}^{u} v^2\, dF_{n\cdot}(v) \tag{3.43}$$

where clearly $K_n(-\infty) = 0$ and $K_n(\infty) = \sigma^2$. Equivalently, equation (3.43) may be written

$$\frac{1}{u^2} dK_n(u) = n dF_{n\cdot}(u) \tag{3.44}$$

so expression (3.42) becomes

$$\int_{-\infty}^{\infty} (e^{itu} - 1) \frac{1}{u^2} dK_n(u) \Rightarrow \log f(t) \tag{3.45}$$

It can be shown by analytical considerations that we can always pick a subsequence from the sequence $K_1(u), K_2(u), \ldots$ of partial variances such that $K_n(u) \Rightarrow K(u)$ where $K_n(u)$ is a member of the subsequence, and $K(u)$, like the $K_n(u)$, is also a function for which $K(-\infty) = 0$ and $K(\infty) = \sigma^2$. Put

$$a_n = \int_{-\infty}^{\infty} \frac{dK_n(u)}{u} = n \int_{-\infty}^{\infty} u dF_{n\cdot}(u)$$

Then by equation (3.44) we have

$$ita_n - \int_{-\infty}^{\infty} itu \frac{1}{u^2} dK_n(u) = 0$$

which we add to the left-hand side of expression (3.45) and obtain

$$ia_n t + \int_{-\infty}^{\infty} (e^{itu} - 1 - itu) \frac{1}{u^2} dK_n(u) \Rightarrow \log f(t)$$

The integral in this expression converges to

$$\int_{-\infty}^{\infty} [e^{itu} - 1 - itu] \frac{1}{u^2} dK(u)$$

and since the whole left-hand side of this expression converges to $\log f(t)$, we must conclude that a_n converges to some number a. Thus we have obtained Kolmogorov's formula (3.40). Q.E.D. (1).

Proof ("if" statement). By hypothesis Kolmogorov's formula (3.40) holds. The integral in this formula is the limit of the Stieltjes sums (where all u_k are chosen different from zero)

$$\sum_{k=1}^{m} (e^{itu_k} - 1 - itu_k) \frac{1}{u_k^2} [K(u_k) - K(u_{k-1})]$$

Since each term is the logarithm of the characteristic function of a Poisson distribution, it follows from Theorem 3.32 that $f(t)$ is infinitely divisible. Q.E.D. (2).

Example 3.40 Let x be a normal variate with mean a and variance σ^2. Then in Kolmogorov's formula (3.40), it is seen that $a = a$, and $K(u) = 0$ for $u < 0$, $K(u) = \sigma^2$ for $u \geqslant 0$, i.e. the partial variance $K(u)$ is a step-function with a single step equal to σ^2 at $u = 0$.

Example 3.41 Let x be a Poisson variate with mean a and variance a. Then in Kolmogorov's formula (3.40) it is seen that $a = a$, $K(u) = 0$ for $u < 1$, $K(u) = a$ for $u \geqslant 1$; i.e. the partial variance $K(u)$ is a step-function with a single step equal to a at $u = 1$.

3.5 Vanishingly small variates

The General Limit Theorem, due to G. M. Bawly (1936) and A. Ja. Khintchine (1937), concerns sums of an increasing number of independent variates. Let $x_1, x_2, \ldots, x_n, \ldots$ be a sequence of variates, each of which is the sum of k_n mutually independent variates $x_{n1}, x_{n2}, \ldots, x_{nk_n}$ where $k_n \to \infty$ as $n \to \infty$; i.e.

$$x_n = x_{n1} + x_{n2} + \ldots + x_{nk_n} \quad (n = 1, 2, 3, \ldots)$$

For the General Limit Theorem it is required that the role of each summand $x_{nj}\,(j = 1, 2, \ldots, k_n)$ must become more and more negligible as $n \to \infty$. In this way no single summand x_{nj} can have a predominant effect on the total x_n. Hence the total sum x_n is not dominated by any one, or a few, of the summands x_{nj}, but instead represents the combined effect of many small influences $x_{n1}, x_{n2}, \ldots, x_{nk_n}$, each and every one of which becomes negligible (or vanishingly small) as $n \to \infty$.

In this section, we wish to define "vanishingly small variates", and prove three important lemmas about them.

Definition 3.50 The variates x'_{nj} $(j = 1, 2, \ldots, k_n)$, with distribution functions $F'_{nj}(u)$, and characteristic functions $f'_{nj}(t)$, are called vanishingly small (or infinitesimal) if

$$\max_{1 \leqslant j \leqslant k_n} P\{|x'_{nj}| \geqslant \epsilon\} \to 0$$

as $n \to \infty$ for arbitrary $\epsilon > 0$.

In other words, the variates x'_{nj} are vanishingly small if and only if x'_{nj} converge (in probability) to zero uniformly in j $(j = 1, 2, \ldots, k_n)$ as $n \to \infty$. In this section, and the next, primes will indicate vanishingly small variates, their distribution functions, and their characteristic functions.

As we have seen in Example 3.10, the characteristic function of a variate degenerate at zero is equal to 1. Hence, if we define α_{nj} to be

$$\alpha_{nj} = f'_{nj}(t) - 1$$

we see that α_{nj} represents the discrepancy between the characteristic function of the vanishingly small variate x'_{nk} and the characteristic function of a variate degenerate at zero. We shall assume throughout that all vanishingly small variates x'_{nj} have zero mean, i.e. $E(x'_{nj}) = 0$, and finite variance, i.e. $D^2(x'_{nj}) < \infty$.

Lemma 3.50:

$$\max_{j=1,2,\ldots,k_n} |\alpha_{nj}| \Rightarrow 0 \quad \text{as} \quad n \to \infty$$

Proof. By hypothesis, x'_{nj} are vanishingly small. Consider, for $\epsilon > 0$

$$\max_j |\alpha_{nj}| = \max_j |f'_{nj}(t) - 1| = \max_j \left| \int_{-\infty}^{\infty} (e^{itu} - 1)\, dF'_{nj}(u) \right|$$

$$\leqslant \max_j \int_{|u|<\epsilon} |e^{itu} - 1|\, dF'_{nj}(u) + 2 \max_j \int_{|u| \geqslant \epsilon} dF'_{nj}(u)$$

Since $|e^{iv} - 1| \leqslant |v|$ for real v, we have

$$\max_j |\alpha_{nj}| \leqslant \epsilon |t| + 2 \max_j P\{|x'_{nj}| \geqslant \epsilon\}$$

Hence, because the x'_{nj} are vanishing small, the right-hand side of this inequality can be made arbitrarily small for large enough n. Q.E.D.

Lemma 3.51:

$$|\log f'_{nj}(t) - \alpha_{nj}| \leqslant |\alpha_{nj}|^2$$

Proof. By Lemma 3.50, it follows that for $n > N$, where N is sufficiently large

$$\max_j |\alpha_{nj}| < 1/2$$

on an arbitrary but fixed interval of t. Since, for real v,

$$\log(1+v) - v = -\frac{v^2}{2} + \frac{v^3}{3} - \frac{v^4}{4} + \ldots$$

we have

$$|\log f'_{nj}(t) - \alpha_{nj}| = |\log(1+\alpha_{nj}) - \alpha_{nj}| \leqslant \sum_{s=2}^{\infty} \frac{1}{s} |\alpha_{nj}|^s$$

$$\leqslant \frac{1}{2} \sum_{s=2}^{\infty} |\alpha_{nj}|^s = \frac{1}{2} \frac{|\alpha_{nj}|^2}{1 - |\alpha_{nj}|} < |\alpha_{nj}|^2 \qquad \text{Q.E.D.}$$

Lemma 3.52:

$$|\alpha_{nj}| \leqslant \frac{t^2}{2} D^2(x'_{nj})$$

Proof. Since

$$E(x'_{nj}) = \int_{-\infty}^{\infty} u\, dF'_{nj}(u) = 0$$

α_{nj} may be written

$$\alpha_{nj} = \int_{-\infty}^{\infty} (e^{itu} - 1)\, dF'_{nj}(u) = \int_{-\infty}^{\infty} (e^{itu} - 1 - itu)\, dF'_{nj}(u)$$

For every real v, the following inequality holds:

$$|e^{itv}-1-itv| \leqslant \frac{t^2 v^2}{2}$$

Consequently,

$$|\alpha_{nj}| \leqslant \frac{t^2}{2}\int_{-\infty}^{\infty} u^2\, dF'_{nj}(u) = \frac{t^2}{2} D^2(x'_{nj}) \qquad \text{Q.E.D.}$$

3.6 The General Limit Theorem

This section is devoted to the proof of

Theorem 3.60 *The Khintchine–Bawly General Limit Theorem* (for sums with bounded variances):

(I) Given a sequence $A_1, A_2, \ldots$ of constants, let $F_n(u)$ be the distribution function of the sum

$$x_n = x_{n1} + x_{n2} + \ldots + x_{nk_n} - A_n \tag{3.60}$$

where the summands x_{nj} ($j = 1, 2, \ldots, k_n$) are independent variates subject to the conditions:

(1) The deviates $x'_{nj} = x_{nj} - E(x_{nj})$, $j = 1, 2, \ldots, k_n$, are vanishingly small as $n \to \infty$.
(2) The variates x_{nj} have finite variances $D^2(x_{nj})$.
(3) The variance of the sum is bounded by C for all n; i.e.

$$D^2(x_n) = D^2(x_{n1}) + \ldots + D^2(x_{nk_n}) < C \tag{3.61}$$

where C is a constant independent of n.

(II) Given a sequence $A_1, A_2, \ldots$ of constants, let $G_n(u)$ be the distribution function corresponding to the characteristic function $e^{\psi_n(t)}$ whose logarithm is

$$\psi_n(t) = -iA_n t + \sum_{j=1}^{k_n}\left\{itE(x_{nj}) + \int_{-\infty}^{\infty}(e^{itu}-1)\,dF_{nj}(u+E(x_{nj}))\right\} \tag{3.62}$$

where $F_{nj}(u)$ is the distribution function of x_{nj}. Then, the following statements are true:

(A) $G_n(u)$ is an infinitely divisible distribution function.

(B) For a given sequence of constants $A_1, A_2, \ldots$ the sequence $F_1(u), F_2(u)$ of distribution functions (which involve $A_1, A_2, \ldots$) converges weakly to a limit distribution function $F(u)$ if and only if the sequence of $G_1(u), G_2(u), \ldots$ of distribution functions (which involve the *same* $A_1, A_2, \ldots$) converges weakly to a limit distribution function $G(u)$. Moreover, the two limit distribution functions coincide, i.e. $F(u) = G(u)$.

(B′) (Concise statement of B.) For a given sequence $A_1, A_2, \ldots$, we have $F_n \Rightarrow F$ if and only if $G_n \Rightarrow F$.

(C) For a given sequence $A_1, A_2, \ldots$, we have $f_n(t) \Rightarrow f(t)$ if and only if $g_n(t) = e^{\psi_n(t)} \Rightarrow f(t)$.

(D) For a given sequence $A_1, A_2, \ldots$, we have $f_n(t) - e^{\psi_n(t)} \Rightarrow 0$.

(E) For a given sequence $A_1, A_2, \ldots$, we have $\log f_n(t) - \psi_n(t) \Rightarrow 0$.

(F) The limit distribution function $F(u) = G(u)$ is infinitely divisible.

Proof of Statement A. Consider the partial variance defined by

$$K_n(u) \equiv \sum_{j=1}^{k_n} \int_{-\infty}^{u} v^2 \, dF'_{nj}(v) \tag{3.63}$$

where $F'_{nj}(v) = F_{nj}\{v + E(x_{nj})\}$ is the distribution function of the deviate $x'_{nj} = x_{nj} - E(x_{nj})$. The partial variance $K_n(u)$ is non-decreasing, $K_n(-\infty) = 0$, and

$$K_n(\infty) = \sum_{j=1}^{k_n} \int_{-\infty}^{\infty} v^2 \, dF_{nj}\{v + E(x_{nj})\} = \sum_{j=1}^{k_n} D^2(x_{nj}) = D^2(x_n)$$

We readily see that $\psi_n(t)$ may be rewritten as

$$\psi_n(t) = it\left[\sum_{j=1}^{k_n} E(x_{nj}) - A_n\right] + \int_{-\infty}^{\infty} (e^{itu} - 1 - itu)\frac{1}{u^2}\, dK_n(u)$$

which is the Canonical Representation (3.40) of the logarithm of an infinitely divisible characteristic function. Q.E.D. (A).

Proof of Statements B, B′, C, D, E. Since it may be readily verified that these statements are equivalent, we shall prove only Statement E. Consider the sum (3.60) which may be written

$$x_n = [x_{n1} - E(x_{n1})] + \ldots + [x_{nk_n} - E(x_{nk_n})] + \sum_{j=1}^{k_n} E(x_{nj}) - A_n$$

Because the summands are independent, the characteristic function of x_n is

$$f_n(t) = f'_{n1}(t) f'_{n2}(t) \ldots f'_{nk_n}(t) \exp\left[it\left\{\sum_j E(x_{nj}) - A_n\right\}\right]$$

where sums on j are from 1 to k_n. By taking logarithms we have

$$\log f_n(t) = \log f'_{n1}(t) + \log f'_{n2}(t) + \ldots$$
$$+ \log f'_{nk_n}(t) + it\sum_j E(x_{nj}) - itA_n$$

which is $$\log f_n(t) + itA_n - \sum_j itE(x_{nj}) = \sum_j \log f'_{nj}(t)$$

We recall that $\alpha_{nj} = f'_{nj}(t) - 1$. Hence we have

$$\log f_n(t) - \left\{-itA_n + \sum_j itE(x_{nj}) + \sum_j \alpha_{nj}\right\} = \sum_j \log f'_{nj}(t) - \sum_j \alpha_{nj}$$

It immediately follows from the definition of α_{nj} that ψ_n given by equation (3.62) is equal to

$$\psi_n(t) = -itA_n + \sum_j itE(x_{nj}) + \sum_j \alpha_{nj}$$

so

$$\log f_n(t) - \psi_n(t) = \sum_j \{\log f'_{nj}(t) - \alpha_{nj}\}$$

Hence, we have

$$|\log f_n(t) - \psi_n(t)| \leqslant \sum_j |\log f'_{nj}(t) - \alpha_{nj}|$$

Applying Lemma 3.51 to the right-hand side of this expression, we have

$$|\log f_n(t) - \psi_n(t)| \leqslant \sum_j |\alpha_{nj}|^2$$

Since $|\alpha_{nj}| \leqslant \max_j |\alpha_{nj}|$, it follows that

$$|\log f_n(t) - \psi_n(t)| \leqslant \max_j |\alpha_{nj}| \sum_j |\alpha_{nj}|$$

Applying Lemma 3.52, we have

$$|\log f_n(t) - \psi_n(t)| \leqslant \max_j |\alpha_{nj}| \frac{t^2}{2} \sum_j D^2(x'_{nj})$$

Since $D^2(x'_{nj}) = D^2(x_{nj})$ and since by equation (3.61)

$$\sum_j D^2(x_{nj}) < C$$

we have

$$|\log f_n(t) - \psi_n(t)| \leqslant \frac{t^2 C}{2} \max_j |\alpha_{nj}|$$

Therefore, by Lemma 3.50, we have $|\log f_n(t) - \psi_n(t)| \Rightarrow 0$, so $\log f_n(t) - \psi_n(t) \Rightarrow 0$. Q.E.D. (B, B′, C, D, E).

Proof of Statement F. Statement F follows immediately from Theorem 3.31. Q.E.D. (F).

From the general limit theorem, conditions for the existence of limit distribution functions $F(u)$ can be established. These conditions are established in

Theorem 3.61 (Gnedenko, 1939): Choose the constants A_n by the formula

$$A_n = \sum_{j=1}^{k_n} E(x_{nj}) - a + o(1) \tag{3.64}$$

where a is any constant. Then the distribution functions $F_n(u)$ of the sums $x_n = x_{n1} + \ldots + x_{nk_n} - A_n$ of independent variates x_{nj} with vanishingly small deviates $x'_{nj} = x_{nj} - E(x_{nj})$ $(j = 1, 2, \ldots, k_n)$ converge weakly to a limit distribution $F(u)$, and the variances $D^2(x_n)$ of these sums converge to the variance σ^2 of the limit distribution $F(u)$, if and only if there exists a non-decreasing function $K(u)$ (with $K(-\infty)=0$ and $K(\infty)=\sigma^2$) such that $K_n(u) \Rightarrow K(u)$ as $n \to \infty$, where $K_n(u)$ is defined by equation (3.63). The limit distribution $F(u)$ has the characteristic function $f(t)$ where $\log f(t)$ is given by Kolmogorov's formula (3.40) with the constant a and partial variance, $K(u)$.

The proof follows readily from Theorem 3.60 and a convergence theorem for infinitely divisible distributions. See Gnedenko and Kolmogorov, Chapter 3, § 19 and Chapter 4, § 21.

3.7 The Central Limit Theorem

The classical limit theorems of statistics and theoretical physics are special cases of the general limit theorem. From Example 3.40 and Theorem 3.61, we obtain

Theorem 3.70 *The Central Limit Theorem.* Choose the constants A_n by formula (3.64), where a is any constant. Then the distribution functions $F_n(u)$ of the sums $x_n = x_{n1} + \ldots + x_{nk_n} - A_n$ of independent variates with vanishingly small deviates converge to the normal distribution (with mean a and variance σ^2) and the variances $D^2(x_n) \to \sigma^2$ as $n \to \infty$ if and only if

$$K_n(u) = \sum_{j=1}^{k_n} \int_{-\infty}^{u} v^2 \, dF'_{nj}(v) \to \begin{cases} 0 & u < 0 \\ \sigma^2 & u > 0 \end{cases}, \quad K_n(\infty) \to \sigma^2 \qquad (3.71)$$

In the special case when $E(x_{nj}) = 0$ and $D^2(x_n) = 1$ for all n, j, the condition (3.71) is seen to be equivalent to the condition

$$K_n(\infty) - K_n(\epsilon) + K_n(-\epsilon) - K_n(-\infty) = \sum_{j=1}^{k_n} \int_{|v| \geqslant \epsilon} v^2 \, dF_{nj}(v) \to 0$$

as $n \to \infty$ where $\epsilon > 0$. Hence we have

Theorem 3.71 *The Lindeberg–Feller Central Limit Theorem.* Let $y_1, y_2, \ldots$ be a sequence of independent variates with distribution functions $F_1, F_2, \ldots$. The distribution functions of the normalized sums

$$x_n = \frac{\sum_{j=1}^{n} [y_j - E(y_j)]}{D_n}, \quad \left(D_n^2 = \sum_{j=1}^{n} D^2(y_j) \right)$$

with vanishingly small summands converge to the standard normal distribution (i.e. a normal distribution with zero mean and unit variance) if and only if Lindeberg's condition

$$\frac{1}{D_n^2} \sum_{j=1}^{n} \int_{|v| \geqslant \epsilon D_n} v^2 \, dF_j\{v + E(y_j)\} \to 0$$

is satisfied for every $\epsilon > 0$.

The classical *Central Limit Theorem of Ljapounov* follows readily from the Lindeberg–Feller Central Limit Theorem 3.71.

3.8 The Poisson Limit Theorem

From Example 3.41 and Theorem 3.61 we obtain

Theorem 3.80 *The Poisson Limit Theorem* (Gnedenko, 1944). Choose the constants A_n by formula (3.64), where a is any constant. Then the distribution functions of the sums

$$x_n = x_{n1} + \ldots + x_{nk_n} - A_n$$

of independent variates with vanishingly small deviates converge weakly to the Poisson distribution (with mean a and variance a) and the variances $D^2(x_n) \to a$ as $n \to \infty$ if and only if $K_n(u) \to 0$ for $u < 1$, $K_n(u) \to a$ for $u > 1$, and $K_n(\infty) \to a$.

CHAPTER IV

HILBERT SPACE

4.1 Linear space

G. Udny Yule (1921, 1927) proposed methods of time-series analysis that were of essentially wider scope than classical methods. The work of André Kolmogorov (1933) and A. Ja. Khintchine (1933) on probability theory made possible the study of (finite or infinite) families of variates: the so-called *stochastic processes.* Khintchine (1934) defined and laid the groundwork of *stationary stochastic processes.* M. Fréchet (1937) observed that a set of variates with finite variances becomes a metric space by defining the distance between two variates, say x and y, as the standard deviation of the difference variate $x-y$. Herman Wold (1938) unified the fertile methods of Yule with the probabilistic work of Kolmogorov and Khintchine. Not least among Wold's results was his use of Fréchet's observation to develop the theory of (second-order) *stationary stochastic processes.* Wold's novel way of treating an infinite number of variates as elements in linear space led to the central result of the theory of stationary processes: the Wold decomposition. Kolmogorov (1939, 1941) generalized Wold's contributions and obtained further deep results. Norbert Wiener (1942), making use of some of his earlier important work (Wiener, 1930), independently arrived at Kolmogorov's results. Furthermore, Wiener developed the direct application of this field to communication and control, which already has had profound effects in diverse sciences.

The rest of this book will be devoted to the Wold–Kolmogorov approach. The reason for the way in which we develop our presentation, which is not the conventional way, is to make clear the relationship of the mathematics of stationary processes with that of quantum mechanics.

The postulational treatment of linear, inner product space is due to John von Neumann (1929, 1932, 1933). Let $\mathbf{V}$ denote an abstract space, with elements x, y, z, w, and let a, b, c be complex numbers. We assume throughout that addition, $x+y$, is defined for every pair of elements x, y in $\mathbf{V}$, and that multiplication, ax, by an arbitrary complex number a, is defined for every x in $\mathbf{V}$.

Definition 4.10 If x and y are two elements of a space $\mathbf{V}$, then $\mathbf{V}$ is called *linear* if $x+y$ and ax also belong to $\mathbf{V}$, and if

(1) $x+y = y+x$ (commutative law of addition),

(2) $x+(y+z) = (x+y)+z$ (associative law of addition),
(3) there exists a unique element 0 in $\mathbf{V}$ such that $x+0 = x$ (role of the null element 0),
(4) there corresponds to every element x a unique element $-x$ such that $x+(-x) = 0$ (role of the inverse),
(5) $a(bx) = (ab)x$ (associative law of multiplication),
(6) $1x = x$ (role of unity),
(7) $(a+b)x = ax+bx$ (distributive law),
(8) $a(x+y) = ax+ay$ (distributive law),

where x, y, z are elements of $\mathbf{V}$ and a, b are complex numbers.

Postulate P_1 The space $\mathbf{V}$ is linear.

4.2 Inner product

Definition 4.20 If a complex number $((x,y))$ exists for each pair of elements x, y of linear space $\mathbf{V}$ such that

(1) $((ax,y)) = a((x,y))$ where a is a complex number (associative law for the first factor),
(2) $((x+y,z)) = ((x,z))+((y,z))$ (distributive law for the first factor),
(3) $((x,y)) = \overline{((y,x))}$ where the bar indicates the complex conjugate (Hermitian symmetry),
(4) $((x,x)) \geqslant 0$, and $= 0$ if and only if $x = 0$ (definite form),

then $((x,y))$ is called the *inner product* of x and y.

Postulate P_2 An inner product $((x,y))$ is defined over the linear space $\mathbf{V}$.

Definition 4.21 The positive square root of $((x,x))$ is called the *length* of the element x and is denoted by $\|x\|$. The *distance* $d(x,y)$ between x and y is defined to be $\|y-x\|$.

Since it is easily verified that the distance function $d(x,y)$ satisfies the metric conditions, the space $\mathbf{V}$ is a metric space with the elements $x, y, \ldots$ as points. Consequently, all concepts defined for points in a metric space are applicable, in particular such concepts as convergence, separability, fundamental sequences, and completeness.

Postulate P_3 The linear space $\mathbf{V}$, with inner product $((x,y))$, is complete.

Definition 4.22 A space $\mathbf{V}$ satisfying Postulates P_1, P_2 and P_3 is called *von Neumann space* $\mathbf{V}$.

We assume throughout that $\mathbf{V}$ is a von Neumann space.

4.3 Orthogonal elements

To investigate the geometry of the von Neumann space **V**, the only trigonometry we shall use is the concept of the right angle, i.e. orthogonality.

Definition 4.30 Two elements x and y in **V** are said to be *orthogonal*, $x \perp y$, if $((x,y)) = 0$. A set **O** of elements in **V** is called *orthonormal* if each element x in **O** has length $\|x\| = 1$, and if each pair of distinct elements $x, y\,(x \neq y)$ in **O** is orthogonal $((x,y)) = 0$. If **O** is an orthonormal set in **V**, **O** is called *maximal* in **V** if it is not a proper subset of any other orthonormal set in **V**.

We see that for any orthonormal set **O** to be maximal in **V**, **V** cannot contain any element x, of length 1, that is orthogonal to every element in **O**. Hence **V** cannot contain any element y, of positive length $\|y\| > 0$, that is, orthogonal to every element in a maximal orthonormal set **O**. For if such a y did exist, then $y/\|y\|$ would be an element of length 1 orthogonal to every element in **O**, which would violate the maximality of **O**.

Let $\theta_1, \ldots, \theta_n$ be an orthonormal set in **V**, (i.e. $((\theta_i, \theta_j)) = 0$ if $i \neq j$, and $= 1$ if $i = j$). If $x = a_1\theta_1 + \ldots + a_n\theta_n$, then it follows that a_i is the Fourier coefficient $a_i = ((x, \theta_i))$. Moreover, if x is arbitrary, let $x' = a_1\theta_1 + \ldots + a_n\theta_n$, where $a_i = ((x, \theta_i))$. Then $x - x'$ is orthogonal to $\theta_1, \ldots, \theta_n$ because $((x - x', \theta_i)) = a_i - a_i = 0$.

Theorem 4.30 Let $\theta_1, \theta_2, \ldots, \theta_n$ be a finite orthonormal set in **V** and let x be any element in **V**. The distance $d = \left\| x - \sum_1^n a_i\theta_i \right\|$ is minimized when $a_i = ((x, \theta_i))$.
If $d_{\min}$ is the minimum value of d, then

$$0 \leqslant d_{\min}^2 = \|x\|^2 - \sum_{i=1}^n |((x, \theta_i))|^2$$

Proof.

$$d^2 = \left(\left(x - \sum_1^n a_i\theta_i,\ x - \sum_1^n a_i\theta_i\right)\right)$$

$$= ((x,x)) - \left(\left(x, \sum_1^n a_i\theta_i\right)\right) - \left(\left(\sum_1^n a_i\theta_i, x\right)\right) + \left(\left(\sum_1^n a_i\theta_i, \sum_1^n a_i\theta_i\right)\right)$$

$$= ((x,x)) - 2Re\left[\sum_1^n a_i\overline{((x, \theta_i))}\right] + \sum^n |a_i|^2$$

$$= \sum_1^n \{|a_i|^2 - 2Re\,[a_i\overline{((x, \theta_i))}] + |((x, \theta_i))|^2\} + \|x\|^2 - \sum_1^n |((x, \theta_i))|^2$$

$$= \sum_1^n |a_i - ((x, \theta_i))|^2 + \|x\|^2 - \sum_1^n |((x, \theta_i))|^2$$

Certainly, d^2 is minimized when a_i is the Fourier coefficient $a_i = ((x, \theta_i))$. For this choice of a_i, we see that the element x is equal to the sum of two orthogonal elements, i.e.

$$x = \left(\sum_1^n a_i\,\theta_i\right) + \left(x - \sum_1^n a_i\,\theta_i\right)$$

where the lengths of the elements

$$x, \sum_1^n a_i\,\theta_i, \quad \text{and} \quad x - \sum_1^n a_i\,\theta_i$$

satisfy the Pythagorean theorem (or Bessel's identity):

$$\|x\|^2 = \sum_1^n |a_i|^2 + d^2_{\min}$$

Since $d^2_{\min} \geqslant 0$, we have *Bessel's inequality*:

$$\|x\|^2 \geqslant \sum_1^n |((x, \theta_i))|^2 \qquad \text{Q.E.D.}$$

4.4. Linear independence

Definition 4.40 A finite number of elements $x_1, x_2, \ldots, x_n$ of von Neumann space $\mathbf{V}$ are called *linearly independent* if the condition $a_1 x_1 + a_2 x_2 + \ldots + a_n x_n = 0$ (where a_i are complex numbers) implies $a_1 = a_2 = \ldots = a_n = 0$. Elements that are not linearly independent are called *linearly dependent.*

Clearly, every finite orthonormal set $\theta_1, \ldots, \theta_n$ is linearly independent, for it follows from $a_1\,\theta_1 + \ldots + a_n\,\theta_n = 0$, by forming the inner product with θ_i, that $a_i = ((0, \theta_i)) = 0$.

Theorem 4.40 (*Gram–Schmidt Orthogonalization*) If $x_1, \ldots, x_n$ are linearly independent elements in $\mathbf{V}$, then there is an orthonormal set $\mathbf{O}$ of elements $\theta_1, \ldots, \theta_n$ in $\mathbf{V}$.

Proof. Using the Gram–Schmidt orthogonalization process, we form

$$\begin{aligned} y_1 &= x_1 & \theta_1 &= y_1/\|y_1\|, \\ y_2 &= x_2 - ((x_2, \theta_1))\theta_1 & \theta_2 &= y_2/\|y_2\|, \\ y_3 &= x_3 - ((x_3, \theta_1))\theta_1 - ((x_3, \theta_2))\theta_2 & \theta_3 &= y_3/\|y_3\|, \ldots \end{aligned}$$

Each θ_j is possible since $\|y_j\| \neq 0$. For if it were not possible, then $y_j = 0$ and x_j would be a linear combination of $\theta_1, \ldots, \theta_{j-1}$, i.e. a linear combination of $x_1, \ldots, x_{j-1}$, contrary to hypothesis. The orthogonalization process may be continued until the set $\theta_1, \ldots, \theta_n$ is obtained. Furthermore, we see that x_j is a linear combination of $\theta_1, \ldots, \theta_j$, and that θ_j is a linear combination of $x_1, \ldots, x_j$, for $j = 1, 2, \ldots, n$. Q.E.D.

4.5 Unitary space

Let $\mathbf{V}$ be a von Neumann space.

Postulate P'_4 The maximum number of linearly independent elements in $\mathbf{V}$ is n; that is, there exist n elements $x_1, \ldots, x_n$ in $\mathbf{V}$ such that $a_1 x_1 + \ldots + a_n x_n = 0$ implies $a_1 = \ldots = a_n = 0$, while if $y_1, \ldots, y_{n+1}$ are arbitrary elements of $\mathbf{V}$, there are complex numbers not all zero such that $a_1 y_1 + \ldots + a_{n+1} y_{n+1} = 0$.

Definition 4.50 Any von Neumann space $\mathbf{V}$ for which Postulate P'_4 is true is called (*n-dimensional*) *unitary space*, and is denoted by $\mathbf{V}_n$.

Definition 4.51 Let $\mathbf{U}_n$ denote the class of all n-tuples $\alpha = (a_1, \ldots, a_n)$ of complex numbers $a_1, \ldots, a_n$. Define the null element to be $0 = (0, \ldots, 0)$. For any n-tuples $\alpha = (a_1, \ldots, a_n)$ and $\beta = (b_1, \ldots, b_n)$ of $\mathbf{U}_n$, define addition by $\alpha + \beta = (a_1 + b_1, \ldots, a_n + b_n)$, multiplication by a complex number c by $c\alpha = (ca_1, \ldots, ca_n)$ and inner product by $((\alpha, \beta)) = a_1 \bar{b}_1 + \ldots + a_n \bar{b}_n$. Then $\mathbf{U}_n$ is called *unitary n-tuple space*, or complex Euclidean n-tuple space.

Since unitary n-tuple space $\mathbf{U}_n$ satisfies Postulates $\mathrm{P}_1, \mathrm{P}_2, \mathrm{P}_3$, and P'_4, $\mathbf{U}_n$ is an n-dimensional unitary space $\mathbf{V}_n$.

Theorem 4.50 (Weyl, 1928) Let $\mathbf{V}_n$ be an arbitrary n-dimensional unitary space. Then $\mathbf{V}_n$ is *isomorphic* with n-tuple space $\mathbf{U}_n$; i.e. $\mathbf{V}_n$ can be put in one-to-one correspondence with $\mathbf{U}_n$ so that if the correspondence is indicated by $x \leftrightarrow \alpha = (a_1, \ldots, a_n)$ and $y \leftrightarrow \beta = (b_1, \ldots, b_n)$, then $x + y \leftrightarrow \alpha + \beta$, $cx \leftrightarrow c\alpha$ (c a complex number), and $((x, y)) = ((\alpha, \beta))$.

Proof. There is an orthonormal set $\theta_1, \ldots, \theta_n$ in $\mathbf{V}_n$ by Theorem 4.40, so that elements x and y of $\mathbf{V}_n$ have the unique Fourier representations $x = a_1 \theta_1 + \ldots + a_n \theta_n$, with $a_i = ((x, \theta_i))$ and $y = b_1 \theta_1 + \ldots + b_n \theta_n$, with $b_i = ((y, \theta_i))$. Hence let $x \leftrightarrow \alpha = (a_1, \ldots, a_n)$ and $y \leftrightarrow \beta = (b_1, \ldots, b_n)$. That is, α is the n-tuple consisting of the Fourier coefficients of x, and β is the n-tuple of Fourier coefficients of y. It follows that $x + y \leftrightarrow \alpha + \beta$ and $cx \leftrightarrow c\alpha$. On the other hand to each n-tuple $(a_1, \ldots, a_n)$ there corresponds an element $x = a_1 \theta_1 + \ldots + a_n \theta_n$ of $\mathbf{V}_n$ since $\mathbf{V}_n$ is linear. Moreover, we have

$$((x, y)) = ((a_1 \theta_1 + \ldots + a_n \theta_n, b_1 \theta_1 + \ldots + b_n \theta_n))$$
$$= a_1 \bar{b}_1 + \ldots + a_n \bar{b}_n = ((\alpha, \beta)).$$

It also follows that an orthonormal set $\theta_1, \ldots, \theta_k$ is maximal in $\mathbf{V}_n$ if and only if $k = n$. Q.E.D.

The Weyl Theorem 4.50 tells us that all unitary spaces of dimension n are formally identical, and hence can only differ by the different interpretations we attach to their elements.

4.6 Linear manifolds

Definition 4.60 A *linear manifold* **M** is a subset of von Neumann space **V** such that if x and y are in **M**, then $ax+by$ is also in **M** for every pair of complex numbers a and b.

Because **V** is a complete metric space, it follows that **M** is a complete metric space (i.e. each fundamental sequence in **M** converges to an element in **M**) if and only if **M** is closed (i.e. **M** contains all its limit elements). A linear manifold that is closed is called a *closed linear manifold.*

Let **A** be any set in **V**. There always exists at least one linear manifold **M** that contains **A**, because the space **V** itself is a linear manifold that contains **A**. The common part, or intersection, of all linear manifolds containing **A** is the smallest linear manifold containing **A**. This smallest linear manifold containing **A** is denoted by $\{\mathbf{A}\}$ and is called the *linear manifold spanned by* **A**.

Similarly, the smallest closed linear manifold containing **A** is denoted by $[\mathbf{A}]$ and is called the *closed linear manifold spanned by* **A**. It follows that $[\mathbf{A}]\supset\{\mathbf{A}\}\supset\mathbf{A}$. Moreover, $\{\mathbf{A}\}$ and $[\mathbf{A}]$ may be constructed as follows: Let x_i belong to **A**. Then $\{\mathbf{A}\}$ is the set of all finite linear combinations $a_1 x_1+\ldots+a_n x_n$ where $a_1,\ldots,a_n$ are complex numbers, and $n = 1,2,\ldots$. If we add to $\{\mathbf{A}\}$ all its limit elements, we obtain $[\mathbf{A}]$. If **A** contains only a finite number of elements $x_1,\ldots,x_n$, then $\{\mathbf{A}\} = [\mathbf{A}]$.

Let **M** be any subset of **V**. Then, clearly,

(1) **M** satisfies Postulate P_1 if and only if **M** is a linear manifold.
(2) **M** always satisfies Postulate P_2 (since we assume, of course, that the definition of $((x,y))$ remains unchanged).
(3) **M** satisfies Postulate P_3 if and only if **M** is closed.

Thus the closed linear manifolds **M** of von Neumann space **V** are in their own right von Neumann spaces. If **V** is unitary space $\mathbf{V}_n$, then clearly any closed linear manifold (subset) of $\mathbf{V}_n$ is also a unitary space with dimension less than or equal to n. If **V** is not finite-dimensional (i.e. **V** violates Postulate P'_4), then any closed linear manifold of **V** may or may not be a finite-dimensional unitary space $\mathbf{V}_n$.

An element x in **V** is said to be orthogonal to a linear manifold **M** in **V** if x is orthogonal to every element of **M**. Two linear manifolds in **V** are said to be orthogonal if every element of one is orthogonal to every element of the other.

4.7 Hilbert space

Let **V** be a von Neumann space.

Postulate P_4 There are arbitrarily many linearly independent elements in **V**; i.e. for every $n = 1,2,\ldots$, there is a set of n linearly independent elements of **V**.

Postulate P_5 The space $\mathbf{V}$ is separable, i.e. there is a sequence $x_1, x_2, \ldots$ in $\mathbf{V}$ that is everywhere dense in $\mathbf{V}$.

Definition 4.70 (von Neumann, 1929) Any von Neumann space $\mathbf{V}$ for which Postulates P_4 and P_5 are true is called a *Hilbert space* and is denoted by $\mathbf{V}_\infty$.

We now wish to outline some theorems about Hilbert space. Complete proofs may be found in von Neumann (1929, 1932, 1933) and Stone (1932).

Theorem 4.70 Each orthonormal set $\mathbf{O}$ in Hilbert space $\mathbf{V}_\infty$ is either finite or countably infinite. If $\mathbf{O}$ is maximal, it is certainly infinite.

Because of this theorem, we see that every orthonormal set $\mathbf{O}$ may be written as a finite or infinite sequence $\theta_1, \theta_2, \ldots$ where $((\theta_i, \theta_i)) = 1$ and $((\theta_i, \theta_j)) = 0$ for $i \neq j$. In order for $\mathbf{O}$ to be maximal, the sequence $\theta_1, \theta_2, \ldots$ must be non-terminating, but a non-terminating orthonormal sequence $\theta_1, \theta_2, \ldots$ is not necessarily maximal. The following theorems about orthonormal sets $\theta_1, \theta_2, \ldots$ are valid for both infinite and finite sets. For finite sets $\theta_1, \ldots, \theta_n$, all terms for $i > n$, of course, must be regarded as absent.

Theorem 4.71 Let $\theta_1, \theta_2, \ldots$ be an orthonormal set in Hilbert space $\mathbf{V}_\infty$. Then we have *Bessel's inequality* $\|x\|^2 \geqslant \sum_1^\infty |((x, \theta_i))|^2$ and all series $\sum_1^\infty ((x, \theta_i))\overline{((y, \theta_i))}$ are absolutely convergent for every x and y in $\mathbf{V}_\infty$.

Proof. By Theorem 4.30, Bessel's inequality holds for $\theta_1, \ldots, \theta_n$. Letting $n \to \infty$, we see that the series $\sum_1^\infty |((x, \theta_i))|^2$ has a sum that cannot exceed $\|x\|^2$ and is convergent, so Bessel's inequality is valid in this case. The second part of the theorem follows from the inequality: $|((x, \theta_i))\overline{((y, \theta_i))}| \leqslant \frac{1}{2}|((x, \theta_i))|^2 + \frac{1}{2}|((y, \theta_i))|^2$. Q.E.D.

Theorem 4.72 Let $\theta_1, \theta_2, \ldots$ be an orthonormal set in Hilbert space $\mathbf{V}_\infty$. Then the series $\sum_1^\infty a_i \theta_i$ converges (i.e. the sequence $\{x_n\} = \left\{\sum_1^n a_i \theta_i\right\}$ converges) if and only if the series $\sum_1^\infty |a_i|^2$ converges. If $x = \sum_1^\infty a_i \theta_i$ (convergence is assumed), then a_i is the Fourier coefficient $a_i = ((x, \theta_i))$.

Theorem 4.73 Let $\theta_1, \theta_2, \ldots$ be an orthonormal set in Hilbert space $\mathbf{V}_\infty$. Then for every x in $\mathbf{V}_\infty$, the Fourier series

$$x' = \sum_1^\infty a_i \theta_i, \quad a_i = ((x, \theta_i)),$$

is always convergent, and $x - x'$ is orthogonal to $\theta_1, \theta_2, \ldots$.

Proof. The convergence follows from Theorems 4.71 and 4.72. By Theorem 4.72, $((x-x', \theta_i)) = ((x, \theta_i)) - ((x', \theta_i)) = a_i - a_i = 0$. Q.E.D.

Theorem 4.74 Let $\theta_1, \theta_2, \ldots$ be an orthonormal set in a closed linear manifold $\mathbf{M}$ of Hilbert space. In order for $\theta_1, \theta_2, \ldots$ to be maximal in $\mathbf{M}$, each of the following conditions is necessary and sufficient:

(1) The closed linear manifold spanned by $\theta_1, \theta_2, \ldots$ is $\mathbf{M}$.

(2) Every x in $\mathbf{M}$ has the Fourier series representation $x = \sum_{i=1}^\infty a_i \theta_i, \quad a_i = ((x, \theta_i))$.

(3) The Parseval identity $((x, y)) = \sum_{i=1}^\infty ((x, \theta_i)) \overline{((y, \theta_i))}$ is true for every x and y in $\mathbf{M}$.

In particular, we may let $\mathbf{M}$ be $\mathbf{V}_\infty$ itself.

Theorem 4.75 To each sequence $x_1, x_2, \ldots$ in Hilbert space there corresponds an orthonormal sequence $\theta_1, \theta_2, \ldots$ (finite or infinite) that spans the same linear manifold as $x_1, x_2 \ldots$.

Proof. Replace $x_1, x_2, \ldots$ by linearly independent $y_1, y_2, \ldots$ which spans the same linear manifold. The sequence $y_1, y_2, \ldots$ may terminate, or not. Use the Gram–Schmidt orthogonalization process (of Theorem 4.40) to construct the orthonormal set $\theta_1, \theta_2, \ldots$. Since each y_j is a linear combination of $\theta_1, \ldots, \theta_j$, and conversely each θ_j is a linear combination of $y_1, \ldots, y_j$, it is clear that $y_1, y_2, \ldots$ and $\theta_1, \theta_2, \ldots$ span the same linear manifold. Q.E.D.

Theorem 4.76 Let $\mathbf{M}$ be a closed linear manifold of Hilbert space. Then there is a maximal orthonormal set $\theta_1, \theta_2, \ldots$ in $\mathbf{M}$ (i.e. the closed linear manifold spanned by the orthonormal set $\theta_1, \theta_2, \ldots$ is $\mathbf{M}$).

Proof. If $\mathbf{M}$ is isometric to n-tuple space, the proof follows readily. Otherwise we must call upon Postulate P_5 ($\mathbf{V}_\infty$ is separable) from which it follows $\mathbf{M}$ is separable. Hence we may find a sequence $x_1, x_2, \ldots$ everywhere dense in $\mathbf{M}$. The closed linear manifold spanned by $x_1, x_2, \ldots$ is equal to $\mathbf{M}$. By Theorem 4.75 we can find an orthonormal sequence $\theta_1, \theta_2, \ldots$ that spans the same closed linear manifold as $x_1, x_2, \ldots$, namely $\mathbf{M}$. Q.E.D.

Definition 4.71 Let **H** denote the class of all sequences $\alpha = (a_1, a_2, \ldots)$ of complex numbers $a_1, a_2, \ldots$, such that $\sum_1^\infty |a_i|^2$ is finite. Define the null element to be $0 = (0, 0, \ldots)$. For any sequences $\alpha = (a_1, a_2, \ldots)$ and $\beta = (b_1, b_2, \ldots)$ define addition by $\alpha + \beta = (a_1 + b_1, a_2 + b_2, \ldots)$, multiplication by a complex number c by $c\alpha = (ca_1, ca_2, \ldots)$, and inner product by

$$((\alpha, \beta)) = a_1 \bar{b}_1 + a_2 \bar{b}_2 + \ldots$$

Then **H** is called *Hilbert sequence space.*

It can be verified that **H** satisfies Postulates P_1, P_2, P_3, P_4, and P_5 and so is a Hilbert space.

Theorem 4.77 (von Neumann, 1929) Let $\mathbf{V}_\infty$ be an arbitrary Hilbert space. Then $\mathbf{V}_\infty$ is isomorphic with Hilbert sequence space **H**: i.e. $\mathbf{V}_\infty$ can be put in one-to-one correspondence with **H** so that if the correspondence is indicated by $x \leftrightarrow \alpha = (a_1, a_2, \ldots)$ and $y \leftrightarrow \beta = (b_1, b_2, \ldots)$, then $x + y \leftrightarrow \alpha + \beta$, $cx \leftrightarrow c\alpha$ (c a complex number), and $((x, y)) = ((\alpha, \beta))$.

Proof. By Theorems 4.70 and 4.76, there is a maximal orthonormal set $\theta_1, \theta_2, \ldots$ in $\mathbf{V}_\infty$, which must be countably infinite. By Theorem 4.74 each element in $\mathbf{V}_\infty$ has the Fourier representation $x = \sum_1^\infty a_i \theta_i$, $a_i = ((x, \theta_i))$. Thus to each element x there is one and only one sequence of Fourier coefficients $(a_1, a_2, \ldots)$ such that, by Theorem 4.71, $\sum_1^\infty |a_i|^2$ is convergent. Conversely, to each sequence $(a_1, a_2, \ldots)$ such that $\sum_1^\infty |a_i|^2$ is finite, there corresponds one and only one $x = \sum_1^\infty a_i \theta_i$ of $\mathbf{V}_\infty$ with this sequence as Fourier coefficients (Theorem 4.72). Moreover, by Parseval's identity (Theorem 4.74) we have $((x, y)) = a_1 \bar{b}_1 + a_2 \bar{b}_2 + \ldots = ((\alpha, \beta))$. Q.E.D.

The von Neumann Theorem 4.77 tells us that all Hilbert spaces are formally identical and hence can only differ by the different interpretations we attach to their elements. Postulate P_4 requires Hilbert space to have infinite dimension, and Postulate P_5 restricts the dimension to be no more than countably infinite. Hence a Hilbert space has countably infinite dimension, i.e. *dimension number aleph nought* $\aleph_0$. A closed linear manifold of a Hilbert space is either a Hilbert space or an n-dimensional unitary space.

CHAPTER V

LINEAR OPERATORS

5.1 Linear operators

In this chapter, unless otherwise stated, we shall assume that the space $\mathbf{V}$ is either Hilbert space $\mathbf{V}_\infty$ (with countably infinite dimension) or unitary space $\mathbf{V}_n$ (with finite dimension n).

Definition 5.10 A single-valued *operator* in $\mathbf{V}$ is a transformation T which is defined over some subset $\mathbf{D}$ of $\mathbf{V}$ and which has exactly one value xT in $\mathbf{V}$ corresponding to each element x of $\mathbf{D}$. The set $\mathbf{D}$ is called the *domain*, and the set of all values xT is called the *range* $\mathbf{R}$, of the operator. An operator is called *linear* if $\mathbf{D}$ and $\mathbf{R}$ are linear manifolds.

We assume throughout that all operators are single-valued, linear operators whose domains $\mathbf{D}$ are everywhere dense. Clearly an operator T is linear if and only if, whenever $x_1, \ldots, x_n$ are elements of $\mathbf{D}$, $a_1 x_1 + \ldots + a_n x_n$ is in $\mathbf{D}$, and

$$(a_1 x_1 + \ldots + a_n x_n) T = a_1 (x_1 T) + \ldots + a_n (x_n T)$$

where $a_1, \ldots, a_n$ are complex numbers.

Suppose that T establishes a one-to-one correspondence between $\mathbf{D}$ and $\mathbf{R}$ (i.e. distinct elements of $\mathbf{D}$ correspond to distinct elements of $\mathbf{R}$). Then there is an operator T^{-1}, whose domain is $\mathbf{R}$ and range is $\mathbf{D}$, such that $xTT^{-1} = x$ for x in $\mathbf{D}$ and $xT^{-1}T = x$ for x in $\mathbf{R}$. The operator T^{-1} is called the *inverse* of T. The operator I such that $xI = x$ for all x in $\mathbf{V}$ is called the *identity operator*.

Example 5.10 In terms of matrix algebra, an n-tuple $\alpha = (a_1, \ldots, a_n)$ in $\mathbf{U}_n$ is called a *row vector*. The inner product of α and $\beta = (b_1, \ldots, b_n)$ is given by the matrix multiplication

$$((\alpha, \beta)) = \alpha\bar{\beta}' = a_1 \bar{b}_1 + \ldots + a_n \bar{b}_n$$

where $\bar{\beta}'$ is the transpose (indicated by a prime) of the complex conjugate (indicated by a bar) of the row vector β. A linear operator T is represented by an *n by n matrix* $T = [t_{ij}]$ whose elements t_{ij} $(i,j = 1, \ldots, n)$ are complex numbers. If $\alpha = (a_1, \ldots, a_n)$, then αT is given by the matrix multiplication

$$\alpha T = \left(\sum_1^n a_i t_{i1}, \ldots, \sum_1^n a_i t_{in}\right)$$

Definition 5.11 Two linear operators T and T^* are called *adjoint* if they have the same domain $\mathbf{D}$, and if, in $\mathbf{D}$, $((xT, y)) = ((x, yT^*))$. The operator H is said to be *Hermitian* if $H^* = H$. A Hermitian operator H is called *definite* if $((xH, x)) \geqslant 0$.

If H is Hermitian, the Hermitian form $((xH, x))$ is always real since it equals $((x, xH)) = \overline{((xH, x))}$.

Example 5.12 Let α, β be n-tuples in unitary space $\mathbf{U}_n$. In order for $((\alpha H, \beta)) = ((\alpha, \beta H))$, i.e. for $\alpha H \bar{\beta}' = \alpha \bar{H}' \bar{\beta}'$, we must have $H = \bar{H}'$. Thus a *Hermitian matrix* H is one that is equal to the transpose of its complex conjugate. A real Hermitian matrix is a *symmetric matrix*.

Definition 5.12 A linear operator U such that $((xU, yU)) = ((x, y))$ for every pair of elements x, y in its domain $\mathbf{D}$ is called an *isometric operator*. An isometric operator whose domain $\mathbf{D}$ and range $\mathbf{R}$ each coincide with the entire space $\mathbf{V}$ (i.e. $\mathbf{D} = \mathbf{R} = \mathbf{V}$) is called a *unitary operator*.

Because isometric and unitary operators preserve inner products, they also preserve distances $\|yU - xU\| = \|y - x\|$.

Theorem 5.10 The inverse U^{-1} and adjoint U^* of a unitary operator U are also unitary operators and are equal (i.e. $U^{-1} = U^*$).

Proof. The equation $\|xU - yU\| = \|x - y\|$ holds for every pair of elements x, y in $\mathbf{V}$. That is, $xU = yU$ if and only if $x = y$ so that U has an inverse U^{-1} with domain and range $\mathbf{V}$. In $((xU, yU)) = ((x, y))$ replace y by zU^{-1} to obtain $((xU, z)) = ((x, zU^{-1}))$ for every x, z in $\mathbf{V}$. Hence U^* exists and coincides with U^{-1}. In $((xU, yU)) = ((x, y))$ replace x and y by wU^{-1} and zU^{-1} to obtain $((w, z)) = ((wU^{-1}, zU^{-1}))$, showing that U^{-1} is a unitary operator. Q.E.D.

Example 5.13 Let α and β be n-tuples in unitary space U_n. A *unitary matrix* U is one for which $((\alpha U, \beta U)) = ((\alpha, \beta))$ for every α, β. Hence we have $\alpha U \bar{U}' \bar{\beta}' = \alpha \bar{\beta}'$ so that $U \bar{U}' = I$, the identity matrix, and $U^* = U^{-1} = \bar{U}'$. A real unitary matrix is an *orthogonal matrix*.

5.2 Projection operators

Definition 5.20 If $\mathbf{M}$ and $\mathbf{N}$ are closed linear manifolds and if $\mathbf{N}$ is a subset of $\mathbf{M}$, then the set of all elements of $\mathbf{M}$ that are orthogonal to $\mathbf{N}$ is denoted by $\mathbf{M} - \mathbf{N}$, and is called the *orthogonal complement of* $\mathbf{N}$ *with respect to* $\mathbf{M}$.

The set $\mathbf{M} - \mathbf{N}$ is a closed linear manifold. In particular, $\mathbf{V} - \mathbf{M}$ is the set of all elements in $\mathbf{V}$ orthogonal to $\mathbf{M}$, and is called the orthogonal complement of $\mathbf{M}$.

Theorem 5.20 *The Projection Theorem.* If $\mathbf{M}$ is a closed linear manifold in $\mathbf{V}$, then every element x of $\mathbf{V}$ can be resolved in one and only one way into two components, $x = y + z$, where y is in $\mathbf{M}$ and z is in $\mathbf{V} - \mathbf{M}$.

Proof. By Theorem 4.76 there is a maximal orthonormal set $\theta_1, \theta_2, \ldots$ in $\mathbf{M}$. Let $y = \sum_i ((x, \theta_i))\, \theta_i$. This series converges by Theorem 4.73, and clearly y is in $\mathbf{M}$. Moreover, $z = x - y$ is orthogonal to $\theta_1, \theta_2, \ldots$, and hence to $\mathbf{M}$, so that z belongs to $\mathbf{V} - \mathbf{M}$. Uniqueness can be established by showing that another resolution $x = y' + z'$ would lead to a contradiction. Q.E.D.

The concept of projection corresponds to that in Euclidean geometry.

Definition 5.21 If $\mathbf{M}$ is a closed linear manifold in $\mathbf{V}$, if x is an element of $\mathbf{V}$, and if $x = y + z$ where y is in $\mathbf{M}$ and z is in $\mathbf{V} - \mathbf{M}$, then y is called the *projection of x in* $\mathbf{M}$ and z is called the *normal from x to* $\mathbf{M}$. The *projection operator* $P_{\mathbf{M}}$ is defined by $xP_{\mathbf{M}} \equiv y$; i.e. the projection operator $P_{\mathbf{M}}$ is the operator that assigns to each x of $\mathbf{V}$ its projection, $xP_{\mathbf{M}}$, in $\mathbf{M}$.

Definition 5.22 A linear operator E, with domain $\mathbf{V}$, is said to be *idempotent* if $xE^2 = xE$ for every x in $\mathbf{V}$. The range of an idempotent operator is called the *eigenmanifold* of the operator.

Theorem 5.21 The projection operator $P_{\mathbf{M}}$ is a Hermitian, idempotent operator with eigenmanifold $\mathbf{M}$.

Proof. Let $x = x_1 + x_2$ and $y = y_1 + y_2$ where x_1 and y_1 are in $\mathbf{M}$ and x_2 and y_2 are in $\mathbf{V} - \mathbf{M}$. Since x_1 and y_1 are orthogonal to x_2 and y_2, we have

$$((xP_{\mathbf{M}}, y)) = ((x_1, y_1 + y_2)) = ((x_1, y_1)) = ((x_1 + x_2, y_1)) = ((x, yP_{\mathbf{M}}))$$

so $P_{\mathbf{M}}$ is Hermitian. Because $xP_{\mathbf{M}}$ belongs to $\mathbf{M}$, it has the unique resolution $xP_{\mathbf{M}} = xP_{\mathbf{M}} + 0$ by Theorem 5.20, and hence $(xP_{\mathbf{M}})\, P_{\mathbf{M}} = xP_{\mathbf{M}}$, so $P_{\mathbf{M}}$ is idempotent. We see that $\mathbf{M}$ is the set of all solutions of the equation $xP_{\mathbf{M}} = x$, and $\mathbf{V} - \mathbf{M}$ is the set of all solutions of the equation $xP_{\mathbf{M}} = 0$. All $xP_{\mathbf{M}}$ belong to $\mathbf{M}$ by Definition 5.21, and each x' of $\mathbf{M}$ is equal to an $xP_{\mathbf{M}}$, in fact to $x'P_{\mathbf{M}}$. Hence the closed linear manifold $\mathbf{M}$ is the range, or eigenmanifold, of $P_{\mathbf{M}}$. Q.E.D.

The following theorem characterizes a projection operator independently of $\mathbf{M}$.

Theorem 5.22 An operator E is a projection operator if and only if E is a Hermitian, idempotent operator. The eigenmanifold $\mathbf{M}$ of a projection operator E is uniquely determined by E, and $E = P_{\mathbf{M}}$.

Proof. A projection operator is Hermitian and idempotent by Theorem 5.21. Conversely, suppose E is Hermitian and idempotent. Let $\mathbf{M}$ be the closed linear manifold spanned by all the xE. Because

$$((xE, y-yE)) = ((xE, y)) - ((xE, yE)) = ((xE, y)) - ((xE^2, y)) = 0$$

we see that $y-yE$ is orthogonal to all xE. The elements (of $\mathbf{V}$) orthogonal to $y-yE$ form a closed linear manifold. Hence they include $\mathbf{M}$ as well as xE, and so $y-yE$ belongs to $\mathbf{V}-\mathbf{M}$. The resolution of y (by Theorem 5.20) is then $y = yE+(y-yE)$ where yE belongs to $\mathbf{M}$, and so $yP_{\mathbf{M}} = yE$, and E is the projection operator $P_{\mathbf{M}}$. The eigenmanifold $\mathbf{M}$ of a projection operator E is the set of all solutions of the equation $xE = x$. Q.E.D.

Let E and F be projection operators with eigenmanifolds $\mathbf{M}$ and $\mathbf{N}$ respectively, i.e. $E = P_{\mathbf{M}}$ and $F = P_{\mathbf{N}}$. The operators E and F are said to *commute* if $EF = FE$. We see that the operator EF is Hermitian, $((xEF, y)) = ((xE, yF)) = ((x, yFE))$, and idempotent

$$(EF)^2 = EFEF = EEFF = E^2F^2 = EF$$

if and only if E and F commute. Hence EF is a projection operator if and only if E and F commute. In this case, the eigenmanifold of the projection operator EF is the closed linear manifold $\mathbf{M} \cup \mathbf{N}$ consisting of elements common to $\mathbf{M}$ and $\mathbf{N}$.

Let E and F be projection operators with eigenmanifolds $\mathbf{M}$ and $\mathbf{N}$, respectively. Then $E+F$ is clearly Hermitian. Also $E+F$ is idempotent, $(E+F)^2 = E^2+EF+FE+F^2 = E+EF+FE+F = E+F$, if and only if $EF+FE = 0$. But, it can be shown that $EF+FE = 0$ if and only if $EF = 0$ (or, equivalently, if and only if $FE = 0$). Hence $E+F$ is a projector if and only if $EF = 0$ (or, equivalently, if and only if $FE = 0$). The condition $EF = 0$ means that $\mathbf{M}$ is orthogonal to $\mathbf{N}$. In this case, the eigenmanifold of the projection operator $E+F$ is the closed linear manifold $[\mathbf{M}, \mathbf{N}]$ spanned by the set consisting of $\mathbf{M}$ and $\mathbf{N}$. The closed linear manifold $[\mathbf{M}, \mathbf{N}]$ is usually denoted by $\mathbf{M}+\mathbf{N}$. Since only a finite number of closed linear manifolds $\mathbf{M}, \mathbf{N}$ are present, and since $\mathbf{M}$ is orthogonal to $\mathbf{N}$, it can be shown that the linear manifold $\{\mathbf{M}, \mathbf{N}\}$ spanned by $\mathbf{M}$ and $\mathbf{N}$ is equal to the closed linear manifold $\mathbf{M}+\mathbf{N}$, the eigenmanifold of $E+F$.

Let E and F be projection operators with eigenmanifolds $\mathbf{M}$ and $\mathbf{N}$, respectively. It can be shown that the operator $E-F$ is a projection operator if and only if $EF = F$ (or, equivalently, if and only if $FE = F$). The condition $EF = F$ (or, equivalently, $FE = F$) means that $\mathbf{N}$ is a subset of $\mathbf{M}$. In this case, the eigenmanifold of the projection operator $E-F$ is the closed linear manifold $\mathbf{M}-\mathbf{N}$, the orthogonal complement of $\mathbf{N}$ with respect to $\mathbf{M}$.

Definition 5.23 The projection operators E and F are called *orthogonal* if their eigenmanifolds are orthogonal. The projection operator F is said to be a part of E (in symbols, $F \leqslant E$) if the eigenmanifold of F is a subset of the eigenmanifold of E.

If the projection operators E and F are orthogonal, then $EF = 0$ or also $FE = 0$, and $E+F$ is a projection operator. If F and E are projection operators and $F \leqslant E$, then $EF = F$, or also $FE = F$, so E and F commute, and EF and $E-F$ are projection operators. We see that $E \leqslant F, I-F \leqslant I-E$, and E orthogonal to $I-F$ are all equivalent.

Definition 5.24 If $E(\lambda)$ is a family of projection operators, defined as a function of the real variable λ, $-\infty < \lambda < \infty$, with properties:

(1) $xE(\lambda) \Rightarrow 0$ as $\lambda \to -\infty$, $xE(\lambda) \Rightarrow x$ as $\lambda \to \infty$, and $xE(\lambda+\epsilon) \Rightarrow xE(\lambda)$ as $\epsilon \to 0$ (ϵ positive),
(2) if $\lambda \leqslant \lambda'$, then $E(\lambda) \leqslant E(\lambda')$,

then $E(\lambda)$ is called a *resolution of the identity operator I.*

As usual, we let the symbol $\Rightarrow$ denote convergence in metric space, which is now the metric space $\mathbf{V}$. We shall introduce the symbolism $E(\lambda) \Rightarrow E(\lambda')$ to stand for $xE(\lambda) \Rightarrow xE(\lambda')$. Then Property (1) of Definition 5.24 may be equivalently written: $E(\lambda) \Rightarrow 0$ as $\lambda \to -\infty$, $E(\lambda) \Rightarrow I$ as $\lambda \to \infty$, and $E(\lambda+\epsilon) \Rightarrow E(\lambda)$ as $\epsilon \to 0$ (ϵ positive). The last condition in Property (1) states that $E(\lambda)$ is continuous on the right. Property (2) states that $E(\lambda)$ is non-decreasing, i.e. the eigenmanifolds of $E(\lambda)$ are non-decreasing as λ increases. Moreover, from Property (2) we see that the increments of $E(\lambda)$ are orthogonal, i.e. if $\lambda_1 < \lambda_2 \leqslant \lambda_3 < \lambda_4$, then the eigenmanifolds of $E(\lambda_4) - E(\lambda_3)$ and $E(\lambda_2) - E(\lambda_1)$ are orthogonal.

5.3 Function space $\mathbf{L}^2$

We now wish to consider specific realizations of von Neumann space $\mathbf{V}$.

Definition 5.30 Let $x(\omega)$ be a complex- alued, measurable function defined on the measure space $(\Omega, \mathbf{F}, \mu)$. Then the *function space* $\mathbf{L}^2$, or $\mathbf{L}^2(\mu)$ is the set of all those $x(\omega)$ for which

$$\int_\Omega |x(\omega)|^2 \mu(d\omega) \tag{5.30}$$

is finite. Two functions $x(\omega)$ and $y(\omega)$ of function space $\mathbf{L}^2$ are considered as identical if and only if $x(\omega) = y(\omega)$ almost everywhere (i.e. except for a set of measure zero).

Theorem 5.30 If we define the null function as the function which vanishes almost everywhere, and the inner product by

$$((x, y)) = \int_\Omega x(\omega)\overline{y(\omega)}\, \mu(d\omega) \tag{5.31}$$

then $\mathbf{L}^2$ is a von Neumann space $\mathbf{V}$ with the functions $x(\omega)$ as elements.

Proof. We must show that $\mathbf{L}^2$ satisfies Postulates P_1, P_2, and P_3. If $x(\omega)$ belongs to $\mathbf{L}^2$, then certainly $cx(\omega)$ belongs to $\mathbf{L}^2$ where c is a complex constant. If $x(\omega)$ and $y(\omega)$ belong to $\mathbf{L}^2$, then $((x, y))$ given by equation (5.31) is finite because $|x(\omega)\overline{y(\omega)}| \leqslant (|x(\omega)|^2+|y(\omega)|^2)/2$. Moreover $x(\omega)+y(\omega)$ belongs to $\mathbf{L}^2$ because of the identity

$$|x(\omega)+y(\omega)|^2 = |x(\omega)|^2+|y(\omega)|^2+2\mathrm{Re}\,[x(\omega)\,\overline{y(\omega)}]$$

From the above it easily follows that $\mathbf{L}^2$ satisfies Postulates P_1 and P_2. That $\mathbf{L}^2$ satisfies Postulate P_3 is the content of the Riesz–Fischer–Weyl–von Neumann Theorem. That is, the reformulation due to Weyl (1909) of the theorem of F. Riesz (1907) and E. Fischer (1907) states that if Ω is the real line and μ is Lebesgue measure, then $\mathbf{L}^2$ is a complete metric space. Von Neumann (1929, 1933) extended this theorem to cover more general measure spaces $(\Omega, \mathbf{F}, \mu)$. Q.E.D.

Convergence $x_n \Rightarrow x$ in the function space $\mathbf{L}^2$ (a metric space by Theorem 5.30!) means that the numbers $d(x_n, x) = \|x-x_n\|$ converge to zero as $n \to \infty$; i.e. that the numbers

$$\int_{\Omega} |x(\omega)-x_n(\omega)|^2\,\mu(d\omega)$$

converge to zero as $n \to \infty$. Hence $x_n \Rightarrow x$ is sometimes called "convergence in the mean" to distinguish convergence ($\Rightarrow$) in function space $\mathbf{L}^2$ from point-wise convergence $x_n(\omega) \to x(\omega)$ for a given point ω.

5.4 Variate space

In the case when the measure is probability measure $\mu = P$, then the functions $x(\omega)$ are variates, and $\mathbf{L}^2(P)$ is a *variate space* whose elements are those variates with finite second absolute moments (equation 5.30). Since P is normed measure, the function $x(\omega) = 1$ for all ω belongs to $\mathbf{L}^2$. Consequently the mean value $E(x)$ of any variate x of $\mathbf{L}^2$ exists, for $E(x) = ((x, 1))$. If $((x, y)) = E(x)\,E(\bar{y})$, the variates x and y are said to be *uncorrelated.* If x and y are uncorrelated, then $x-E(x)$ and $y-E(y)$ are orthogonal. Therefore we shall suppose throughout that all variates of variate space $\mathbf{L}^2$ are centred by subtraction of their mean values. Then the inner product $((x, y))$ is the *covariance* of x and y, the inner product $((x, x))$ is the *variance* $D^2(x)$, and the length $\|x\|$ is the *standard deviation* $D(x)$.

Example 5.40 Let the sample space Ω be countable; i.e. let Ω be a finite or infinite sequence of points $\omega_1, \omega_2, \ldots$, with probabilities $P_1, P_2, \ldots$, respectively. A variate $x(\omega)$ is a function of each point ω_j, and hence can be represented as a finite or infinite sequence $a_1, a_2, \ldots$, where $a_j = x(\omega_j)$. Variates $x(\omega)$ belong to $\mathbf{L}^2$ if $\Sigma_j |a_j|^2 P_j$ is finite. This condition, of course, is automatically satisfied if the number of points in Ω is finite. The inner product is $((x,y)) = \Sigma_j a_j b_j P_j$, where x is represented by the sequence $a_1, a_2, \ldots$ and y by $b_1, b_2, \ldots$. The $\mathbf{L}^2$ space is certainly separable (Postulate P_5), and Postulate P_4' or P_4 holds, depending upon whether there are a finite number n or an infinite number of points ω_j with $P_j > 0$. Hence this variate space $\mathbf{L}^2$ is either n-dimensional unitary space or Hilbert space.

Example 5.41 Let the sample Ω be n-dimensional Cartesian space $\mathbf{E}_n$, and let P be Lebesgue–Stieltjes measure which is absolutely continuous with respect to n-dimensional Lebesgue measure $d\omega$. Then, by the Radon–Nikodym Theorem, we have

$$P(A) = \int_A f(\omega)\, d\omega$$

where $f(\omega)$ can be chosen to be a real, non-negative, measurable function. Variates $x(\omega)$ belong to $\mathbf{L}^2$ if

$$\|x\|^2 = \int_\Omega |x(\omega)|^2 P(d\omega) = \int_\Omega |x(\omega)|^2 f(\omega)\, d\omega$$

is finite. Unless $\mathbf{L}^2$ contains only the null variate 0, and hence has dimension 0, it can be shown that $\mathbf{L}^2$ has infinite dimension and thus satisfies Postulate P_4. The separability of $\mathbf{L}^2$ follows from the separability of $\mathbf{E}_n$ (von Neumann, 1929, 1933) and hence $\mathbf{L}^2$ satisfies Postulate P_5. Therefore this variate space $\mathbf{L}^2$ is a Hilbert space.

5.5 The stochastic integral

Let $\mathbf{V}$ be a variate space $\mathbf{L}^2$, and let $E(\lambda)$ be a resolution of the identity (Definition 5.24). If x is a variate of $\mathbf{V}$, then the family of variates $xE(\lambda)$, $-\infty < \lambda < \infty$, belongs to $\mathbf{V}$ and is a stochastic process. The increments $xE(\lambda_4) - xE(\lambda_3)$ and $xE(\lambda_2) - xE(\lambda_1)$, where $\lambda_1 < \lambda_2 \leqslant \lambda_3 < \lambda_4$, belong to eigenmanifolds of $E(\lambda_4) - E(\lambda_3)$ and $E(\lambda_2) - E(\lambda_1)$ respectively. Since these eigenmanifolds are orthogonal, the increments are orthogonal, and the stochastic process $xE(\lambda)$ is said to have *orthogonal increments.* With suitable restrictions on the λ-function $\Phi(\lambda)$, the integral (over λ) given by $\int \Phi(\lambda)\, dxE(\lambda)$ is a variate of $\mathbf{V}$ and is called a *stochastic integral,* a concept introduced in a different way by Wiener (1923).

CHAPTER VI

SPECTRAL REPRESENTATION

6.1 Multiple regression

We assume throughout the chapter that $\mathbf{V}$ is either n-dimensional unitary space or Hilbert space, as the case requires. When we refer to $\mathbf{V}$ as variate space, we mean variate space $\mathbf{L}^2(P)$.

In preparation for his generalization of the analysis of a finite number of random variables to the infinite number of the stationary process, Wold (1938) gave the following interpretation of multiple regression analysis as a linear approximation in variate space. In this way he showed that the multiple regression theory founded and developed by the English statistical school (see Yule and Kendall, 1937) can be interpreted as a particular branch of the theory of approximations in linear, inner product space (i.e. von Neumann space).

Let $x_1, \ldots, x_n$ be linearly independent variates in variate space $\mathbf{V}$ and let x be any variate in $\mathbf{V}$. The linear manifold $\mathbf{M}$ spanned by $x_1, \ldots, x_n$ is finite-dimensional and necessarily closed. We wish to approximate x in the least-squares sense by an element of $\mathbf{M}$. Every variate in $\mathbf{M}$ can be expressed uniquely by a finite linear combination $b_1 x_1 + \ldots + b_n x_n$. The complex numbers $b_1, \ldots, b_n$ are called the co-ordinates of the variate relative to the basis $x_1, \ldots, x_n$. In linear, least-squares regression analysis, we find the constants $b_1, \ldots, b_n$ that minimize the distance $d = \|x - b_1 x_1 - \ldots - b_n x_n\|$ from x to $\mathbf{M}$. We write

$$d^2 = ((x,x)) - \sum_1^n \bar{b}_i ((x, x_i)) - \sum_1^n b_i ((x_i, x)) + \sum_{i=1}^n \sum_{j=1}^n b_i \bar{b}_j ((x_i, x_j))$$

and set the derivatives of d^2 with respect to the real and imaginary parts of $b_1, \ldots, b_n$ equal to zero. Thus we obtain a system of linear equations, called the normal equations. The solution of the normal equations gives the constants $b_1, \ldots, b_n$, which are called the regression coefficients. With these coefficients, the linear combination $b_1 x_1 + \ldots + b_n x_n$ is the best linear approximation to x (by an element of $\mathbf{M}$) in the sense of the principle of least-squares.

Wold (1938) points out that the same result may be obtained in the following way. Apply the Gram–Schmidt orthogonalization process (Theorem 4.40) to $x_1, \ldots, x_n$ to obtain the orthonormal set $\theta_1, \ldots, \theta_n$ which spans the same linear manifold $\mathbf{M}$ (Theorem 4.75). By Theorem 4.30, the distance $d = \|x - a_1 \theta_1 - \ldots - a_n \theta_n\|$ from x to $\mathbf{M}$ is minimized when a_i is the Fourier coefficient $a_i = ((x, \theta_i))$. With these coefficients, the linear combination $a_1\theta_1+\ldots+a_n\theta_n$ is the linear, least-squares approximation to x (by an element of **M**), and is identical with the linear combination $b_1x_1+\ldots+b_nx_n$, where $b_1,\ldots,b_n$ are the regression coefficients.

Example 6.10 We wish to approximate the variate x as closely as possible by b_1x_1 in the least-squares sense. The minimum of $d=\|x-b_1x_1\|$ is obtained when $b=((x,x_1))/((x_1,x_1))$, the regression coefficient. The best linear approximation to x is

$$b_1x_1=((x,x_1))\,x_1/((x_1,x_1))$$

Example 6.11 Continuing Example 6.10, the Gram–Schmidt orthogonalization of x_1 is $\theta_1=x_1/\|x_1\|$. The Fourier coefficient of x is $a_1=((x,\theta_1))$, and the best linear approximation to x is $a_1\theta_1=((x,x_1))\,x_1/((x_1,x_1))$, as above.

Let $x_1,x_2,\ldots$ be an infinite sequence of linearly independent variates in **V** and let x be any variate in **V**. Let **M** be the closed linear manifold spanned by $x_1,x_2,\ldots$. We wish to approximate x in a least-squares sense by an element of **M**. We recall that **M** is the set of all finite linear combinations $b_1x_1+\ldots+b_nx_n$ ($n=1,2,\ldots$; $b_1,\ldots,b_n$ arbitrary complex numbers) together with all the limit variates of these finite linear combinations (see Section 4.6). These limit variates of **M** in general cannot be expressed by a convergent infinite series of the form $\sum_1^\infty b_ix_i$. Consequently, we cannot assume that the linear, least-squares approximation, which may be any variate in **M**, can be represented by such an infinite series $b_1x_1+b_2x_2+\ldots$. Thus, for an infinite number of variates $x_1,x_2,\ldots$, multiple regression analysis fails in that, in general, we cannot speak about an infinite number of regression coefficients $b_1,b_2,\ldots$ for an infinite regression $b_1x_1+b_2x_2+\ldots$.

We may proceed, however, along the alternative route. By the Gram–Schmidt orthogonalization process, construct the infinite orthonormal sequence $\theta_1,\theta_2,\ldots$ that spans **M** (by Theorem 4.75). By Theorem 4.74, $\theta_1,\theta_2,\ldots$ is maximal in **M**, and hence every variate in **M** has a Fourier series representation $a_1\theta_1+a_2\theta_2+\ldots$. Application of Theorem 4.71 shows that the linear least-squares approximation to x is given by the (convergent) Fourier series $a_1\theta_1+a_2\theta_2+\ldots$, where a_i is the Fourier coefficient $a_i=((x,\theta_i))$. That is, the least-squares approximation $((x,\theta_1))\,\theta_1+((x,\theta_2))\,\theta_2+\ldots$ is equal to the projection, $xP_{\mathbf{M}}$, of x on **M** (Theorems 4.73, 5.20).

6.2 Multivariate analysis and the eigenvalue problem

Returning to the finite case, let $x_1,\ldots,x_n$ be n linearly independent variates that span the closed linear manifold **M**. Consequently, every variate x in **M** can be expressed uniquely as a finite linear combination $x=a_1x_1+\ldots+a_nx_n$, where $a_1,\ldots,a_n$ are the *co-ordinates relative to* *the* (*non-orthonormal*) *basis* $x_1, \ldots, x_n$. Let the n-tuple α be these co-ordinates, i.e. $\alpha = (a_1, \ldots, a_n)$ when and only when

$$x = a_1 x_1 + \ldots + a_n x_n$$

and call the resulting unitary n-tuple space $\mathbf{U}_n$. There is a one-to-one correspondence $x \leftrightarrow \alpha$ between the variates x of $\mathbf{M}$ and the n-tuples α of $\mathbf{U}_n$, such that if $x \leftrightarrow \alpha$ and $y \leftrightarrow \beta$ (x, y in $\mathbf{M}$, and α, β in $\mathbf{U}_n$), then $x+y \leftrightarrow \alpha+\beta$ and $cx \leftrightarrow c\alpha$ (c a complex number). Inner products, however, are not preserved by this correspondence. That is, if

$$x = a_1 x_1 + \ldots + a_n x_n \leftrightarrow \alpha = (a_1, \ldots, a_n)$$

and $$y = b_1 x_1 + \ldots + b_n x_n \leftrightarrow \beta = (b_1, \ldots, b_n)$$

then $((x, y)) = ((\alpha H, \beta))$ where $H = [((x_i, x_j))]$ $(i, j = 1, \ldots, n)$ is a Hermitian matrix, and is called the *covariance matrix*. The variance of the variate x in $\mathbf{M}$ is equal to $((x, x)) = ((\alpha H, \alpha))$.

A variate $x = a_1 x_1 + \ldots + a_n x_n$ in $\mathbf{M}$ is said to be *normalized* if its co-ordinate n-tuple $\alpha = (a_1, \ldots, a_n)$ has length one, i.e. $\|\alpha\| = 1$. The normalized variate in $\mathbf{M}$ that has maximum variance μ_n is called the *first principal component* z_n of the variates $x_1, \ldots, x_n$. The normalized variate in $\mathbf{M}$ that is orthogonal to z_n and that has maximum variance μ_{n-1} is called the *second principal component* z_{n-1}. The normalized variate in $\mathbf{M}$ that is orthogonal to z_n, z_{n-1} and that has maximum variance μ_{n-2} is called the *third principal component* z_{n-2}. Now this process can be repeated (see Kendall, 1957) until we obtain n principal components $z_1, \ldots, z_n$ with variances $\mu_1, \ldots, \mu_n$ where $\mu_1 \leqslant \mu_2 \leqslant \ldots \leqslant \mu_n$. The principal components form a maximal orthonormal set in $\mathbf{M}$.

If the principal component z_i $(i = 1, \ldots, n)$ has the representation $z_i = c_{i1} x_1 + \ldots + c_{in} x_n$, denote its co-ordinate n-tuple by $\phi_i = (c_{i1}, \ldots, c_{in})$. Then its variance is $\mu_i = ((z_i, z_i)) = ((\phi_i H, \phi_i))$. In matrix notation, this equation becomes $\phi_i H \bar{\phi}_i' = \mu_i$. Let U be the matrix whose ith row is the row vector (n-tuple) ϕ_i $(i = 1, \ldots, n)$, and let Λ be the diagonal matrix with μ_i in the ith diagonal position $(i = 1, \ldots, n)$ and zeros off the diagonal. We note that since $\mu_1 \leqslant \mu_2 \leqslant \ldots \leqslant \mu_n$ the diagonal of Λ has the same ordering. Since $U\bar{U}' = I$, the matrix U is a unitary matrix. The relationship between μ_i and ϕ_i is then expressed by $UHU^{-1} = \Lambda$, or $UH = \Lambda U$, or $\phi_i H = \mu_i \phi_i$ $(i = 1, \ldots, n)$. We recognize this equation to be the characteristic equation of the Hermitian matrix H, the covariance matrix. In other words, the problem of finding the principal components in $\mathbf{M}$ is the well-known eigenvalue problem in $\mathbf{U}_n$, which always has a solution.

The eigenvalue problem in unitary n-tuple space

Given the Hermitian matrix H in unitary n-tuple space $\mathbf{U}_n$, find all (non-null) solutions ϕ of the *characteristic equation* $\phi H = \mu \phi$ where ϕ is an n-tuple and μ is a real number (i.e. H given, ϕ and μ to be determined). From these solutions form a maximal orthonormal set $\phi_1, \ldots, \phi_n$. The $\phi_1, \ldots, \phi_n$ are called the *eigensolutions*, and their corresponding real numbers $\mu_1, \ldots, \mu_n$, the *eigenvalues*, of the matrix H. The totality of the eigenvalues is called the *spectrum*.

The eigensolution ϕ_i is the co-ordinate n-tuple (with respect to the basis $x_1, \ldots, x_n$) of the principal component z_i, and the corresponding eigenvalue μ_i is the variance of z_i. Equivalently, as we have just seen, the eigenvalue problem is the problem of finding a unitary matrix U that reduces H to a real diagonal matrix, i.e. $UHU^{-1} = \Lambda$.

6.3 The spectral representation in unitary space

The eigenvalues μ_i need not be distinct, and any value λ_j that occurs k_j times among $\mu_1, \ldots, \mu_n$ is said to be an eigenvalue of multiplicity k_j. From now on, when we refer to the set of *distinct eigenvalues* of the Hermitian matrix H, we shall denote them by $\lambda_1, \lambda_2, \ldots, \lambda_m$, where $\lambda_1 < \lambda_2 < \ldots < \lambda_m$ $(m \leqslant n)$, and their multiplicities by $k_1, k_2, \ldots, k_m$, where $k_1 + k_2 + \ldots + k_m = n$.

The eigensolutions ϕ_i are not unique. If λ_j has multiplicity one, then its corresponding eigensolution is uniquely determined except for a factor e^{it} (t a real number) of magnitude one. If λ_j has multiplicity k_j greater than one, then its corresponding eigensolutions are not uniquely determined. (As we shall see, each λ_j has an eigenmanifold of dimension k_j associated with it. The only requirement on the eigensolutions corresponding to λ_j is that they be a maximal orthonormal set in the eigenmanifold of λ_j, and hence the eigensolutions are not uniquely determined. The eigenmanifolds of two distinct eigenvalues are orthogonal, and the eigenmanifolds of all the λ_j span the whole space.) This lack of uniqueness in the eigensolutions is serious enough to prevent any generalization of the solution of the eigenvalue problem to Hilbert space by some limiting process $n \to \infty$ in n-tuple space. That is, the arbitrarily large variations in the eigensolutions that are possible because of the incompleteness of their determination are enough to prevent any limiting process involving them from converging.

Although the unitary matrix U, whose rows are the eigensolutions, is not uniquely determined by H, Hilbert (1912) and von Neumann (1929, 1932) exploited the unitary nature of U to reformulate the eigenvalue problem. The Hilbert–von Neumann reformulation which is called the *spectral resolution problem* yields the same results as the eigenvalue problem in n-tuple space, but has the advantage that it can be extended to Hilbert space.

If α and β are n-tuples of $\mathbf{U}_n$, then the inner product

$$((\alpha, \beta)) = ((\alpha U^{-1}, \beta U^{-1}))$$

is a numerical invariant, regardless of the particular choice we make for the non-unique eigensolution matrix U, because U (and hence U^{-1})

is unitary. To decompose $((\alpha, \beta))$ into numerically invariant components, we shall establish a correspondence between $\mathbf{U}_n$, whose elements, we recall, are the co-ordinate n-tuples $\alpha = (a_1, \ldots, a_n)$ of $x = a_1 x_1 + \ldots + a_n x_n$ relative to the (non-orthonormal) basis $x_1, \ldots, x_n$, and the n-tuple space $\mathbf{W}_n$ whose elements we define to be the co-ordinate n-tuples $\rho = (r_1, \ldots, r_n)$ of $\alpha = r_1 \phi_1 + \ldots + r_n \phi_n$ relative to the orthonormal basis $\phi_1, \ldots, \phi_n$. This correspondence is possible because the eigensolutions $\phi_1, \ldots, \phi_n$ form a maximal orthonormal set in $\mathbf{U}_n$. By the Weyl Theorem 4.50, this correspondence,

$$\alpha = r_1 \phi_1 + \ldots + r_n \phi_n \leftrightarrow \rho = (r_1, \ldots, r_n)$$

(the $r_i = ((\alpha, \phi_i)) = \alpha \bar{\phi}'_i$ being the Fourier coefficients), establishes an isomorphism between $\mathbf{U}_n$ and $\mathbf{W}_n$ that preserves inner product. That is, if $\alpha \leftrightarrow \rho = (r_1, \ldots, r_n)$ and $\beta \leftrightarrow \sigma = (s_1, \ldots, s_n)$, then

$$((\alpha, \beta)) = ((\rho, \sigma)) = r_1 \bar{s}_1 + \ldots + r_n \bar{s}_n$$

In matrix notation, we have

$$\rho = (r_1, \ldots, r_n) = (\alpha \bar{\phi}'_1, \ldots, \alpha \bar{\phi}'_n) = \alpha \bar{U}' = \alpha U^{-1}$$

and similarly $\sigma = \beta U^{-1}$.

The conditions of the eigenvalue problem, which are expressed by the characteristic equation $\phi_i H = \mu_i \phi_i (i = 1, \ldots, n)$, or $H = U^{-1} \Lambda U$, may equivalently be written (since $U^* = U^{-1}$)

$$((\alpha H, \beta)) = ((\alpha U^{-1} \Lambda U, \beta)) = ((\rho \Lambda, \sigma)) = \mu_1 r_1 \bar{s}_1 + \ldots + \mu_n r_n \bar{s}_n \quad (6.30)$$

The Cartesian character of the co-ordinate system in $\mathbf{W}_n$ is expressed by

$$((\alpha, \beta)) = ((\alpha U^{-1}, \beta U^{-1})) = ((\rho I, \sigma)) = r_1 \bar{s}_1 + \ldots + r_n \bar{s}_n \quad (6.31)$$

We recall that the diagonal matrix Λ was defined to be the matrix with the eigenvalues, ordered by increasing magnitude, along its main diagonal, and zeros off the diagonal. Because of this ordering, the matrix Λ is uniquely determined for any given Hermitian matrix H. Consequently, for given H we have the unique decomposition $\Lambda = \lambda_1 Q_1 + \ldots + \lambda_m Q_m$, where Q_i is a diagonal matrix with one's in those places on its main diagonal in which λ_j occurs in Λ, and zeros elsewhere. Clearly $I = Q_1 + \ldots + Q_m$.

Using the decomposition $I = Q_1 + \ldots + Q_m$, equation (6.31) becomes

$$((\alpha, \beta)) = ((\rho (Q_1 + \ldots + Q_m), \sigma)) = ((\rho Q_1, \sigma)) + \ldots + ((\rho Q_m, \sigma)) \quad (6.32)$$

where each $((\rho Q_j, \sigma))$ $(j = 1, \ldots, m)$ is the sum of those $r_i \bar{s}_i$ (k_j in number) that are associated with the eigenvalue λ_j (of multiplicity k_j). Each $((\rho Q_j, \sigma))$ is numerically invariant under the variations of the non-unique eigensolutions ϕ_i. Thus for given α, β, H, equation (6.32) is a decomposition of $((\alpha, \beta))$ into components numerically invariant with respect to the particular choice of the non-uniquely determined eigensolutions.

Let us define P_j to be $P_j = U^{-1}Q_jU$ $(j = 1, 2, \ldots, m)$. Then the invariant components in equation (6.32) are equal to

$$((\rho Q_j, \sigma)) = ((\alpha U^{-1}Q_j, \beta U^{-1})) = ((\alpha P_j, \beta)) \quad (j = 1, \ldots, m)$$

and equation (6.32) becomes

$$((\alpha, \beta)) = ((\alpha P_1, \beta)) + \ldots + ((\alpha P_m, \beta)) = \sum_{\lambda_j < \infty} ((\alpha P_j, \beta))$$

This equation leads us to define the Hermitian form

$$((\alpha K(\lambda), \beta)) = \sum_{\lambda_j \leqslant \lambda} ((\alpha P_j, \beta)) \tag{6.33}$$

which, we see, is invariant for fixed $\alpha, \beta, H, \lambda$. Moreover, given $((\alpha K(\lambda), \beta))$ we may retrace our steps to obtain the eigensolutions and eigenvalues of H. In the Hilbert–von Neumann reformulation of the eigenvalue problem (i.e. in the spectral resolution problem), the $((\alpha K(\lambda), \beta))$ only will appear, instead of the eigensolutions and eigenvalues, so that the reformulation will be a unique formulation. To obtain the Hilbert–von Neumann spectral resolution problem, then, we must convert equations (6.30) and (6.31), which express the conditions of the eigenvalue problem, into equivalent conditions on $((\alpha K(\lambda), \beta))$. We proceed as follows:

First, we wish to show that equation (6.31) means that $K(\lambda)$ is a *resolution of the identity* (Definition 5.24). From equation (6.33) we see that $K(\lambda)$ is given by

$$K(\lambda) = \sum_{\lambda_j \leqslant \lambda} P_j = \sum_{\lambda_j \leqslant \lambda} U^{-1}Q_jU = U^{-1}\left(\sum_{\lambda_j \leqslant \lambda} Q_j\right)U \tag{6.34}$$

Clearly, from the definition of Q_j, we see that Q_j is a projection matrix (i.e. Hermitian and idempotent) and $Q_iQ_j = 0$ for $i \neq j$. Because

$$((\alpha P_j, \beta)) = ((\alpha U^{-1}Q_jU, \beta)) = ((\alpha, \beta U^{-1}Q_jU)) = ((\alpha, \beta P_j))$$

we see that P_j is Hermitian, and because

$$P_jP_j = U^{-1}Q_jUU^{-1}Q_jU = P_j$$

we see that P_j is idempotent, so P_j is a projection matrix. Since, for $i \neq j$, $P_iP_j = U^{-1}Q_iUU^{-1}Q_jU = 0$, it follows that $P_i + P_j$ is a projection matrix, and the eigenmanifold of P_i is orthogonal to the eigenmanifold of P_j. Since, by equation (6.34),

$$K(\lambda) = P_1 + P_2 + \ldots + P_j, \quad \lambda_j \leqslant \lambda < \lambda_{j+1}$$

it follows that $K(\lambda)$ is a family of projection matrices (i.e. idempotent and Hermitian) with the property that the eigenmanifolds of $K(\lambda)$ are non-decreasing as λ increases. Hence if $\lambda \leqslant \lambda'$, the eigenmanifold of $K(\lambda)$ is a subset of the eigenmanifold of $K(\lambda')$, and so $K(\lambda) \leqslant K(\lambda')$.

Moreover, equation (6.31) means that if λ is larger than all the eigenvalues (i.e. $\lambda \geqslant \lambda_m$), then $K(\lambda) = I$. If λ is smaller than all the eigenvalues (i.e. $\lambda < \lambda_1$), then $K(\lambda) = 0$. Also $K(\lambda)$ is seen to be continuous on the right. (In fact, $K(\lambda)$ is always constant, except at the eigenvalues $\lambda_1, \ldots, \lambda_m$ where it changes discontinuously.) Therefore $K(\lambda)$ satisfies the properties of Definition 5.24, and hence is a resolution of the identity.

Next, we wish to show that equation (6.30) means that $K(\lambda)$ is the so-called *resolution of the identity belonging to the Hermitian matrix H*. Using the decomposition $\Lambda = \lambda_1 Q_1 + \ldots + \lambda_m Q_m$, equation (6.30) becomes

$$
\begin{aligned}
&((\alpha H, \beta)) \\
&\quad = ((\rho(\lambda_1 Q_1 + \ldots + \lambda_m Q_m), \sigma)) = \lambda_1((\rho Q_1, \sigma)) + \ldots + \lambda_m((\rho Q_m, \sigma)) \\
&\quad = \lambda_1((\alpha P_1, \beta)) + \ldots + \lambda_m((\alpha P_m, \beta))
\end{aligned}
$$

where the components $((\alpha P_j, \beta))$, as we have seen, are numerically invariant for given α, β, H. This equation may be rewritten as

$$((\alpha H, \beta)) = \sum_{i=1}^{m} \lambda_j [((\alpha K(\lambda_j), \beta)) - ((\alpha K(\lambda_{j-1}), \beta))]$$

which, using the Stieltjes integral, is

$$((\alpha H, \beta)) = \int_{-\infty}^{\infty} \lambda d((\alpha K(\lambda), \beta)) \tag{6.35}$$

This equation is called the *spectral representation* of the Hermitian matrix H. Symbolically, for the spectral representation (6.35) we write (which for n-tuple space $\mathbf{U}_n$ is a bona fide matrix equation, as indicated by the right-hand side):

$$H = \int_{-\infty}^{\infty} \lambda dK(\lambda) = \lambda_1 P_1 + \ldots + \lambda_m P_m \tag{6.36}$$

We also write

$$\alpha = \int_{-\infty}^{\infty} d\alpha K(\lambda) = \alpha P_1 + \ldots + \alpha P_m \tag{6.37}$$

$$\alpha H = \int_{-\infty}^{\infty} \lambda d\, \alpha K(\lambda) = \lambda_1 \alpha P_1 + \ldots + \lambda_m \alpha P_m \tag{6.38}$$

That is, the *spectral representation* (6.37) exhibits α as the sum of its m projections, αP_j, on the m (orthogonal) *eigenmanifolds* of the projection matrices P_j (also called the eigenmanifolds of λ_j ($j = 1, \ldots, m$)). If each of these projections αP_j is submitted to an appropriate magnification λ_j to yield $\lambda_j \alpha P_j$, then the *spectral representation* (6.38) exhibits αH as the sum of these $\lambda_j \alpha P_j$. The eigenmanifold of P_j has dimension equal to k_j (i.e. the multiplicity of its eigenvalue λ_j), and the linear manifold spanned by all these eigenmanifolds ($j = 1, \ldots, m$) is the whole space $\mathbf{U}_n$. From the characteristic equation, we see that the eigensolutions corresponding to λ_j form a maximal orthonormal set in the λ_j-eigenmanifold. From the spectral representation (6.36) we have

$$H^2 = (\lambda_1 P_1 + \ldots + \lambda_m P_m)^2 = \lambda_1^2 P_1 + \ldots + \lambda_m^2 P_m = \int_{-\infty}^{\infty} \lambda^2 \, dK(\lambda)$$

so numerically

$$((\alpha H^2, \beta)) = \int_{-\infty}^{\infty} \lambda^2 \, d((\alpha K(\lambda), \beta))$$

When $\alpha = \beta$, this equation becomes

$$\|\alpha H\|^2 = \int_{-\infty}^{\infty} \lambda^2 \, d\|\alpha K(\lambda)\|^2 \tag{6.39}$$

since $((\alpha H^2, \alpha)) = ((\alpha H, \alpha H))$ and $((\alpha K(\lambda), \alpha)) = ((\alpha K(\lambda), \alpha K(\lambda)))$. Since $\|\alpha K(\lambda)\|^2$ is a non-decreasing function of λ, and λ^2 is non-negative, the integral in equation (6.39) must be either a finite non-negative number or plus infinity. If the integral is finite, then αH has finite length $\|\alpha H\|$, and so αH is defined. In the unitary space setting, αH always has finite length, and so αH is defined for all α. Thus the condition that the integral in equation (6.39) must be finite will have independent meaning only when these results are extended to Hilbert space. With this motivation, we have

Definition 6.30 Given the Hermitian operator H in (unitary or Hilbert) space $\mathbf{V}$, a resolution of the identity $K(\lambda)$, such that:

(1) the domain of H consists of those elements x of $\mathbf{V}$ for which the integral

$$\int_{-\infty}^{\infty} \lambda^2 \, d\|xK(\lambda)\|^2$$

is finite (i.e. xH is defined if and only if this integral is finite);

(2) when xH is defined, it has the spectral representation

$$((xH, y)) = \int_{-\infty}^{\infty} \lambda \, d((xK(\lambda), y))$$

for every y of $\mathbf{V}$,

is called the *resolution of the identity belonging to H.*

Clearly $K(\lambda)$, equation (6.34), is the unique resolution of the identity belonging to the Hermitian matrix H.

Then the Hilbert–von Neumann reformulation of the eigenvalue problem is

The spectral resolution problem for unitary n-tuple space. Given the Hermitian matrix H in $\mathbf{U}_n$, find the resolution of the identity belonging to H.

We have just seen that the solution of the eigenvalue problem for H in unitary space $\mathbf{U}_n$, as expressed by equations (6.30) and (6.31), means that there is a unique resolution of the identity $K(\lambda)$ belonging to H. It is not difficult to show the converse: from $K(\lambda)$ we may obtain the eigenvalues and a set of eigensolutions. The eigenvalue problem and the spectral resolution problem always have a solution in unitary space, and moreover the solution $K(\lambda)$ of the spectral resolution problem is unique.

6.4 Eigenvalue problem in Hilbert space

Corresponding to the eigenvalue problem in unitary space, we have

The eigenvalue problem in Hilbert space. Given the Hermitian operator H in Hilbert space $\mathbf{V}_\infty$, find all (non-null) solutions ϕ of the characteristic equation $\phi H = \lambda\phi$, where ϕ is an element of $\mathbf{V}_\infty$ and λ is a real number (i.e. H given, ϕ and λ to be determined). From these solutions form a maximal orthonormal set $\phi_1, \phi_2, \ldots$. The ϕ_i are called the *eigensolutions*, and the corresponding real numbers λ_i, the *eigenvalues* of the Hermitian operator H.

The solution of the eigenvalue problem in Hilbert space does not always exist, as the following examples show.

Example 6.40 Suppose $\mathbf{V}_\infty$ is the function space $\mathbf{L}^2$ (with Lebesgue measure) where the configuration space Ω is the real line with points the real numbers u. Let H be the operator that multiplies the function $x(u)$ by u; i.e. H is defined by $x(u)H \equiv x(u)\,u$. Because

$$((xH, y)) = \int_{-\infty}^{\infty} x(u)\,u\,\overline{y(u)}\,du = ((x, yH))$$

we see that H is Hermitian. The characteristic equation is $\phi(u)u = \lambda\phi(u)$, the solution of which is $\phi(u) = 0$ for $u \neq \lambda$, and $= k$ for $u = \lambda$, where k is arbitrary. Since $\phi(u)$ vanishes at every point of the real line, except for the point $u = \lambda$, which has Lebesgue measure zero, we see that $\phi(u)$ vanishes almost everywhere. Hence $\phi(u)$ is the null function, and so $\phi(u)$ is not an eigensolution. Thus the operator H has no eigensolutions whatever.

Example 6.41 Let $\mathbf{V}_\infty$ be the $\mathbf{L}^2$ (with Lebesgue measure) of Example 6.40 and let the operator H be defined by

$$xH \equiv \frac{h}{2\pi i}\frac{\partial x}{\partial u} \quad (-\infty < u < \infty) \tag{6.40}$$

where h is Planck's constant. For x and y in the domain of H it can be shown that $x(u)\overline{y(u)} \to 0$ as $u \to \infty$ or $-\infty$. Then we have

$$((xH, y)) - ((x, yH)) = \frac{h}{2\pi i}\int_{-\infty}^{\infty} [x'(u)\overline{y(u)} + x(u)\overline{y'(u)}]\,du$$
$$= \frac{h}{2\pi i}[x(u)\overline{y(u)}]_{-\infty}^{\infty} = 0$$

so H is Hermitian. The solutions of the characteristic equation $\phi H = \lambda\phi$ are the functions

$$\phi(u) = c(\lambda)\, e^{2\pi i \lambda u/h} \tag{6.41}$$

where c is an arbitrary function of λ. Because

$$\int_{-\infty}^{\infty} |c(\lambda)|^2\, du = \infty$$

these functions do not have finite length, and so do not belong to $\mathbf{L}^2$ (unless $c = 0$, in which case $\phi = 0$). Hence the operator H has no eigensolutions whatever.

In Hilbert space then, we may or may not be able to find a sufficient number of eigensolutions of an Hermitian operator H to form a maximal orthonormal set, i.e. the eigenvalue problem may or may not have a solution. If we cannot form a maximal orthonormal set (i.e. the eigenvalue problem for H does not have a solution), then we say that the *continuous spectrum* of the operator H exists. Any eigenvalues $\lambda_1, \lambda_2, \ldots$ that the Hermitian operator H might have are called the *discrete spectrum* of H. We see that in both Examples 6.40 and 6.41 the operator H has a *purely continuous spectrum* (i.e. the discrete spectrum is absent).

6.5 The spectral representation in Hilbert space

Because the eigenvalue problem in Hilbert space does not in general have a solution, von Neumann (1929, 1932) turned his attention to

> *The spectral resolution problem in Hilbert space.* Given the Hermitian operator H in $\mathbf{V}_\infty$, find the resolution of the identity belonging to H.

It is the spectral resolution that is required both in quantum mechanics and stationary processes, rather than the solution of the eigenvalue problem as such. The solution of the spectral resolution problem in Hilbert space is due to John von Neumann (1929). M. H. Stone (1929, 1930) arrived at similar results on unbounded Hermitian operators independently. The results of von Neumann are as follows:

An Hermitian operator H is *maximal* if it has no proper extensions, i.e. if xH is defined for all elements x where it could be defined without violation of the Hermitian nature of H. If H is non-maximal, then von Neumann showed that the spectral resolution problem has no solution. Nevertheless, every non-maximal Hermitian operator can be always extended to a maximal one, so we need to consider only maximal Hermitian operators.

Let U be the *Cayley–Moebius transformation* of the maximal Hermitian operator H. That is, U is symbolically defined by

$$U = \frac{I+iH}{I-iH} \tag{6.50}$$

which interpreted means this: zU is defined for an arbitrary element z if and only if $z = x(I-iH)$ where x belongs to the domain of H, and then $zU = x(I+iH)$. The domain $\mathbf{D}$ of U is then the set of all $x-ixH$, and the range $\mathbf{R}$ of U is the set of all $x+ixH$. Since H and U are linear, $\mathbf{D}$ and $\mathbf{R}$ are linear manifolds, and moreover they are closed linear manifolds. Because H is Hermitian, we have

$$\|z\| = \|x-ixH\| = \|x+ixH\| = \|zU\|$$

so U is an isometric operator. In fact, von Neumann showed that the Cayley–Moebius transformation establishes a one-to-one correspondence between the (closed) Hermitian operators H and the *isometric* operators U. Consequently, from the isometric operator U one may recover the Hermitian operator H by the *inverse Cayley–Moebius transformation*:

$$xH = \frac{1}{2i}(zU-z) \tag{6.51}$$

for all
$$x = \tfrac{1}{2}(zU+z) \tag{6.52}$$

Given the Hilbert space $\mathbf{V}_\infty$, the Hermitian operator H, and its corresponding isometric operator U, where U has domain $\mathbf{D}$ and range $\mathbf{R}$, the following relationships are true:

(1) H is non-maximal if and only if $\mathbf{D} \neq \mathbf{V}_\infty$ and $\mathbf{R} \neq \mathbf{V}_\infty$;
(2) H is maximal if and only if $\mathbf{D} = \mathbf{V}_\infty$ and/or $\mathbf{R} = \mathbf{V}_\infty$.

Von Neumann calls H *hypermaximal* if and only if $\mathbf{D} = \mathbf{V}_\infty$ *and* $\mathbf{R} = \mathbf{V}_\infty$, in which case, by definition, U is *unitary*. We may then write:

(2′) H is maximal, but not hypermaximal, if and only if $\mathbf{D} = \mathbf{V}_\infty$, $\mathbf{R} \neq \mathbf{V}_\infty$ *or* $\mathbf{D} \neq \mathbf{V}_\infty, \mathbf{R} = \mathbf{V}_\infty$;
(2″) H is hypermaximal if and only if U is unitary (i.e. $\mathbf{D} = \mathbf{V}_\infty$ *and* $\mathbf{R} = \mathbf{V}_\infty$).

Thus the Cayley–Moebius transformation (6.50) of a hypermaximal Hermitian operator H is a unitary operator U; and conversely, the inverse Cayley–Moebius transformation (6.51 6.52) of a unitary operator U is a hypermaximal Hermitian operator H.

If H is maximal, but not hypermaximal, then von Neumann showed that the spectral resolution problem has no solution for H. Moreover, since H is maximal, no proper extensions of H exist, and so the non-solvability of the spectral resolution problem for H is a final state of affairs. Thus we need to consider only hypermaximal Hermitian operators H.

We now turn our attention to a unitary operator U. Since, because of linearity, $\|xU-yU\| = \|(x-y)U\| = \|x-y\|$, each unitary operator U is always *continuous* (i.e. bounded in the sense of Hilbert (1912)). Hermitian operators, however, are not necessarily continuous; for example, the operators H in Examples 6.40 and 6.41, which are important in quantum mechanics, are *discontinuous* (i.e. unbounded in the sense of Hilbert). Making use of the continuity of U, von Neumann (1929) solved the spectral resolution problem for U; i.e. he established

Theorem 6.50 *The von Neumann spectral representation for the unitary operator U in Hilbert space.* Given the unitary operator U in Hilbert space $\mathbf{V}_\infty$, there is a unique family of projection operators $E(\nu)$ $(-0{\cdot}5 \leqslant \nu \leqslant 0{\cdot}5)$ with properties:

(1) $E(-0{\cdot}5) = 0$, $E(0{\cdot}5) = I$, $E(\nu)$ continuous on the right, and if $\nu \leqslant \nu'$, then $E(\nu) \leqslant E(\nu')$;
(2) zU has the *spectral representation*

$$((zU, w)) = \int_{-0\cdot5}^{0\cdot5} e^{2\pi i\nu}\, d((zE(\nu), w)) \tag{6.53}$$

for every z, w in $\mathbf{V}_\infty$.

Property (1) says that $E(\nu)$ is a resolution of the identity, and property (2) says that $E(\nu)$ belongs to the unitary operator U. Because a unitary operator is defined everywhere (i.e. its domain is $\mathbf{V}_\infty$), a description of its domain, corresponding to the condition (1) in Definition 6.30 for Hermitian operators, is unnecessary. Thus Theorem 6.50 states that, given the unitary operator U, there is a *unique resolution of the identity, $E(\nu)$, belonging to U.*

The following theorem gives rules of calculation for the Hermitian operator H.

Theorem 6.51 (von Neumann, 1931) If the Hermitian operator H has the (symbolic) spectral representation

$$H = \int_{-\infty}^{\infty} \lambda\, dK(\lambda)$$

and $\Phi(\lambda)$ and $\Psi(\lambda)$ are functions of λ, then we may symbolically write the operators

$$\Phi(H) = \int_{-\infty}^{\infty} \Phi(\lambda)\, dK(\lambda), \quad \Psi(H) = \int_{-\infty}^{\infty} \Psi(\lambda)\, dK(\lambda)$$

which have the properties

$$\Phi(H)\Psi(H) = \int_{-\infty}^{\infty} \Phi(\lambda)\Psi(\lambda)\,dK(\lambda), \quad [\Phi(H)]^* = \int_{-\infty}^{\infty} \overline{\Phi(\lambda)}\,dK(\lambda)$$

$$\Phi(H)+\Psi(H) = \int_{-\infty}^{\infty} [\Phi(\lambda)+\Psi(\lambda)]\,dK(\lambda)$$

Proof. Reference is made to von Neumann (1931) for the precise foundation of this function concept. The relations, however, may be verified by formal manipulations. For example, because $K(\lambda)$ is a projection operator with orthogonal increments, we have

$$K(\lambda)\Phi(H) = \int_{-\infty}^{\infty} \Phi(\mu)\,dK(\lambda)K(\mu) = \int_{-\infty}^{\lambda} \Phi(\mu)\,dK(\mu)$$

and so
$$\Psi(H)\Phi(H) = \int_{-\infty}^{\infty} \Psi(\lambda)\,dK(\lambda)\Phi(H)$$

$$= \int_{-\infty}^{\infty} \Phi(\lambda)\,d\int_{-\infty}^{\lambda} \Phi(\mu)\,dK(\mu)$$

which equals the desired result. Q.E.D.

Similar rules of calculation hold for the unitary operator U. If U has the spectral representation (6.53), and if $\Phi(e^{2\pi i\nu})$ is a function of $e^{2\pi i\nu}$, then

$$\Phi(U) = \int_{-0\cdot5}^{0\cdot5} \Phi(e^{2\pi i\nu})\,dE(\nu)$$

is an operator with properties corresponding to those given in Theorem 6.51. We may summarize these rules in this way: In the spectral representation of H, the (quasi-eigenvalue) λ represents the Hermitian operator H, and thus the (quasi-eigenvalue) $\Phi(\lambda)$ represents the operator $\Phi(H)$. In the spectral representation of U, the (quasi-eigenvalue) $e^{2\pi i\nu}$ represents the unitary operator U, and thus the (quasi-eigenvalue) $\Phi(e^{2\pi i\nu})$ represents $\Phi(U)$.

The following theorem makes the solution of the spectral resolution problem for the (possibly discontinuous) Hermitian operator H depend upon the solution of the spectral resolution problem for the corresponding operator U.

Theorem 6.52 (von Neumann, 1929) Let H be a (closed) Hermitian operator and let U be the Cayley–Moebius transformation of H. Then there is a unique resolution of the identity, $K(\lambda)$, belonging to H (Definition 6.30) if and only if U is unitary.

Proof. (1) We observe that all the eigenvalues of a Hermitian operator are real. Hence a bilinear mapping that transforms the real axis in the λ-plane into the circumference of the unit circle, $e^{2\pi i\nu}$, $-0\cdot5 \leqslant \nu \leqslant 0\cdot5$, would transform Hermitian operators into operators whose eigenvalues all lie on the unit circle (i.e. the isometric/unitary operators). The Cayley–Moebius transformation is such a mapping, i.e.

$$\lambda = \frac{-i(e^{2\pi i\nu}-1)}{e^{2\pi i\nu}+1} = \tan \pi\nu, \quad \nu = \frac{1}{\pi}\tan^{-1}\lambda \tag{6.54}$$

maps the intervals $0<\nu<1$ and $-\infty<\lambda<\infty$ monotonically and one-to-one onto each other. (To see the relationship between the above equation and equations (6.50), (6.51) and (6.52), replace the quasi-eigenvalue λ by H and the quasi-eigenvalue $e^{2\pi i\nu}$ by U.) If we let

$$K(\lambda) = E\left(\frac{1}{\pi}\tan^{-1}\lambda\right), \quad E(\nu) = K(\tan \pi\nu) \tag{6.55}$$

it then follows that $K(\lambda)$ $(-\infty<\lambda<\infty)$ is a resolution of the identity if and only if $E(\nu)$ $(-0{\cdot}5\leqslant\nu\leqslant 0{\cdot}5)$ is a resolution of the identity. Q.E.D. (1).

(2) Now assume that U is a unitary operator with spectral resolution (6.53), and let H be obtained from U by the inverse Cayley–Moebius transformation (6.51) and (6.52). Equation (6.52) gives

$$x = \tfrac{1}{2}(zU+z) = \frac{1}{2}\int_{-0\cdot 5}^{0\cdot 5}(e^{2\pi i\nu}+1)\,dzE(\nu) = \int_{-0\cdot 5}^{0\cdot 5}\cos \pi\nu\, e^{\pi i\nu}\,dzE(\nu)$$

Since $E(\nu)$ is a projection operator with orthogonal increments, we have

$$xE(\nu) = \int_{-0\cdot 5}^{0\cdot 5}\cos \pi\nu'\, e^{\pi i\nu'}\,dzE(\nu')\,E(\nu) = \int_{-0\cdot 5}^{\nu}\cos \pi\nu'\, e^{\pi i\nu'}\,dzE(\nu') \tag{6.56}$$

Equation (6.51) gives

$$xH = \frac{1}{2i}(zU-z) = \frac{1}{2i}\int_{-0\cdot 5}^{0\cdot 5}(e^{2\pi i\nu}-1)\,dzE(\nu) = \int_{-0\cdot 5}^{0\cdot 5}\sin \pi\nu\, e^{\pi i\nu}\,dzE(\nu)$$

which is
$$xH = \int_{-0\cdot 5}^{0\cdot 5}\tan \pi\nu \cos \pi\nu\, e^{\pi i\nu}\,dzE(\nu)$$
$$= \int_{-0\cdot 5}^{0\cdot 5}\tan \pi\nu\, d\int_{-0\cdot 5}^{\nu}\cos \pi\nu'\, e^{\pi i\nu'}\,dzE(\nu')$$

Hence, by equations (6.56), and by equations (6.54), (6.55), we have

$$xH = \int_{-0\cdot 5}^{0\cdot 5}\tan \pi\nu\, dxE(\nu) = \int_{-\infty}^{\infty}\lambda\, dxK(\lambda)$$

which is the spectral representation of H. We see that

$$\|xH\|^2 = \int_{-\infty}^{\infty}\lambda^2\, d\|xK(\lambda)\|^2 = \int_{-0\cdot 5}^{0\cdot 5}\sin^2 \pi\nu\, d\,\|zE(\nu)\|^2$$

is finite because the integral on the right is dominated absolutely by $\|z\|^2$. Hence these manipulations were justified, and so $K(\lambda)$ is the resolution of the identity belonging to H. Q.E.D. (2).

(3) Now assume that $K(\lambda)$ is the resolution of the identity belonging to H (i.e. that H has a unique spectral representation). We write, for arbitrary z,

$$x = \int_{-\infty}^{\infty} \frac{1}{1-i\lambda}\, dzK(\lambda) = \int_{-0\cdot 5}^{0\cdot 5} \frac{1}{1-i\tan \pi\nu}\, dzE(\nu)$$
$$= \int_{-0\cdot 5}^{0\cdot 5} \cos \pi\nu\, e^{\pi i\nu}\, dzE(\nu)$$

These integrals converge because their integrands are bounded. Then we have $xK(\lambda) = xE(\nu)$, and using the representation of x given in the above equation, we obtain the same result as before for $xE(\nu)$, namely equation (6.56). Then from the spectral representation of H, we proceed in the opposite direction as before, again using equation (6.56), to obtain

$$xH = \int_{-\infty}^{\infty} \lambda dxK(\lambda) = \int_{-0\cdot 5}^{0\cdot 5} \tan \pi\nu\, dxE(\nu) = \int_{-0\cdot 5}^{0\cdot 5} \sin \pi\nu\, e^{\pi i\nu}\, dzE(\nu)$$

Therefore

$$x - ixH = \int_{-0\cdot 5}^{0\cdot 5} dzE(\nu) = z, \quad x + ixH = \int_{-0\cdot 5}^{0\cdot 5} e^{2\pi i\nu}\, dzE(\nu) = zU$$

This equation defines zU for all z, because z was arbitrary. If w is arbitrary, then

$$\overline{((zU, w))} = \int_{-0\cdot 5}^{0\cdot 5} e^{-2\pi i\nu}\, d\overline{((zE(\nu), w))} = \int_{-0\cdot 5}^{0\cdot 5} e^{-2\pi i\nu}\, d((wE(\nu), z))$$

and since $\overline{((zU, w))} = ((wU^*, z))$, we see

$$U^* = \int_{-0\cdot 5}^{0\cdot 5} e^{-2\pi i\nu}\, dE(\nu) = U^{-1}$$

so U is unitary and $E(\nu)$ is the resolution of the identity belonging to U. Q.E.D. (3).

Because the Cayley–Moebius transformation U of the Hermitian operator H is unitary if and only if H is hypermaximal, we have

Theorem 6.53 *The von Neumann spectral representation for the hypermaximal Hermitian operator H in Hilbert space.* There is a unique resolution of the identity belonging to the Hermitian operator H (Definition 6.30) if and only if H is hypermaximal.

6.6 Discrete and continuous spectra

Let $K(\lambda)$ be the resolution of the identity belonging to the hypermaximal Hermitian operator H in Hilbert space $\mathbf{V}_\infty$.

Definition 6.60 An *eigenvalue* of a Hermitian operator H in Hilbert space is a real number λ for which the characteristic equation $\phi H = \lambda\phi$ has non-null solutions ϕ in the Hilbert space. Orthonormal ϕ are called the *eigensolutions* corresponding to λ.

Theorem 6.60 The real number λ_0 is an eigenvalue of H if and only if λ_0 is a point of discontinuity of $K(\lambda)$.

Proof ("only if" statement). By hypothesis there is a non-null solution ϕ in $\mathbf{V}_\infty$ of the characteristic equation with λ_0. Hence for y in $\mathbf{V}_\infty$ we have:

$$\begin{aligned} 0 &= \lambda_0((\phi, y)) - ((\phi H, y)) \\ &= \lambda_0 \int_{-\infty}^{\infty} d((\phi K(\lambda), y)) - \int_{-\infty}^{\infty} \lambda d((\phi K(\lambda), y)) \\ &= \int_{-\infty}^{\infty} (\lambda_0 - \lambda)\, d((\phi K(\lambda), y)) \end{aligned}$$

First, if we set $y = \phi K(\lambda_0)$, we have

$$\begin{aligned} 0 &= \int_{-\infty}^{\infty} (\lambda_0 - \lambda)\, d((\phi K(\lambda), \phi K(\lambda_0))) \\ &= \int_{-\infty}^{\infty} (\lambda_0 - \lambda)\, d((\phi K(\lambda) K(\lambda_0), \phi)) \\ &= \int_{-\infty}^{\infty} (\lambda_0 - \lambda)\, d((\phi K(\min\{\lambda, \lambda_0\}), \phi)) \\ &= \int_{-\infty}^{\infty} (\lambda_0 - \lambda)\, d\|\phi K(\min\{\lambda, \lambda_0\})\|^2 \\ &= \int_{-\infty}^{\lambda_0} (\lambda_0 - \lambda)\, d\|\phi K(\lambda)\|^2 + \int_{\lambda_0}^{\infty} (\lambda_0 - \lambda)\, d\|\phi K(\lambda_0)\|^2 \\ &= \int_{-\infty}^{\lambda_0} (\lambda_0 - \lambda)\, d\|\phi K(\lambda)\|^2 \end{aligned} \tag{6.60}$$

since $\|\phi K(\lambda_0)\|^2$ is a constant, so $d\|\phi K(\lambda_0)\|^2 = 0$. Second, if we set $y = \phi$, then

$$\begin{aligned} 0 &= \int_{-\infty}^{\infty} (\lambda_0 - \lambda)\, d((\phi K(\lambda), \phi)) \\ &= \int_{-\infty}^{\infty} (\lambda_0 - \lambda)\, d\|\phi K(\lambda)\|^2 \end{aligned} \tag{6.61}$$

If we subtract equation (6.61) from equation (6.60), we obtain

$$0 = \int_{\lambda_0}^{\infty} (\lambda - \lambda_0)\, d\|\phi K(\lambda)\|^2 \tag{6.62}$$

In both equations (6.60) and (6.62) the integrand is non-negative, and the function behind the differential sign is a monotone, non-decreasing function of λ. Consequently, for an arbitrary positive ϵ, we have

$$0 = \int_{-\infty}^{\lambda_0} (\lambda_0 - \lambda)\, d\|\phi K(\lambda)\|^2 \geqslant \int_{-\infty}^{\lambda_0-\epsilon} (\lambda_0 - \lambda)\, d\|\phi K(\lambda)\|^2$$

$$\geqslant \int_{-\infty}^{\lambda_0-\epsilon} \epsilon d\,\|\phi K(\lambda)\|^2 = \epsilon\|\phi K(\lambda_0 - \epsilon)\|^2$$

and

$$0 = \int_{\lambda_0}^{\infty} (\lambda - \lambda_0)\, d\|\phi K(\lambda)\|^2 \geqslant \int_{\lambda_0+\epsilon}^{\infty} (\lambda - \lambda_0)\, d\|\phi K(\lambda)\|^2$$

$$\geqslant \int_{\lambda_0+\epsilon}^{\infty} \epsilon d\|\phi K(\lambda)\|^2 = \epsilon\|\phi\|^2 - \epsilon\|\phi K(\lambda_0 + \epsilon)\|^2$$

$$= \epsilon\|\phi - \phi K(\lambda_0 + \epsilon)\|^2$$

The expressions on the right-hand sides are clearly non-negative, and yet (because of the left-hand sides) are non-positive, and so are equal to zero. Thus $\phi K(\lambda_0 - \epsilon) = 0$ and $\phi K(\lambda_0 + \epsilon) = \phi$ for arbitrary positive ϵ. We recall that $K(\lambda)$ is continuous on the right. We therefore have the condition

$$\phi K(\lambda) = \begin{cases} 0 & \text{for} \quad \lambda < \lambda_0 \\ \phi & \text{for} \quad \lambda \geqslant \lambda_0 \end{cases} \tag{6.63}$$

which characterizes the discontinuity of $K(\lambda)$ at the point λ_0. Q.E.D. (1).

Proof ("if" statement). By hypothesis $K(\lambda)$ is discontinuous at the point λ_0 (i.e. condition (6.63) is true for some non-null ϕ in $\mathbf{V}_\infty$). Then, clearly, the Stieltjes integral reduces to

$$\int_{-\infty}^{\infty} \lambda d((\phi K(\lambda), y)) = \lambda_0((\phi, y))$$

Thus $((\phi H, y)) = \lambda_0((\phi, y))$, or $((\phi H - \lambda_0\phi, y)) = 0$ so $\phi H = \lambda_0\phi$. Q.E.D. (2).

If $\lambda \to \lambda_0$ (with $\lambda > \lambda_0$), then $K(\lambda) \Rightarrow K(\lambda_0 + 0) = K(\lambda_0)$ because $K(\lambda)$ is continuous on the right. If $\lambda \to \lambda_0$ (with $\lambda < \lambda_0$), then $K(\lambda) \Rightarrow K(\lambda_0 - 0)$. Thus λ_0 is a point of discontinuity if and only if $K(\lambda_0 - 0) \neq K(\lambda_0)$. Moreover, because $K(\lambda) \leqslant K(\lambda_0)$ for $\lambda < \lambda_0$, it follows that $K(\lambda_0 - 0) \leqslant K(\lambda_0)$. Hence $K(\lambda_0) - K(\lambda_0 - 0)$ is a projection operator, and this projection operator is $\neq 0$ if and only if λ_0 is a point of discontinuity of $K(\lambda)$.

Lemma 6.60 Condition (6.63) is true if and only if $\phi K(\lambda_0 - 0) = 0$ and $\phi K(\lambda_0) = \phi$.

Proof. The "only if" statement is obvious. If $\phi K(\lambda_0 - 0) = 0$, then $\phi K(\lambda) = 0$ for $\lambda < \lambda_0$, since $K(\lambda) \leqslant K(\lambda_0 - 0)$ for $\lambda < \lambda_0$. If

$\phi = \phi K(\lambda_0)$, then $\phi K(\lambda) = \phi K(\lambda_0) K(\lambda)$. Because for $\lambda_0 \leqslant \lambda$ we have $K(\lambda_0) \leqslant K(\lambda)$, it follows that $K(\lambda_0) K(\lambda) = K(\lambda_0)$. Hence

$$\phi K(\lambda) = \phi K(\lambda_0) = \phi$$

for $\lambda_0 \leqslant \lambda$. Q.E.D.

Theorem 6.61 The non-null solutions ϕ in $\mathbf{V}_\infty$ corresponding to the eigenvalue λ_0 form a closed linear manifold $\mathbf{M}_{\lambda_0}$, which is the eigenmanifold of the projection operator $K(\lambda_0) - K(\lambda_0 - 0)$.

Proof. By Theorem 6.60 and Lemma 6.60, the condition $\phi K(\lambda_0 - 0) = 0$ and $\phi K(\lambda_0) = \phi$ is true if and only if $\phi H = \lambda_0 \phi$ for non-null ϕ in $\mathbf{V}_\infty$. In other words, the condition

$$\phi\,[K(\lambda_0) - K(\lambda_0 - 0)] = \phi$$

characterizes the solutions ϕ corresponding to λ_0. These solutions form a closed linear manifold, the eigenmanifold of $K(\lambda_0) - K(\lambda_0 - 0)$. Q.E.D.

Theorem 6.62 The eigenmanifolds $\mathbf{M}_{\lambda_0}$ and $\mathbf{M}_{\mu_0}$ of the distinct eigenvalues λ_0 and μ_0 are orthogonal.

Proof. If $\lambda_0 < \mu_0$, then $K(\lambda_0) - K(\lambda_0 - 0) \leqslant K(\lambda_0) \leqslant K(\mu_0 - 0)$ and $K(\mu_0) - K(\mu_0 - 0) \leqslant I - K(\mu_0 - 0)$. Because the eigenmanifolds of $K(\mu_0 - 0)$ and $I - K(\mu_0 - 0)$ are orthogonal, it follows that the eigenmanifolds of $K(\lambda_0) - K(\lambda_0 - 0)$ and $K(\mu_0) - K(\mu_0 - 0)$ are also orthogonal. Q.E.D.

Example 6.60 Suppose that the eigenvalue problem in Hilbert space for H (Section 6.4) has a solution, i.e. there exists a maximal orthonormal set $\phi_1, \phi_2, \ldots$ of eigensolutions corresponding respectively to the (not necessarily distinct) eigenvalues $\lambda_1, \lambda_2, \ldots$. We define

$$K(\lambda) = \sum_{\lambda_i \leqslant \lambda} P_{[\phi_i]}$$

It is not difficult to verify that $K(\lambda)$ is a resolution of the identity. To see that $K(\lambda)$ belongs to H, represent x by its Fourier series $x = a_1\phi_1 + a_2\phi_2 + \ldots$. We see that $xH = \lambda_1 a_1 \phi_1 + \lambda_2 a_2 \phi_2 + \ldots$, which is defined if $\lambda_1^2 |a_1|^2 + \lambda_2^2 |a_2|^2$ is finite. Because

$$xK(\lambda) = \sum_{\lambda_i \leqslant \lambda} xP_{[\phi_i]} = \sum_{\lambda_i \leqslant \lambda} ((x, \phi_i))\phi_i = \sum_{\lambda_i \leqslant \lambda} a_i \phi_i$$

we see that

$$\int_{-\infty}^{\infty} \lambda^2 d\|xK(\lambda)\|^2 = \int_{-\infty}^{\infty} \lambda^2 d\Big(\sum_{\lambda_i \leqslant \lambda} |a_i|^2\Big) = \sum_{i=1}^{\infty} \lambda_i^2 |a_i|^2$$

$$\int_{-\infty}^{\infty} \lambda d((xK(\lambda), y)) = \int_{-\infty}^{\infty} \lambda d\Big(\sum_{\lambda_i \leqslant \lambda} a_i \bar{b}_i\Big) = \sum_{i=1}^{\infty} \lambda_i a_i \bar{b}_i = ((xH, y))$$

so $K(\lambda)$ belongs to H. If $\mathbf{N}_0$ is the closed linear manifold spanned by ϕ_i for $\lambda_i \leqslant \lambda_0$, then $K(\lambda_0)$ is the projection operator with eigenmanifold $\mathbf{N}_0$.

Example 6.61 Returning to Example 6.40, we see that a linear combination of solutions $\phi(u)$ for $\lambda_1, \ldots, \lambda_n$ would be any function that may be non-zero only for the points $\lambda_1, \ldots, \lambda_n$. Hence, heuristically, a linear combination of all solutions $\phi(u)$ for $\lambda \leqslant \lambda_0$ would be any function that may be non-zero only for $\lambda \leqslant \lambda_0$. (For the pure discrete spectrum (Example 6.60), $K(\lambda_0)$ was the projection operator with eigenmanifold $\mathbf{N}_0$, the closed linear manifold spanned by all the eigensolutions ϕ_i corresponding to eigenvalues $\lambda_i \leqslant \lambda_0$. That is, $\mathbf{N}_0$ consisted of all linear combinations of all solutions of the characteristic equation $\phi H = \lambda\phi$ for which $\lambda \leqslant \lambda_0$.) Hence we might now expect that $K(\lambda_0) = P_{\mathbf{N}_0}$, where $\mathbf{N}_0$ now consists of all those functions that may be non-zero only for $u \leqslant \lambda_0$. Then $\mathbf{V}_\infty - \mathbf{N}_0$ consists of all functions that vanish for $u \leqslant \lambda_0$. Hence $x(u)K(\lambda_0) = x(u)$ for $u \leqslant \lambda_0$ and $= 0$ for $u > \lambda_0$. It is readily seen that $K(\lambda)$ is a resolution of the identity. To show that $K(\lambda)$ belongs to H, we see that

$$\int_{-\infty}^{\infty} \lambda^2 d\|xK(\lambda)\|^2 = \int_{-\infty}^{\infty} \lambda^2 d\left[\int_{-\infty}^{\lambda} |x(u)|^2 du\right]$$

$$= \int_{-\infty}^{\infty} \lambda^2 |x(\lambda)|^2 d\lambda = \int_{-\infty}^{\infty} |ux(u)|^2 du = \|xH\|^2 \qquad (6.64)$$

is finite for x in the domain of H (i.e. for those $x(u)$ for which $|ux(u)|^2$ is integrable). The spectral representation is then

$$\int_{-\infty}^{\infty} \lambda d((xK(\lambda), y)) = \int_{-\infty}^{\infty} \lambda d\left[\int_{-\infty}^{\lambda} x(u)\overline{y(u)}\, du\right] = ((xH, y)) \qquad (6.65)$$

Example 6.62 Let the Hilbert space be the $\mathbf{L}^2$ space (with Lebesgue measure) of Examples 6.40, 6.41, 6.61. If $y(v)$ is a function of $\mathbf{L}^2$, then $y(v)$ has the Fourier transform

$$y(v)L = x(u) = \frac{1}{\sqrt{h}} \int_{-\infty}^{\infty} e^{2\pi i u v/h} y(v)\, dv \qquad (6.66)$$

where h is Planck's constant. Plancherel's theorem states that

$$\int_{-\infty}^{\infty} |y(v)|^2 dv = \int_{-\infty}^{\infty} |x(u)|^2 du$$

which is $\|y\| = \|yL\|$, so that $x(u)$ is also a function of $\mathbf{L}^2$, and the Fourier transform L is an isometric operator. Furthermore, from the theory of the Fourier transform, it is true that

$$\frac{1}{\sqrt{h}} \int_{-\infty}^{\infty} e^{2\pi i u v/h} x(u)\, du = y(-v)$$

which is $y(v)LL = y(-v)$. Hence the equation $y(v)L = x(u)$ has the solution $y(v) = x(-u)L$ for each $x(u)$ of $\mathbf{L}^2$, i.e. L maps $\mathbf{L}^2$ onto itself. Consequently, the Fourier transform L is a unitary operator, and $x(u)L^{-1} = x(u)L^* = x(-u)L$.

Example 6.63 Continuing Example 6.41, let us now designate the Hermitian operator defined by equation (6.40) by H' (instead of by H, as is done in equation (6.40)). Proceeding in a heuristic way, the functions

$$x(u) = \int_{-\infty}^{\lambda_0} c(\lambda)\, e^{2\pi i\lambda u/h}\, d\lambda \qquad (6.67)$$

represent linear combinations of all solutions given by equation (6.41). Among the functions (6.67) are functions that are in $\mathbf{L}^2$; for example, if we set $c(\lambda) = 1$ for $\lambda_1 \leqslant \lambda \leqslant \lambda_0$ and $= 0$ elsewhere, the resulting $x(u)$ belongs to $\mathbf{L}^2$. Let $\mathbf{N}_0'$ be the set of those $x(u)$ which belong to $\mathbf{L}^2$ and whose inverse Fourier transforms $x(u)L^{-1} = y(v)$ are equal to zero for $v > \lambda_0$. (Comparing equations (6.66) and (6.67), we see that $y(v) = \sqrt{(h)}\, c(v)$.) But such $y(v)$ (i.e. functions of $\mathbf{L}^2$ that vanish for $v > \lambda_0$) as we have seen in Example 6.61, form the closed linear manifold $\mathbf{N}_0$. Hence $\mathbf{N}_0'$, which is the image of $\mathbf{N}_0$ under the Fourier transform L (i.e. if $y(v)$ belongs to $\mathbf{N}_0$, then $y(v)L = x(u)$ belongs to $\mathbf{N}_0'$), is also a closed linear manifold. As we have seen in Example 6.61, $K(\lambda_0)$ is the projection operator with eigenmanifold $\mathbf{N}_0$. We shall form the projection operator $K'(\lambda_0)$ from $K(\lambda_0)$ as we formed $\mathbf{N}_0'$ from $\mathbf{N}_0$ (i.e. by transforming the $\mathbf{L}^2$ space by L). That is,

$$x(u)K'(\lambda_0) = y(v)K(\lambda_0)L = x(u)L^{-1}K(\lambda_0)L$$

so $K'(\lambda_0) = L^{-1}K(\lambda_0)L$. Now we claim that $K'(\lambda_0)$, which is a projection operator with eigenmanifold $\mathbf{N}_0'$, is the resolution of the identity belonging to H'. From Examples 6.40 and 6.61, we know that $K(\lambda)$ is the resolution of the identity belonging to the operator H (i.e. the operator that multiplies the function $y(v)$ by v). Because $K'(\lambda) = L^{-1}K(\lambda)L$ it is readily verified that $K'(\lambda)$ is a resolution of the identity. If $x(u) = y(v)L$, then

$$\begin{aligned} x(u)H' &= \frac{h}{2\pi i}\frac{\partial}{\partial u}\left(\frac{1}{\sqrt{h}}\int_{-\infty}^{\infty} e^{2\pi i u v/h}\, y(v)\, dv\right) \\ &= \frac{\sqrt{h}}{2\pi i}\int_{-\infty}^{\infty} y(v)\frac{\partial}{\partial u}(e^{2\pi i u v/h})\, dv \\ &= \frac{1}{\sqrt{h}}\int_{-\infty}^{\infty} y(v)\, v\, e^{2\pi i u v/h}\, dv \\ &= [y(v)\, v]L = x(u)L^{-1}HL \end{aligned}$$

so $H' = L^{-1}HL$. Now, from Example 6.61, H has the spectral representation (6.65) for all x for which equation (6.64) is finite, and these equations are still valid after transformation of $\mathbf{L}^2$ by L. Hence they hold for $H' = L^{-1}HL$ and $K'(\lambda) = L^{-1}K(\lambda)L$. That is, if $x(u)$ is differentiable and

$$\int_{-\infty}^{\infty} \left| \frac{h}{2\pi i} x'(u) \right|^2 du$$

is finite, then

$$\int_{-\infty}^{\infty} \lambda^2 \, d\| xK'(\lambda)\|^2$$

is finite, and H' has the spectral representation

$$((xH', y)) = \int_{-\infty}^{\infty} \lambda d((xK'(\lambda), y))$$

Hence $K'(\lambda)$ belongs to H'.

CHAPTER VII

STATIONARY PROCESSES

7.1. Stationary processes and quantum mechanics

Definition 7.10 A sequence $x(t)$, $-\infty<t<\infty$, in Hilbert space $\mathbf{V}_\infty$ is called a *stationary process* if the inner product

$$R(s) = ((x(t+s), x(t)))$$

does not depend on t. The set of sequences

$$x_1(t), \ldots, x_n(t), \quad -\infty<t<\infty$$

is called a (*multiple*) *stationary process* if the inner products

$$R_{ij}(s) = ((x_i(t+s), x_j(t))) \quad (i,j = 1, \ldots, n)$$

do not depend on t. Both s and t are assumed to be integers. The variable s is called the *lag*. (Note: $R_{ij}(s)$ may also be denoted by $R_{x_i x_j}(s)$.)

If $\mathbf{V}_\infty$ is variate space $\mathbf{L}^2(P)$, then $x_i(t)$ is called a (second-order) *stationary random* (or *stochastic*) *process*, and if t represents time, a (second-order) *stationary time-series*. If the variates of a stationary stochastic process are centred at their means, then $R_{ii}(s)$ is called the *autocovariance*; $R_{ii}(s)/R_{ii}(0)$ the *autocorrelation*; $R_{ij}(s)$ the *cross-covariance*; and $R_{ij}(s)/[R_{ii}(0)\,R_{jj}(0)]^{1/2}$ the *cross-correlation*. We shall assume that all variates are centred, unless otherwise stated.

There is an elegant relationship between stationary processes and unitary operators due to André Kolmogorov.

Theorem 7.10 (Kolmogorov, 1941) Let

$$x_1(t), \ldots, x_n(t), \quad -\infty<t<\infty$$

be a multiple stationary process in Hilbert space, and let $\mathbf{X}_{12\ldots n}$ be the closed linear manifold spanned by all the elements of the stationary process. Then the equation

$$x_j(t)\,U = x_j(t+1), \quad (j = 1, \ldots, n), \quad -\infty<t<\infty$$

uniquely determines the unitary operator $U \equiv U_{12\ldots n}$ with domain and range $\mathbf{X}_{12\ldots n}$.

This theorem shows that each stationary process is characterized by a uniquely determined unitary operator. As we have seen in Chapter VI, there is a one-to-one correspondence between unitary operators and hypermaximal Hermitian operators, and hence each stationary process is characterized by a uniquely determined hypermaximal Hermitian operator. John von Neumann (1932) has shown that the spectral representation enters so essentially into all quantum mechanical concepts that its existence cannot be dispensed with. Accordingly, in the mathematical foundations of quantum mechanics he admits only hypermaximal Hermitian operators. Thus quantum mechanical concepts are associated with hypermaximal Hermitian operators. Therefore we see that stationary processes and quantum mechanics have a common mathematical framework, namely the hypermaximal Hermitian operators.

7.2 Time-domain and frequency-domain

Let $x(t)$, $-\infty<t<\infty$, be a stationary time-series. The closed linear manifold $\mathbf{X}$ spanned by $x(t)$, $-\infty<t<\infty$, is called the *time-domain* of the time-series. By Kolmogorov's Theorem 7.10, the time-series is characterized by the unitary operator U with domain $\mathbf{X}$ and range $\mathbf{X}$ where $x(t+1)=x(t)U$ for all integers t. Using the von Neumann spectral representation of U (Theorem 6.50), we obtain

$$x(t+1)=x(t)U=\int_{-0\cdot5}^{0\cdot5} e^{2\pi i\nu}\,dx(t)E(\nu)$$

Hence the spectral representation of the time-series is

$$x(t)=\int_{-0\cdot5}^{0\cdot5} e^{2\pi i\nu t}\,dx(0)E(\nu) \tag{7.20}$$

a result given by Kolmogorov (1941). This integral is a stochastic integral (Section 5.5). The stochastic process $f(\nu)$ defined by $f(\nu)=x(0)E(\nu)$, $-0\cdot5\leqslant\nu\leqslant0\cdot5$, has orthogonal increments. By considering the properties of $f(\nu)$ (instead of the Spectral Representation Theorem, 6.50), Cramér (1942) independently arrived at the spectral representation (7.20) (with $x(0)E(\nu)$ replaced by $f(\nu)$).

Define the function $F(\nu)$ to be

$$F(\nu)=\|x(0)E(\nu)\|^2,\qquad -0\cdot5\leqslant\nu\leqslant0\cdot5 \tag{7.21}$$

Because $E(\nu)$ is a resolution of the identity, it follows that

$$F(-0\cdot5)=\|0\|^2=0,\quad F(0\cdot5)=\|x(0)I\|^2=R(0)$$

and $F(\nu)$ is a real, monotone, non-decreasing function of ν that is continuous on the right. Since $F(\nu)$ has the properties of a distribution function (except that $F(0\cdot5)$, although finite, is not necessarily one), the function $F(\nu)$ is called the *spectral distribution function* of the time-series. The variable ν is called the *frequency*.

From the spectral representation (7.20) we may deduce the *Khintchine–Wold Theorem*:

Theorem 7.20 (Khintchine, 1934; Wold, 1938) Let $x(t)$ be a stationary time-series with autocovariance $R(s)$ and spectral distribution function $F(\nu)$. Then $R(s)$ may be determined from $F(\nu)$ by

$$R(s) = \int_{-0\cdot5}^{0\cdot5} e^{2\pi i\nu s}\, dF(\nu) \tag{7.22}$$

and, conversely, $F(\nu)$ may be determined from $R(s)$ by $F(\nu) = W(\nu+0)+C$ where

$$W(\nu) = R(0)\nu - \sum_{s\neq 0} \frac{R(s)}{2\pi i s} e^{-2\pi i\nu s}$$

and the constant C is calculated from $F(-0\cdot5) = 0$.

Proof. From the spectral representation (7.20) we see that

$$R(s) = ((x(s), x(0))) = \left(\left(\int_{-0\cdot5}^{0\cdot5} e^{2\pi i\nu s}\, dx(0)E(\nu), \quad x(0)\right)\right)$$

$$= \int_{-0\cdot5}^{0\cdot5} e^{2\pi i\nu s}\, d((x(0)E(\nu), x(0))) = \int_{-0\cdot5}^{0\cdot5} e^{2\pi i\nu s}\, dF(\nu)$$

Since $F(0\cdot5)-F(-0\cdot5)$ is finite, $F(\nu)$ has the Fourier series expansion $F(\nu) = W(\nu+0)+C$, where $W(\nu)$ and C are determined as above. Thus the spectral distribution function determines the autocorrelation, and conversely. Q.E.D.

The spectral distribution function $F(\nu)$ defines the Lebesgue–Stieltjes measure on the real line segment $-0\cdot5 \leqslant \nu \leqslant 0\cdot5$ (see Example 1.20). Then the function space $\mathbf{L}^2(F)$, for the line segment $-0\cdot5 \leqslant \nu \leqslant 0\cdot5$ with F-measure, is called the *frequency-domain* of the time-series. That is, the frequency-domain $\mathbf{L}^2(F)$ is the space of all complex-valued, measurable functions $\Phi(\nu)$ of the frequency ν for which

$$\int_{-0\cdot5}^{0\cdot5} |\Phi(\nu)|^2\, dF(\nu)$$

is finite.

Theorem 7.21 *The Isomorphism Theorem* (Stone, 1932; Kolmogorov, 1941) Given a stationary time-series

$$x(t), \quad -\infty < t < \infty$$

define a correspondence between an element x of the time-domain $\mathbf{X}$ and an element $\Phi(\nu)$ of the frequency-domain $\mathbf{L}^2(F)$ by

$$x = \int_{-0\cdot5}^{0\cdot5} \Phi(\nu)\, dx(0)E(\nu) \leftrightarrow \Phi(\nu) \tag{7.23}$$

Then under this correspondence the time-domain and the frequency-domain are isomorphic in the sense that:

(1) The correspondence is one-to-one between the whole $\mathbf{X}$ space and the whole $\mathbf{L}^2(F)$ space;
(2) $cx \leftrightarrow c\Phi$ if $x \leftrightarrow \Phi$ (c a complex number);
(3) $x+y \leftrightarrow \Phi+\Psi$ if $x \leftrightarrow \Phi$ and $y \leftrightarrow \Psi$;
(4) $((x,y)) = ((\Phi, \Psi)) = \int_{-0\cdot 5}^{0\cdot 5} \Phi\overline{\Psi}\, dF(\nu)$;
(5) A sequence of elements in $\mathbf{X}$ converges and has the limit

$$x = \int_{-0\cdot 5}^{0\cdot 5} \Phi(\nu)\, dx(0)\, E(\nu)$$

in $\mathbf{X}$ if and only if the corresponding sequence of functions in $\mathbf{L}^2(F)$ converges and has the limit $\Phi(\nu)$ in $\mathbf{L}^2(F)$.

Proof. The set $\mathbf{M}$ of all elements of $\mathbf{X}$ of the form

$$x = \int_{-0\cdot 5}^{0\cdot 5} \Phi(\nu)\, dx(0)\, E(\nu)$$

where Φ belongs to $\mathbf{L}^2(F)$ is a closed linear manifold in $\mathbf{X}$, by Theorem 6.2 of Stone (1932). Since $x(t)$ is given by the spectral representation (7.20) and since the functions $e^{2\pi i\nu t}$, $-\infty < t < \infty$, belong to $\mathbf{L}^2(F)$, we see that $\mathbf{M}$ contains all the $x(t)$, $-\infty < t < \infty$. But $\mathbf{X}$ is the closed linear manifold spanned by all the $x(t)$, $-\infty < t < \infty$, so therefore $\mathbf{M} \supset \mathbf{X}$. But, by the definition of $\mathbf{M}$, $\mathbf{M} \subset \mathbf{X}$. Therefore $\mathbf{M} = \mathbf{X}$. The remainder of the theorem follows directly from Theorem 6.2 of Stone (1932). The verity of statements (2), (3), and (4) may be seen by formal manipulation; for example, in the case of (4), we have:

$$\begin{aligned}
((x,y)) &= \int_{-0\cdot 5}^{0\cdot 5} \Phi(\nu)\, d((x(0)E(\nu), y)) \\
&= \int_{-0\cdot 5}^{0\cdot 5} \Phi(\nu)\, d((x(0), yE(\nu))) \\
&= \int_{-0\cdot 5}^{0\cdot 5} \Phi(\nu)\, d\overline{((yE(\nu), x(0)))} \\
&= \int_{-0\cdot 5}^{0\cdot 5} \Phi(\nu)\, d \int_{-0\cdot 5}^{0\cdot 5} \overline{\Psi(\sigma)}\, d\overline{((x(0)E(\sigma)E(\nu), x(0)))} \\
&= \int_{-0\cdot 5}^{0\cdot 5} \Phi(\nu)\, d \int_{-0\cdot 5}^{\nu} \overline{\Psi(\sigma)}\, d\overline{((x(0)E(\sigma), x(0)))} \\
&= \int_{-0\cdot 5}^{0\cdot 5} \Phi(\nu)\overline{\Psi(\nu)}\, dF(\nu) \qquad \text{Q.E.D.}
\end{aligned}$$

We recall Definition 5.30, which states that two functions of $\mathbf{L}^2(F)$ are considered as identical if and only if they are equal for all sets of frequencies ν that have positive measure. That is, two functions that differ only on a ν-set of F-measure zero are identical in $\mathbf{L}^2(F)$. From the definition of the F-measure (Example 1.20), we see that a ν-set has positive measure if and only if it is a set on which $F(\nu)$ increases. That is, any set of frequencies on which $F(\nu)$ does not increase has zero F-measure, and conversely. For

$$\begin{aligned}\|x(0)[E(\nu)-E(\sigma)]\|^2 &= ((x(0)[E(\nu)-E(\sigma)],\, x(0)[E(\nu)-E(\sigma)]))\\ &= ((x(0)E(\nu),\, x(0)E(\nu))) - ((x(0)E(\nu),\, x(0)E(\sigma)))\\ &\quad - ((x(0)E(\sigma),\, x(0)E(\nu))) + ((x(0)E(\sigma),\, x(0)E(\sigma)))\\ &= \|x(0)E(\nu)\|^2 - \|x(0)E(\sigma)\|^2 = F(\nu)-F(\sigma)\end{aligned}$$

we write symbolically

$$\|dx(0)E(\nu)\|^2 = dF(\nu) \tag{7.24}$$

Definition 7.20 Let $x(t)$ be a stationary time-series with unitary operator U which has the resolution of the identity, $E(\nu)$, $-0{\cdot}5 \leqslant \nu \leqslant 0{\cdot}5$. Those frequencies ν in whose neighbourhoods $E(\nu)$ is not constant form the *spectrum* S of the time-series $x(t)$. Those frequencies σ for which $E(\nu)$ is discontinuous form the *point* (or *discrete*) *spectrum*. The spectrum consists of the point spectrum and the *continuous spectrum*. The set of frequencies that do not belong to the spectrum forms the *resolvent* of the time-series.

From equation (7.24) we see that the spectrum S is the set of those frequencies ν $(-0{\cdot}5 \leqslant \nu \leqslant 0{\cdot}5)$ for which the spectral distribution function $F(\nu)$ increases, and the resolvent is the set of those frequencies for which the spectral distribution function $F(\nu)$ does not increase. Accordingly, two functions $\Phi(\nu)$ and $\Psi(\nu)$, $-0{\cdot}5 \leqslant \nu \leqslant 0{\cdot}5$, that are equal for frequencies $\nu \in S$, but which may not be equal for frequencies ν in the resolvent, are considered as identical functions in $\mathbf{L}^2(F)$. Also, all integrals on ν for $\nu = -0{\cdot}5$ to $\nu = 0{\cdot}5$ may be replaced by integrals on ν for $\nu \in S$. Because it is common practice to use the terms "spectrum" and "spectral distribution function" interchangeably, the reader should note the distinction between these terms in our usage: the spectrum S is a point-set of frequencies ν on the real line segment $-0{\cdot}5 \leqslant \nu \leqslant 0{\cdot}5$, whereas the spectral distribution function $F(\nu)$ is a distribution function on the real line segment $-0{\cdot}5 \leqslant \nu \leqslant 0{\cdot}5$, the set of increase of $F(\nu)$ being S.

The Isomorphism Theorem 7.21 allows us to work with functions $\Phi(\nu)$ defined on the spectrum S instead of the variates x defined on the abstract space Ω. For example, the time-series $x(t)$ at time t corresponds to a complex sinusoidal function; i.e.

$$x(t) \leftrightarrow e^{2\pi i\nu t} = \cos 2\pi\nu t + i \sin 2\pi\nu t, \quad \nu \in S$$

In particular, we see that

$$x(0) \leftrightarrow 1, \quad \nu \in S \tag{7.25}$$

The stochastic process $x(0)E(\sigma)$ (which has orthogonal increments) corresponds to the characteristic function $\gamma_\sigma(\nu)$ of the set $-0{\cdot}5 \leqslant \nu \leqslant \sigma$;

i.e.
$$x(0)E(\sigma) \leftrightarrow \gamma_\sigma(\nu) = \begin{cases} 1, & -0{\cdot}5 \leqslant \nu \leqslant \sigma, \quad \nu \in S \\ 0, & \sigma < \nu \leqslant 0{\cdot}5, \quad \nu \in S \end{cases} \tag{7.26}$$

Moreover, if $x \leftrightarrow \Phi$, then

$$xU^t \leftrightarrow \Phi(\nu)\, e^{2\pi i\nu t}, \quad \nu \in S$$

so that operation on the variate x in the time-domain by the unitary operator U corresponds to multiplication of the function $\Phi(\nu)$ in the frequency-domain by the complex sinusoidal function $e^{2\pi i\nu}$. If $x \leftrightarrow \Phi$, then

$$xE(\sigma) = \int_{-0\cdot5}^{0\cdot5} \Phi(\nu)\, dx(0)E(\nu)E(\sigma) = \int_{-0\cdot5}^{\sigma} \Phi(\nu)\, dx(0)E(\nu)$$

so $xE(\sigma) \leftrightarrow \Phi(\nu)\gamma_\sigma(\nu)$. Because, if $x \leftrightarrow \Phi$ and $y \leftrightarrow \Psi$,

$$((x, y)) = \int_\Omega x(\omega)\overline{y(\omega)}\, P(d\omega) = \int_S \Phi(\nu)\overline{\Psi(\nu)}\, dF(\nu) = ((\Phi, \Psi))$$

we see that

$$((xE(\sigma), y)) = \int_S \Phi\gamma_\sigma\overline{\Psi}\, dF(\nu) = \int_{-0\cdot5}^{\sigma} \Phi(\nu)\overline{\Psi(\nu)}\, dF(\nu) \tag{7.27}$$

7.3 The Gibbs ergodic theorem

The first systematic exposition of the foundations of statistical mechanics and its applications to thermodynamics was given by Willard Gibbs (1902). It is a tribute to the originality of his ideas and the magnitude of his contribution that his ergodic theorem remained unproven for thirty years. B. O. Koopman (1931) perceived the relationship between the ergodic theorem and function space, which led John von Neumann (1932a), using Hilbert function space $\mathbf{L}^2$, to give the first proof of the ergodic theorem. Carleman did the fundamental work for the $\mathbf{L}^2$ version independently, and Birkoff gave a proof for function space $\mathbf{L}^1$. Khintchine (1934) applied the ergodic theorem to stationary processes. For second-order stationary processes, von Neumann's $\mathbf{L}^2$ version, in the framework of Khintchine, is required.

Theorem 7.30 *The Gibbs ergodic theorem* Let $x(t)$ be a stationary time-series with spectral representation (7.20). Then as $m-n \to \infty$, the time average

$$\frac{1}{m-n} \sum_{t=n+1}^{m} x(t)\, e^{-2\pi i\sigma t} \Rightarrow x(0)[E(\sigma) - E(\sigma - 0)] \tag{7.30}$$

Proof. Since $cx \leftrightarrow c\Phi$ if $x \leftrightarrow \Phi$, we have

$$x(t)\,\mathrm{e}^{-2\pi i\sigma t} = \int_{-0\cdot5}^{0\cdot5} \mathrm{e}^{2\pi i\,(\nu-\sigma)\,t}\,dx(0)\,E(\nu)$$

Hence

$$\frac{1}{m-n}\sum_{n+1}^{m} x(t)\,\mathrm{e}^{-2\pi i\sigma t} = \int_{-0\cdot5}^{0\cdot5}\left[\frac{1}{m-n}\sum_{n+1}^{m}\mathrm{e}^{2\pi i\,(\nu-\sigma)\,t}\right]dx(0)\,E(\nu)$$

Consider the integrand in the above integral (i.e. the expression in the brackets). Clearly, when $\nu = \sigma$, the integrand is equal to $(1+1+\ldots+1)/(m-n) = 1$, whatever be the value of $m-n$. That is, when $\nu = \sigma$, the components of the complex sinusoidal functions $\cos 2\pi(\nu-\sigma)\,t + i\sin 2\pi(\nu-\sigma)\,t$ are exactly in phase, so their sum (for $t = n+1$ to m) divided by $m-n$ is precisely equal to 1. On the other hand, when $\nu \neq \sigma$, the components of the complex sinusoidal functions are out of phase, and hence their sum (for $t = n+1$ to m) suffers interference effects. Consequently, their sum divided by $m-n$ tends to zero as $m-n$ tends to infinity. Thus when $m-n \to \infty$, the integrand tends to zero for all ν except for $\nu = \sigma$, where the integrand is 1, and equation (7.30) results. Q.E.D.

Definition 7.30 An *eigenvalue* of a unitary operator U in Hilbert space is a number $\mathrm{e}^{2\pi i\sigma}\,(-0\cdot5 \leqslant \sigma \leqslant 0\cdot5)$ for which the characteristic equation

$$\phi U = \mathrm{e}^{2\pi i\sigma}\,\phi \tag{7.31}$$

has non-null solutions ϕ in the Hilbert space. Orthonormal ϕ are called *eigensolutions.*

Theorem 7.31 Let $x(t)$ be a stationary time-series with unitary operator U. Then the time average

$$\frac{1}{m-n}\sum_{t=n+1}^{m} x(t)\,\mathrm{e}^{-2\pi i\sigma t} \Rightarrow 0$$

as $m-n \to \infty$ if and only if $\mathrm{e}^{2\pi i\sigma}$ is not an eigenvalue of U.

Proof. The unitary-operator counterpart of Theorem 6.60 states that $\mathrm{e}^{2\pi i\sigma}$ is an eigenvalue of U if and only if σ is a point of discontinuity of the resolution of the identity $E(\nu)$ belonging to U; i.e. if and only if $E(\sigma)-E(\sigma-0) \neq 0$. Application of Theorem 7.30 then gives the desired result. Q.E.D.

Theorem 7.32 Let $x(t)$ be a stationary time-series with unitary operator U. If $\mathrm{e}^{2\pi i\sigma}$ is an eigenvalue of U, then the non-null solutions ϕ (in $\mathbf{X}$) of the characteristic equation (7.31) form a closed linear manifold, which is the eigenmanifold of the projection operator $E(\sigma)-E(\sigma-0)$.

Theorem 7.32 is the unitary-operator counterpart of Theorem 6.61. We see that $x(0)\,[E(\sigma)-E(\sigma-0)]$ of expression (7.30) is the projection of $x(0)$ on the eigenmanifold of $e^{2\pi i\sigma}$. Corresponding to Theorem 6.62, eigenmanifolds of distinct eigenvalues of U are orthogonal.

Theorem 7.33 Let $x(t)$ be a stationary time-series with autocovariance $R(t)$ and spectral distribution function $F(\nu)$. Then, as $m-n\to\infty$, the time-average

$$\frac{1}{m-n}\sum_{t=n+1}^{m} R(t)\,e^{-2\pi i\sigma t}\to F(\sigma)-F(\sigma-0) \tag{7.32}$$

Proof. By using equation (7.22) the proof is similar to the proof of Theorem 7.30, except that we now have point-wise convergence ($\to$) instead of metric space convergence ($\Rightarrow$). Since

$$\|x(0)\,[E(\sigma)-E(\sigma-0)]\|^2 = F(\sigma)-F(\sigma-0) \tag{7.33}$$

it follows that the right-hand side of expression (7.32) is non-zero if and only if $e^{2\pi i\sigma}$ is an eigenvalue. Q.E.D.

In practice the autocorrelation or the autocovariance is estimated by time-averages over a *fixed realization* (i.e. a given observation) of a stationary time-series. We recall that a variate is a numerical-valued function of the points ω of Ω, and so, more specifically, we should write $x(t;\omega)$ instead of $x(t)$ in order to show the dependence of the variate $x(t)$ on the point ω. A fixed realization, then, is $x(t;\omega_0)$, $-\infty<t<\infty$, where ω_0 is a single fixed point. Thus we may think of each point ω as a (two-sided) sequence

$$\omega = (\ldots, u_{-1}, u_0, u_1, u_2, \ldots)$$

i.e. each ω represents a fixed realization u_t, $-\infty<t<\infty$ (where u_t are complex numbers). Then the random variable $x(t;\omega)$ is the *co-ordinate function* that assigns to ω the tth co-ordinate of ω, i.e. $x(t;\omega)=u_t$.

By inspection we see that the time-average given in expression (7.34) below is unbiased (i.e. has mean equal to $R(u)$). The following theorem gives the condition for the time-average to be consistent (i.e. to be asymptotically equal to $R(u)$).

Theorem 7.34 (Wiener, 1942; Doob, 1953) Let $x(t)$, $-\infty<t<\infty$, be a stationary time-series in $\mathbf{L}^2(P)$ with autocovariance $R(s)$. Define $y(t)=x(t+u)\overline{x(t)}$, $-\infty<t<\infty$, where u is a fixed integer. (Note: The variates $y(t)$ will not in general be centred at their means.) If $y(t)$ is a stationary time-series in $\mathbf{L}^2(P)$, i.e. if the inner products

$$S(s) = ((x(t+u+s)\overline{x(t+s)}, x(t+u)\overline{x(t)}))$$

do not depend on t, then the time-average

$$\frac{1}{m-n}\sum_{t=n+1}^{m} x(t+u)\overline{x(t)}\Rightarrow R(u)+0 \tag{7.34}$$

(where 0 is the null variate) as $m-n \to \infty$ if and only if

$$\frac{1}{m-n} \sum_{s=n+1}^{m} S(s) \to |R(u)|^2$$

as $m-n \to \infty$.

Proof. We have

$$E[y(t)] = ((y(t), 1)) = \int_{\Omega} x(t+u)\overline{x(t)}\, P(d\omega) = R(u)$$

Hence the time-series $y(t) - R(u)$ is centred (i.e. has zero mean) and so application of Theorem 7.30 (with $\sigma = 0$) shows that, as $m-n \to \infty$, the time-average

$$\frac{1}{m-n} \sum_{t=n+1}^{m} [y(t) - R(u)] = \frac{1}{m-n} \sum_{t=n+1}^{m} x(t+u)\overline{x(t)} - R(u)$$

converges ($\Rightarrow$) to a limit variate in $\mathbf{L}^2(P)$. By Theorem 7.33 (equations (7.32) and (7.33)) this limit is the null variate of $\mathbf{L}^2(P)$ if and only if

$$\frac{1}{m-n} \sum_{s=n+1}^{m} ((y(t+s) - R(u), y(t) - R(u)))$$

$$= \frac{1}{m-n} \sum_{s=n+1}^{m} [((y(t+s), y(t))) - \overline{R(u)}((y(t+s), 1)) \\ - R(u)((1, y(t))) + |R(u)|^2]$$

$$= \frac{1}{m-n} \sum_{s=n+1}^{m} S(s) - |R(u)|^2$$

converges ($\to$) to zero as $m-n \to \infty$. We recall that $[R(u) + \text{null variate}]$ is a function equal to $R(u)$ with probability 1 (i.e. equal to $R(u)$ except on an ω-set of probability zero). Q.E.D.

7.4 Multiple stationary time-series

Let $x_1(t)$ and $x_2(t)$, $-\infty < t < \infty$, be a multiple stationary time-series. Then, by Theorem 7.10,

$$x_j(t) = \int_{-0\cdot5}^{0\cdot5} e^{2\pi i\nu t}\, dx_j(0)\, E(\nu) \quad (j = 1, 2) \tag{7.40}$$

where $E \equiv E_{12}$ is the resolution of the identity belonging to the unitary operator $U \equiv U_{12}$. The *cross-spectral function* is defined to be

$$F_{12}(\nu) = ((x_1(0)\, E(\nu), x_2(0))), \quad -0\cdot5 \leqslant \nu \leqslant 0\cdot5 \tag{7.41}$$

It follows that $F_{12}(\nu)$ is a function continuous on the right and has bounded variation. From $E(-0\cdot5) = 0$ we see that $F_{12}(-0\cdot5) = 0$. The counterpart of Theorem 7.20 is that

$$R_{12}(s) = \int_{-0\cdot5}^{0\cdot5} e^{2\pi i\nu s}\, dF_{12}(\nu)$$

and, conversely, $F_{12}(\nu) = W_{12}(\nu) + C$ where

$$W_{12}(\nu) = R_{12}(s)\nu - \sum_{s \neq 0} \frac{R_{12}(s)}{2\pi i s} e^{-2\pi i \nu s}$$

and C is calculated from $F_{12}(-0{\cdot}5) = 0$. We see that

$$R_{12}(s) = \bar{R}_{21}(-s), \quad F_{12}(\nu) = \bar{F}_{21}(\nu)$$

Moreover, it is true that the increments $\Delta F_{ij} = F_{ij}(\sigma) - F_{ij}(\nu), \sigma > \nu$, satisfy $|\Delta F_{12}|^2 \leqslant \Delta F_{11} \Delta F_{22}$.

Definition 7.40 A stationary time-series $x_2(t)$, $-\infty < t < \infty$, is called *subordinated* to a stationary time-series $x_1(t)$, $-\infty < t < \infty$, if $x_1(t)$ and $x_2(t)$ form a multiple stationary time-series and if $x_2(t)$, $-\infty < t < \infty$, belongs to the time-domain $\mathbf{X}_1$ of $x_1(t)$.

If $x_2(t)$ is subordinated to $x_1(t)$, then $\mathbf{X}_{12} = \mathbf{X}_1$, $U_{12} = U_1$, and $E_{12} = E_1$.

Theorem 7.40 (Kolmogorov, 1941) For each stationary time-series $x_2(t)$ subordinated to $x_1(t)$ there corresponds one and only one function $\Phi_2^{(1)}(\nu)$ in the frequency-domain $\mathbf{L}^2(F_{11})$ for which

$$F_{22}(\nu) = \int_{-0\cdot5}^{\nu} |\Phi_2^{(1)}(\nu)|^2 \, dF_{11}(\nu) \tag{7.42}$$

$$F_{21}(\nu) = \int_{-0\cdot5}^{\nu} \Phi_2^{(1)}(\nu) \, dF_{11}(\nu) \tag{7.43}$$

The correspondence $x_2(0) \leftrightarrow \Phi_2^{(1)}(\nu)$ is one-to-one between the class of all stationary time-series $x_2(t)$ subordinated to $x_1(t)$ and the frequency-domain $\mathbf{L}^2(F_{11})$ of $x_1(t)$. For any $x_2(t)$ and $x_3(t)$ in the class of stationary time-series subordinated to $x_1(t)$ we have

$$F_{23}(\nu) = \int_{-0\cdot5}^{\nu} \Phi_2^{(1)} \overline{\Phi_3^{(1)}} \, dF_{11}(\nu) \tag{7.44}$$

Proof. If $x_2(t)$ is a stationary time-series subordinated to $x_1(t)$, then $x_2(0)$ belongs to $\mathbf{X}_1$. According to the Isomorphism Theorem 7.21, there is a function $\Phi_2^{(1)}$ of $\mathbf{L}^2(F_{11})$ for which $y(0) \leftrightarrow \Phi_2^{(1)}$. Hence, by (7.41) we have

$$F_{23}(\nu) = ((x_2(0)E_{23}(\nu), x_3(0))) = ((x_2(0)E_1(\nu), x_3(0)))$$

which, by equation (7.27), gives equation (7.44). In the particular case when $x_3(t) = x_1(t)$, then $x_3(0) = x_1(0) \leftrightarrow 1 = \Phi_3^{(1)}$ by equation (7.25), and hence (7.43) results from (7.44). In the particular case when $x_3(t) = x_2(t)$, then $x_3(0) = x_2(0) \leftrightarrow \Phi_2^{(1)}$ and hence equation (7.42) results from (7.44). That $x_2(t)$ determines $\Phi_2^{(1)}$ uniquely follows from equation (7.43), by which

$$\Phi_2^{(1)}(\nu) = \frac{dF_{21}(\nu)}{dF_{11}(\nu)} \qquad \text{Q.E.D.}$$

Theorem 7.41 (Kolmogorov, 1941) Let $x_1(t)$ and $x_2(t)$ be a multiple stationary time-series. Then $x_2(t)$ is subordinated to $x_1(t)$ if and only if there is a function Φ in $\mathbf{L}^2(F_{11})$ for which

$$F_{22}(\nu) = \int_{-0.5}^{\nu} |\Phi(\nu)|^2 \, dF_{11}(\nu) \tag{7.45}$$

$$F_{21}(\nu) = \int_{-0.5}^{\nu} \Phi(\nu) \, dF_{11}(\nu) \tag{7.46}$$

Proof. The "only if" statement follows from Theorem 7.40. To establish the "if" statement, construct the time-series $x_3(t)$ by the correspondence $x_3(0) \leftrightarrow \Phi$. Since, by hypothesis, Φ is in $\mathbf{L}^2(F_{11})$ and equations (7.45) and (7.46) are satisfied, the one-to-one correspondence of Theorem 7.40 shows that $x_3(t)$ is subordinated to $x_1(t)$. Then it can be shown that $x_3(t)$ coincides with $x_2(t)$. Q.E.D.

Definition 7.41 Two stationary time-series $x_1(t)$ and $x_2(t)$ are called *equivalent* if they are mutually subordinated, i.e. if $x_1(t)$ and $x_2(t)$ form a multiple stationary time-series and $\mathbf{X}_1 = \mathbf{X}_2$.

Theorem 7.42 The stationary time-series $x_2(t)$, subordinated to $x_1(t)$, is equivalent to $x_1(t)$ if and only if $\Phi_2^{(1)} \neq 0$ for ν in the spectrum S_1 of $x_1(t)$ (i.e. if and only if $\Phi_2^{(1)} = 0$ only on a ν-set of F_{11}-measure zero).

Proof. The proof can be established from the following observation: If $x_2(t)$ is equivalent to $x_1(t)$ then by equation (7.42) F_{22}-measure is absolutely continuous with respect to F_{11}-measure, and (by interchanging indices) the converse is also true. Hence the spectra S_1 and S_2 of equivalent time-series $x_1(t)$ and $x_2(t)$ coincide, and

$$\Phi_2^{(1)}(\nu) = \frac{1}{\Phi_1^{(2)}(\nu)}, \quad \nu \epsilon S_1 = S_2 \tag{7.47}$$

Q.E.D.

7.5 The spectral decomposition

Let S be the spectrum, and $F(\nu)$ the spectral distribution function, of the stationary time-series $x(t)$. Let S be divided into two non-overlapping sets S_1 and S_2 (i.e. $S = S_1 \cup S_2$, $S_1 \cap S_2 =$ null set) each of which is measurable with respect to F-measure. If the spectral representation of $x(t)$ is given by equation (7.20), define $x_j(t)$ $(j = 1, 2)$ to be

$$x_j(t) = \int_{S_j} e^{2\pi i\nu t} \, dx(0) E(\nu) = \int_S e^{2\pi i\nu t} \Lambda_j(\nu) \, dx(0) E(\nu) \tag{7.50}$$

where $\Lambda_j(\nu)$ is the characteristic function of the set S_j (i.e. $\Lambda_j(\nu) = 1$ for ν in S_j, and $= 0$ for ν not in S_j). We see that $x_j(t) \leftrightarrow e^{2\pi i\nu t} \Lambda_j(\nu)$ in the isomorphism of $\mathbf{X}$ and $\mathbf{L}^2(F)$ (Theorem 7.21). We note that

$\Lambda_1(\nu)+\Lambda_2(\nu) = 1$ for ν in S, and $\Lambda_1(\nu)\Lambda_2(\nu) = 0$ for ν in S. By Theorem 7.21 we have

$$((x_j(t+s), x_k(t))) = \int_S e^{2\pi i\nu s}\Lambda_j(\nu)\Lambda_k(\nu)\,dF(\nu)$$

which $= 0$ for $j \neq k$ and which does not depend on t for $j = k$. Hence $x_1(t)$ and $x_2(t)$ form a multiple stationary time-series, and $x_1(t)$ and $x_2(s)$ are orthogonal for all t and s. The spectrum of $x_1(t)$ is S_1 and its spectral distribution function is

$$F_{11}(\nu) = \int_{-0\cdot5}^{\nu} \Lambda_1(\nu)\,dF(\nu) \tag{7.51}$$

and similarly for $F_{22}(\nu)$. The cross-spectral function $F_{12}(\nu)$ is equal to zero. Thus the decomposition

$$x(t) = \int_{S_1} e^{2\pi i\nu t}\,dx(0)E(\nu) + \int_{S_2} e^{2\pi i\nu t}\,dx(0)E(\nu) = x_1(t)+x_2(t)$$

divides $x(t)$ into two orthogonal time-series, where

$$F(\nu) = F_1(\nu)+F_2(\nu)$$

If $F_{1x}(\nu)$ is the cross-spectral function of $x_1(t)$ and $x(t)$, then from the decomposition we see that $F_{1x}(\nu) = F_{11}(\nu)$. Hence from equations (7.43) and (7.51) we have

$$\Phi_1^{(x)}(\nu) = \frac{dF_{1x}(\nu)}{dF(\nu)} = \frac{dF_{11}(\nu)}{dF(\nu)} = \Lambda_1(\nu)$$

and similarly for $\Phi_2^{(x)}(\nu)$. Corresponding orthogonal decompositions can be made by dividing S into any finite or countable number of non-overlapping sets $S_1, S_2, S_3, \ldots$.

If σ is in the *point spectrum* of $x(t)$ then $e^{2\pi i\sigma}$ is an eigenvalue of U, and conversely. For each eigenvalue of U, choose one and only one eigensolution. In this way we obtain an orthonormal set in $\mathbf{X}$. By Theorem 4.70 this set must be either finite or countable. Since, by construction, there is a one-to-one correspondence between the eigensolutions in this set and the eigenvalues of U, the eigenvalues must be either finite or countable. Hence the point spectrum of a time-series must be either a finite or countable set, and so has Lebesgue measure zero.

Suppose that the *continuous spectrum* of $x(t)$ exists, and let the interval $\nu_1 \leqslant \nu \leqslant \nu_2$ be in the continuous spectrum of $x(t)$. That is, in this interval $E(\nu)$ increases continuously, so $E(\nu_1) < E(\nu_2)$. For $\sigma \leqslant \nu_1$, we have $E(\sigma)-E(\sigma-0) \leqslant E(\sigma) \leqslant E(\nu_1)$; for $\nu_1 < \sigma \leqslant \nu_2$,

$$E(\sigma)-E(\sigma-0) = 0$$

by the continuity; and for $\nu_2 < \sigma$,

$$E(\sigma)-E(\sigma-0) \leqslant I-E(\sigma-0) \leqslant I-E(\nu_2)$$

Hence $E(\sigma)-E(\sigma-0)$ is always orthogonal to $E(\nu_2)-E(\nu_1)$. Let $\mathbf{N}$ be the eigenmanifold of the projection operator $E(\nu_2)-E(\nu_1)$. Thus the eigenmanifolds of all the eigenvalues are orthogonal to $\mathbf{N}$. Now $\mathbf{N}$ contains elements other than the null element (because, by hypothesis, $E(\nu_2)-E(\nu_1)\neq 0$). Hence the eigensolutions (which, by definition, belong to the eigenmanifolds of the eigenvalues) cannot form a maximal orthonormal set in $\mathbf{X}$. In other words, if the continuous spectrum of $x(t)$ exists, then $x(t)$ cannot be expressed by a Fourier series in the eigensolutions.

Let $\sigma_1, \sigma_2, \ldots$ be the point spectrum of $x(t)$. Then, by definition, the eigenmanifolds of the eigenvalues are the eigenmanifolds of the projection operators $E(\sigma_1)-E(\sigma_1-0)$, $E(\sigma_2)-E(\sigma_2-0)$, ... Let $\mathbf{M}$ be the closed linear manifold spanned by these eigenmanifolds. Alternatively, $\mathbf{M}$ is the closed linear manifold spanned by all the eigensolutions. The projection of $x(t)$ on $\mathbf{M}$, which we call $x_1(t)$, is a stationary time-series with time-domain $\mathbf{M}$, and has the spectral representation

$$x_1(t) = x(t)P_{\mathbf{M}} = \sum_n e^{2\pi i\sigma_n t}x(0)[E(\sigma_n)-E(\sigma_n-0)]$$

Because of equation (7.33), the spectral distribution function $F(\nu)$ of $x(t)$ has a jump (i.e. $F(\nu)-F(\nu-0)>0$) at and only at $\nu=\sigma_n$, i.e. a frequency in the point spectrum.

The normal $x(t)[I-P_{\mathbf{M}}]$, which we call $x_2(t)+x_3(t)$, is a stationary time-series with time-domain $\mathbf{X}-\mathbf{M}$, which is the eigenmanifold of the projection operators corresponding to the continuous spectrum of $x(t)$. Because of equation (7.33) the spectral distribution function $F(\nu)$ is continuous if and only if ν is a point of the continuous spectrum of $x(t)$.

The continuous spectrum of $x(t)$ may be decomposed into a set S_3 of frequencies where $F(\nu)$ is absolutely continuous with respect to Lebesgue measure $\mu(d\nu) = d\nu$, and into a set S_2 of frequencies where $F(\nu)$ is not absolutely continuous. Because $F(\nu)$ is a distribution function (i.e. $F(\nu)$ has bounded variation: $F(0{\cdot}5)-F(-0{\cdot}5)<\infty$), the set S_2 necessarily has Lebesgue measure zero. If we let S_1 denote the point spectrum $\sigma_1, \sigma_2, \ldots$, then we have decomposed the spectrum S of $x(t)$ into three non-overlapping sets: S_1, the point spectrum; S_2, the continuous but non-absolutely continuous spectrum; and S_3, the absolutely-continuous spectrum. That is, $S = S_1\cup S_2\cup S_3$, where $S_1\cap S_2\cap S_3 =$ null set. Both S_1 and S_2 have Lebesgue measure zero. Hence the following theorem may be readily established:

Theorem 7.50 *The Spectral Decomposition* (Cramér, 1942) Let $x(t)$, $-\infty<t<\infty$, be a stationary time-series with spectrum S and spectral distribution function $F(\nu)$. Then $x(t)$ has the unique decomposition

$$x(t) = x_1(t)+x_2(t)+x_3(t)$$

where the components $x_j(t)$ $(j = 1, 2, 3)$ form a multiple stationary time-series with the properties:

(1) The $x_j(t)$ are mutually orthogonal; i.e. $((x_j(t), x_k(s))) = 0$ for $j \neq k$.

(2) The spectrum S_1 of $x_1(t)$ is the point spectrum of $x(t)$; the spectrum S_2 of $x_2(t)$ is the continuous, non-absolutely continuous spectrum of $x(t)$; and the spectrum S_3 of $x_3(t)$ is the absolutely-continuous spectrum of $x(t)$.

(3) $S_1 \cup S_2 \cup S_3 = S$; $S_1 \cap S_2 \cap S_3 =$ null set.

(4) If the spectral representation of $x_j(t)$ is

$$x_j(t) = \int_{-0\cdot 5}^{0\cdot 5} e^{2\pi i\nu t}\, dx_j(0)\, E_j(\nu)$$

then $dx_j(0)\,E_j(\nu) = dx(0)\,E(\nu)$ for $\nu \epsilon S_j$ and $= 0$ otherwise.

(5) The spectral distribution functions satisfy

$$F(\nu) = F_{11}(\nu) + F_{22}(\nu) + F_{33}(\nu)$$

$F_{11}(\nu)$ is a step-function and $F_{22}(\nu)$ and $F_{33}(\nu)$ are continuous functions. The derivative of $F_{11}(\nu)$ does not exist for ν in S_1 (but $F_{11}(\nu) - F_{11}(\nu - 0) = F(\nu) - F(\nu - 0)$ for ν in S_1), and $F'_{11}(\nu) = 0$ for all ν, $-0\cdot 5 \leqslant \nu \leqslant 0\cdot 5$, except for the set S_1 of Lebesgue measure zero. The derivative $F'_{22}(\nu)$ does not exist for ν in S_2, but $F'_{22}(\nu) = 0$ for all ν except for the set S_2 of Lebesgue measure zero. The derivative $F'_{33}(\nu)$ exists for all ν, and $F'_{33}(\nu) = F'(\nu) > 0$ for ν in S_3 and $F'_{33}(\nu) = 0$ elsewhere.

(6) Any of the components $x_1(t)$, $x_2(t)$, and $x_3(t)$ may be the null variate. If $x_3(t)$ is non-null then the spectrum S_3 has positive Lebesgue measure. The spectra S_1 and S_2 always have zero Lebesgue measure.

The derivative $F'_{33}(\nu)$ is called the *spectral power function*, or *spectral density function* of $x(t)$. To verify property (6) of Theorem 7.50 we note that

$$F_{33}(0\cdot 5) = \int_{S_3} F'_{33}(\nu)\, d\nu$$

will equal zero unless S_3 has positive Lebesgue measure. We also note that the component $x_j(t)$ may be constructed from $F_{jj}(t)$ by use of Theorem 7.21; i.e. form the frequency domains $\mathbf{L}^2(F_{jj})$ and then let $x_j(t) \leftrightarrow e^{2\pi i\nu t}$ in $\mathbf{L}^2(F_{jj})$. The time-domains $\mathbf{X}_j$ $(j = 1, 2, 3)$ are mutually orthogonal and satisfy $\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2 + \mathbf{X}_3$; and the frequency-domains $\mathbf{L}^2(F_{jj})$ are mutually orthogonal and satisfy

$$\mathbf{L}(F) = \mathbf{L}(F_{11}) + \mathbf{L}(F_{22}) + \mathbf{L}(F_{33})$$

7.6 White noise

Definition 7.60 A time-series $x(t)$ is said to be a *running average* of a stationary time-series $\theta(t)$ if it is possible to choose coefficients a_s such that, as $m-n\to\infty$,

$$\sum_{s=n+1}^{m} a_s\,\theta(t-s) \Rightarrow x(t)$$

Clearly, if $x(t)$ is a running average of a stationary time-series $\theta(t)$, then $x(t)$ is stationary and subordinated to $\theta(t)$.

Definition 7.61 A stationary time-series $\theta(t)$ is called *white noise* if $\theta(t)$, $-\infty<t<\infty$, is an orthonormal sequence (i.e. if its autocovariance $R_{\theta\theta}(s)=1$ if $s=0$, and $=0$ if $s\neq 0$).

From Theorem 7.20 we see that if $\theta(t)$ is white noise, then $F_{\theta\theta}=\nu+0{\cdot}5$, and conversely any stationary time-series with a spectral distribution function equal to $\nu+0{\cdot}5$ is white noise. If $\theta(t)$ is white noise, then $\theta(t)$, $-\infty<t<\infty$, is a maximal orthonormal set in the time-domain $\mathbf{\Theta}$ of $\theta(t)$, and hence every element in this time-domain has a unique Fourier series representation in terms of $\theta(t)$ (Theorem 4.74). Thus, for a stationary time-series $x(t)$ to be a running average of white noise $\theta(t)$ it is necessary and sufficient that $x(t)$ be subordinated to $\theta(t)$; in this case, $x(t)$ has the unique Fourier representation

$$x(t)=\sum_{s=-\infty}^{\infty} b_s\,\theta(t-s),\quad b_s=((x(t),\theta(t-s))) \tag{7.60}$$

Because $dF_{\theta\theta}(\nu)=d\nu$, we see that the spectrum S_θ of white noise is the whole interval $-0{\cdot}5\leqslant\nu\leqslant 0{\cdot}5$, except for a set of Lebesgue measure zero. Since the Lebesgue measure of a line segment is equal to the length of the segment, S_θ has Lebesgue measure equal to $0{\cdot}5-(-0{\cdot}5)=1$. The frequency domain $\mathbf{L}^2(F_{\theta\theta})$, which (because $F_{\theta\theta}=0{\cdot}5+\nu$) we shall denote by $\mathbf{L}^2(0{\cdot}5+\nu)$, is the $\mathbf{L}^2$ space with Lebesgue measure $d\nu$.

Theorem 7.60 (Kolmogorov, 1941) Given the stationary time-series $x(t)$; if $x(t)$ is a running average of white noise $\theta(t)$, then $F_{xx}(\nu)$ is absolutely continuous with respect to Lebesgue measure. If $F_{xx}(\nu)$ is absolutely continuous with respect to Lebesgue measure, and if the orthogonal complement $\mathbf{V}_\infty-\mathbf{X}$ has infinite dimension, then $x(t)$ is a running average of white noise $\theta(t)$.

Proof (1). By hypothesis, $x(t)$ has the representation (7.60), and so $x(t)$ is subordinated to $\theta(t)$. By Theorem 7.40, we have

$$F_{xx}(\nu)=\int_{-0\cdot5}^{\nu}|\Phi_x^{(\theta)}(\nu)|^2\,d\nu,\quad F_{x\theta}(\nu)=\int_{-0\cdot5}^{\nu}\Phi_x^{(\theta)}(\nu)\,d\nu \tag{7.61}$$

which shows that F_{xx} is absolutely continuous with respect to Lebesgue measure. We also note that, under the isomorphism between $\boldsymbol{\Theta}$ and $\mathbf{L}^2(0{\cdot}5+\nu)$, we have $x(0) \leftrightarrow \Phi_x^{(\theta)}$ and $\theta(t) \leftrightarrow e^{2\pi i\nu t}$. Hence, by Property (4) of Theorem 7.21, the coefficient b_s in the representation (7.60) is given by

$$b_s = ((x(0), \theta(-s))) = \int_{-0\cdot5}^{0\cdot5} \Phi_x^{(\theta)} e^{2\pi i\nu s}\, d\nu$$

and so $\Phi_x^{(\theta)}$ has the Fourier series in $\mathbf{L}^2(0{\cdot}5+\nu)$ given by

$$\Phi_x^{(\theta)} = \sum_{s=-\infty}^{\infty} b_s e^{-2\pi i\nu s} \tag{7.62}$$

with b_s as Fourier coefficients. From equation (7.61) the spectral density function is

$$F'_{xx}(\nu) = |\Phi_x^{(\theta)}(\nu)|^2 = \left| \sum_{s=-\infty}^{\infty} b_s e^{-2\pi i\nu s} \right|^2 \tag{7.63}$$

Q.E.D. (1).

Proof (2). By hypothesis, F_{xx} is absolutely continuous with respect to Lebesgue measure, and $\mathbf{V}_\infty - \mathbf{X}$ has infinite dimension. Hence, by the Radon–Nikodym theorem, there exists a non-negative function $|\Upsilon(\nu)|^2$ such that

$$F_{xx}(\nu) = \int_{-0\cdot5}^{0\cdot5} |\Upsilon(\nu)|^2\, d\nu \tag{7.64}$$

This shows that $\Upsilon(\nu)$ (which is any square root of $|\Upsilon(\nu)|^2$) belongs to $\mathbf{L}^2(0{\cdot}5+\nu)$. Define the functions

$$F_{11}(\nu) = F_{xx}(\nu), \quad F_{12}(\nu) = \int_{-0\cdot5}^{\nu} \Upsilon(\nu)\, d\nu, \quad F_{22}(\nu) = 0{\cdot}5+\nu$$

These functions satisfy $|\Delta F_{12}|^2 \leqslant \Delta F_{11} \Delta F_{22}$ and so represent the spectral functions of a multiple stationary time-series. Because $\mathbf{V}_\infty - \mathbf{X}$ is infinite, there always exists a time-series $\theta(t)$ such that $x(t)$ and $\theta(t)$ are multiple-stationary and such that

$$F_{x\theta}(\nu) = F_{12}(\nu) = \int_{-0\cdot5}^{\nu} \Upsilon(\nu)\, d\nu, \quad F_{\theta\theta}(\nu) = F_{22}(\nu) = 0{\cdot}5+\nu \tag{7.65}$$

Because $F_{\theta\theta} = 0{\cdot}5+\nu$, $\theta(t)$ is white noise. Equations (7.64) and (7.65), by Theorem 7.41, show that $x(t)$ is subordinated to $\theta(t)$. (Note: Two cases may occur: (1) $\Upsilon(\nu) \neq 0$ almost everywhere with respect to Lebesgue measure. Then by Theorem 7.42, $x(t)$ is equivalent to $\theta(t)$, i.e. $\mathbf{X} = \boldsymbol{\Theta}$, in which case use of the infinite dimension of $\mathbf{V}_\infty - \mathbf{X}$ is not needed. (2) $\Upsilon(\nu) = 0$ on a set of positive Lebesgue measure. Then $x(t)$ is not equivalent to $\theta(t)$, i.e. $\mathbf{X}$ is properly contained in $\boldsymbol{\Theta}$, and so $\boldsymbol{\Theta} - \mathbf{X}$ has infinite dimension. But $\boldsymbol{\Theta} - \mathbf{X}$ is contained in $\mathbf{V}_\infty - \mathbf{X}$, and so the condition that $\mathbf{V}_\infty - \mathbf{X}$ have infinite dimension is needed.) Because $x(t)$ is subordinated to $\theta(t)$, we can use equation (7.60) to express $x(t)$ as a running average. Q.E.D. (2).

Case (1) in the proof above, by noting that $F'_{xx}(\nu) = |\Psi(\nu)|^2$, becomes

Theorem 7.61 A stationary time-series $x(t)$ is a running average of white noise $\theta(t)$ with $\mathbf{X} = \mathbf{\Theta}$ if and only if $F_{xx}(\nu)$ is absolutely continuous with respect to Lebesgue measure and $F'_{xx}(\nu)$ is positive almost everywhere with respect to Lebesgue measure.

7.7 The Wold decomposition

Definition 7.70 Let $x(s)$, $-\infty < s < \infty$, be a stationary time-series. A given integer t is called the *present time*; integers $s < t$, the *past time*, and integers $s > t$, the *future time*.

Definition 7.71 The closed linear manifold $\mathbf{X} = \mathbf{X}(\infty)$ spanned by $x(s)$, $-\infty < s < \infty$, is called the *time-domain*. The closed linear manifold $\mathbf{X}(t)$ spanned by the present and past variates $x(s)$, $s \leqslant t$, is called the *present-and-past time-domain*. The common part $\mathbf{X}(-\infty)$ of all the $\mathbf{X}(t)$, i.e. $\mathbf{X}(-\infty) \equiv \bigcap_{-\infty}^{\infty} \mathbf{X}(t)$, is called the *remote-past time-domain*. The orthogonal complement $\mathbf{X}(t) - \mathbf{X}(-\infty)$ is called the *present-and-recent-past time-domain*.

The remote-past time-domain is a closed linear manifold, and so is the present-and-recent-past time-domain. Since $x(t)\,U^s = x(t+s)$, we may write $\mathbf{X}(t)\,U^s = \mathbf{X}(t+s)$. We also have

$$\mathbf{X}(-\infty)\,U^s = \mathbf{X}(-\infty)$$

Definition 7.72 A stationary time-series $x(t)$ is said to be *singular* if $\mathbf{X}(-\infty) = \mathbf{X}$. A stationary time-series that is not singular is said to be *non-singular*. A stationary time-series $x(t)$ is said to be *regular* if $x(t) \neq 0$ and $\mathbf{X}(-\infty) = 0$.

We see that a stationary time-series is singular if and only if $\mathbf{X}(t) = \mathbf{X}$ for all t. In fact, if $\mathbf{X}(t) = \mathbf{X}$ for a given t, it must be true for all t (because $x(t)$ is stationary), and

$$\mathbf{X}(-\infty) = \mathbf{X}(t) = \mathbf{X}(t+s) = \mathbf{X}$$

Thus a stationary time-series is singular if and only if its present-and-recent-past time-domain is equal to the null variate; i.e. $\mathbf{X}(t) - \mathbf{X}(-\infty) = 0$. A non-null stationary time-series is regular if and only if its remote-past time-domain is equal to the null variate; i.e. $\mathbf{X}(-\infty) = 0$.

Theorem 7.70 *The Wold decomposition* (Wold, 1938) Let $x(t)$, $-\infty < t < \infty$, be a non-singular stationary time-series. Then $x(t)$ has the unique decomposition

$$x(t) = y(t) + z(t), \quad \text{with} \quad z(t) = b_0\,\theta(t) + b_1\,\theta(t-1) + \dots \tag{7.70}$$

where the components $\theta(t), y(t), z(t)$, $-\infty<t<\infty$, form a multiple stationary time-series with the properties:

(1) Each of the variates $\theta(t), y(t), z(t)$ is in $\mathbf{X}(t)$.

(2) The stationary time-series $y(t)$ is singular and the stationary time-series $z(t)$ is regular.

(3) $\theta(t)$ is white noise.

(4) The variates $y(t)$, $-\infty<t<\infty$, are orthogonal to the variates $\theta(t)$ and $z(t)$, $-\infty<t<\infty$; i.e.

$$((y(s), \theta(t))) = ((y(s), z(t))) = 0$$

for $s=t$ as well as $s \neq t$.

(5) $b_0>0$ and $b_0^2+|b_1|^2+|b_2|^2+\ldots<\infty$.

Proof. Application of the Projection Theorem 5.20 shows that $x(t)$ can be resolved in one and only one way into two orthogonal components

$$x(t) = x(t) P_{\mathbf{X}(t-1)} + w(t)$$

where the first component is in $\mathbf{X}(t-1)$ and the second component $w(t)$, the normal to $\mathbf{X}(t-1)$, is in $\mathbf{X}(t)-\mathbf{X}(t-1)$. Because $x(t)$ is regular, it follows that $\mathbf{X}(t) \neq \mathbf{X}(t-1)$, and so $\|w(t)\|>0$. Define $\theta(t)$ to be $w(t)/\|w(t)\|$. Since $\theta(t)$ is in $\mathbf{X}(t)-\mathbf{X}(t-1)$, and since $\theta(t-s), s \geqslant 1$, is in $\mathbf{X}(t-1)$, we see that $((\theta(t), \theta(t-s)))=0$ for $s \geqslant 1$, and so the $\theta(t)$, $-\infty<t<\infty$, form an orthonormal set. The $\theta(t)$ are called the (white noise) *innovations*. The closed linear manifold spanned by $\theta(t)$ is $[\theta(t)]=\mathbf{X}(t)-\mathbf{X}(t-1)$.

Let $\mathbf{\Theta}(t)$ be the closed linear manifold spanned by the present and past innovations $\theta(s)$, $s \leqslant t$. Suppose $y \in \mathbf{X}(t)-\mathbf{\Theta}(t)$; i.e. $y \in \mathbf{X}(t)$ and $y \perp \mathbf{\Theta}(t)$. Then $y \perp [\theta(t)] = \mathbf{X}(t)-\mathbf{X}(t-1)$ and so $y \in \mathbf{X}(t-1)$. In a similar way, it follows that $y \in \mathbf{X}(t-2), y \in \mathbf{X}(t-3), \ldots$, and so by induction $y \in \mathbf{X}(-\infty)$. Hence $\mathbf{X}(t)-\mathbf{\Theta}(t) \supset \mathbf{X}(-\infty)$. Conversely, suppose $y \in \mathbf{X}(-\infty)$. Since $\mathbf{X}(-\infty) \subset \mathbf{X}(s-1)$, it follows that $y \in \mathbf{X}(s-1)$. But $\mathbf{X}(s-1) \perp \theta(s)$, so $y \perp \theta(s)$ for arbitrary s. Thus $y \perp \mathbf{\Theta}(t)$, and so $y \in \mathbf{X}(t)-\mathbf{\Theta}(t)$. Hence $\mathbf{X}(-\infty) \supset \mathbf{X}(t)-\mathbf{\Theta}(t)$. Therefore $\mathbf{X}(-\infty)=\mathbf{X}(t)-\mathbf{\Theta}(t)$, and so

$$\mathbf{X}(t)-\mathbf{X}(-\infty) = \mathbf{X}(t)-[\mathbf{X}(t)-\mathbf{\Theta}(t)] = \mathbf{\Theta}(t)$$

which says that *the present-and-recent-past time-domain is equal to the closed linear manifold spanned by the present and past innovations.* In other words, the present and past innovations, $\theta(s)$ for $s \leqslant t$, form a maximal orthonormal set in the present-and-recent-past time-domain.

Thus we have $\mathbf{X}(t)=\mathbf{X}(-\infty)+\mathbf{\Theta}(t)$ where $\mathbf{X}(-\infty) \perp \mathbf{\Theta}(t)$. By the Projection Theorem 5.20, the time-series $x(t)$ has the unique decomposition

$$x(t) = x(t) P_{\mathbf{X}(-\infty)} + x(t) P_{\mathbf{\Theta}(t)}$$

which we denote by $x(t) = y(t) + z(t)$. The time-series $y(t)$ and $z(t)$ are stationary. It can be readily shown that the present-and-past time-domain $\mathbf{Y}(t)$ of $y(t)$ is $\mathbf{X}(-\infty)$, and that the present-and-past time-domain $\mathbf{Z}(t)$ of $z(t)$ is $\mathbf{\Theta}(t)$. Since $\mathbf{Y}(t) = \mathbf{X}(-\infty)$ for all t, it follows that the time-series $y(t)$ is singular: $\mathbf{Y}(t) = \mathbf{Y}(-\infty)$.

By Theorems 4.74 and 5.20, $z(t)$ has the Fourier series representation

$$z(t) = \sum_{s=0}^{\infty} b_s \theta(t-s)$$

where the Fourier coefficients are given by

$$b_s = ((x(t), \theta(t-s))), \quad s \geqslant 0; \quad \sum_{s=0}^{\infty} |b_s|^2 < \infty$$

Because $\mathbf{Z}(-\infty) = \mathbf{\Theta}(-\infty) = \mathbf{X}(-\infty) - \mathbf{X}(-\infty) = 0$, we see that $z(t)$ is regular. We have

$$b_0 = ((x(t), \theta(t))) = ((x(t) P_{\mathbf{X}(t-1)} + w(t), \theta(t))) = \|w(t)\|$$

so b_0 is real and positive, and $z(t)$ is non-null. Q.E.D.

Theorem 7.71 In the Wold decomposition (7.70), the stationary time-series $z(t)$ and $\theta(t)$ are equivalent (i.e. $\mathbf{Z} = \mathbf{\Theta}$) and their spectra S_z and S_θ coincide.

Theorem 7.72 In the Wold decomposition (7.70) the spectrum S_z of $z(t)$ is a set of Lebesgue measure one.

Proof. The spectrum S_θ of white noise has Lebesgue measure one (Section 7.6) and $S_\theta = S_z$, by Theorem 7.71. Q.E.D.

Theorem 7.73 In the Wold decomposition (7.70) the spectrum S_y of $y(t)$ is a set of Lebesgue measure zero.

Proof. By hypothesis, $y(t)$ and $z(t)$ are orthogonal, and each subordinated to $x(t)$. From Theorem 7.40 we know that

$$F_{yy}(\nu) = \int_{-0.5}^{\nu} |\Phi_y^{(x)}(\nu)|^2 \, dF_{xx}(\nu)$$

$$F_{yx}(\nu) = \int_{-0.5}^{\nu} \Phi_y^{(x)}(\nu) \, dF_{xx}(\nu)$$

and similarly for F_{zz} and F_{zx}. But because $y(t)$ and $z(t)$ are orthogonal, $R_{yx}(s) = R_{yy}(s) + R_{yz}(s) = R_{yy}(s)$ and so $F_{yx} = F_{yy}$. Hence, by the above equations,

$$|\Phi_y^{(x)}(\nu)|^2 = \Phi_y^{(x)}(\nu), \quad |\Phi_z^{(x)}(\nu)|^2 = \Phi_z^{(x)}(\nu) \tag{7.71}$$

almost everywhere with respect to F_{xx}-measure. Moreover,

$$\Phi_y^{(x)}(\nu) + \Phi_z^{(x)}(\nu) = \frac{dF_{yy}}{dF_{xx}} + \frac{dF_{zz}}{dF_{xx}} = \frac{dF_{xx}}{dF_{xx}} = 1 \tag{7.72}$$

For equations (7.71) and (7.72) to be satisfied, we must have

$$\Phi_y^{(x)}(\nu)\,\Phi_z^{(x)}(\nu) = 0$$

almost everywhere with respect to F_{xx}-measure. Hence the spectra S_y of $y(t)$ and S_z of $z(t)$ do not overlap. That is, $S_x = S_y \cup S_z$ and $S_y \cap S_z =$ null set. But by Theorem 7.72, S_z has Lebesgue measure one. Since S_x at most can have Lebesgue measure one, S_x does have Lebesgue measure one, and S_y Lebesgue measure zero. We note that in the isomorphism between $\mathbf{X}$ and $\mathbf{L}^2(F_{xx})$

$$y(0) \leftrightarrow \Phi_y^{(x)}(\nu) = \Lambda_y(\nu), \quad z(0) \leftrightarrow \Phi_z^{(x)}(\nu) = \Lambda_z(\nu) \tag{7.73}$$

where $\Lambda_y(\nu)$ is the characteristic function of S_y (i.e. $\Lambda_y(\nu) = 1$ for ν in S_y and $= 0$ for ν not in S_y) and $\Lambda_z(\nu)$ is the characteristic function of S_z. Q.E.D.

Theorem 7.74 Given a non-singular stationary time-series $x(t)$ with spectral decomposition $x(t) = x_1(t)+x_2(t)+x_3(t)$ and Wold decomposition $x(t) = y(t)+z(t)$, then $y(t) = x_1(t)+x_2(t)$ and $z(t) = x_3(t)$.

Proof. Because $F_3(\nu)$ is absolutely continuous with respect to Lebesgue measure, $x_3(t)$ is a running average of white noise, by Theorem 7.60. Because $z(t)$ is a running average of white noise, $F_{zz}(\nu)$ is absolutely continuous with respect to Lebesgue measure. Because S_1, S_2, and S_y have Lebesgue measure zero, F_{11}, F_{22}, and F_{yy} are not absolutely continuous with respect to Lebesgue measure. Since $S = S_1 \cup S_2 \cup S_3 = S_y \cup S_z$, we have $S_1 \cup S_2 = S_y$ (which has Lebesgue measure zero) and $S_3 = S_z$ (which has Lebesgue measure one). Q.E.D.

Theorem 7.75 Given a non-singular stationary time-series $x(t)$, then its spectral power function $F'(\nu)$ exists and is strictly positive, except on a ν-set of Lebesgue measure zero.

Proof. The spectrum $S_3 = S_z$ has Lebesgue measure one, and so by Property (5) of Theorem 7.50, $F'(\nu) > 0$ for ν in S_3. Q.E.D.

Theorem 7.76 In the Wold decomposition (7.70) we have

$$\theta(t) \leftrightarrow \frac{\Lambda_z(\nu)}{B(\nu)}\, e^{2\pi i\nu t} \tag{7.74}$$

in the isomorphism between the time-domain $\mathbf{X}$ and the frequency-domain $\mathbf{L}^2(F_{xx})$, where

$$\Lambda_z(\nu) = \begin{cases} 0, & \nu \in S_y \\ 1, & \nu \in S_z \end{cases}, \quad B(\nu) = \sum_{s=0}^{\infty} b_s\, e^{-2\pi i\nu s}$$

Proof. Let $\theta(0) \leftrightarrow \Psi(\nu)$ in the isomorphism between $\mathbf{X}$ and $\mathbf{L}^2(F_{xx})$. Then by Property (5) of Theorem 7.21 we have

$$z(0) = \sum_{s=0}^{\infty} b_s \theta(-s) \leftrightarrow \sum_{s=0}^{\infty} b_s \Psi(\nu) \mathrm{e}^{-2\pi i \nu s}$$

in the isomorphism. But by equation (7.73), $z(0) \leftrightarrow \Lambda_z(\nu)$, and solving for $\Psi(\nu)$ we get the desired result. Q.E.D.

A non-singular stationary time-series $x(t)$ is regular if and only if $y(t) = 0$ in its Wold decomposition. Then $x(t)$ is equivalent to $\theta(t)$, and has the one-sided (i.e. $b_s = 0$ for $s < 0$) representation

$$x(t) = \sum_{s=0}^{\infty} b_s \theta(t-s), \quad b_s = ((x(t), \theta(t-s))) \tag{7.75}$$

By equation (7.62) we have

$$x(0) \leftrightarrow \Phi_x^{(\theta)}(\nu) = \sum_{s=0}^{\infty} b_s \mathrm{e}^{-2\pi i \nu s} \equiv B(\nu) \tag{7.76}$$

which defines $B(\nu)$. Since

$$\|x(t)\|^2 = \sum_{s=0}^{\infty} |b_s|^2 = R_{xx}(0)$$

is finite, the power series $\sum_{0}^{\infty} b_s \xi^s$ is analytic within the unit circle $|\xi| < 1$, and its limit value on the boundary of the unit circle is $B(\nu)$ where $\xi = \mathrm{e}^{-2\pi i \nu}$ for $|\xi| = 1$. Since $x(t) \neq 0$, the power series is not identically zero, and hence $B(\nu) \neq 0$ almost everywhere for $-0{\cdot}5 \leqslant \nu \leqslant 0{\cdot}5$ (i.e. $B(\nu) \neq 0$ for ν in S_θ). Hence by Theorem 7.42, $x(t)$ is equivalent to $\theta(t)$. By equation (7.63), $F'(\nu) = |B(\nu)|^2 > 0$ almost everywhere.

Fundamental results of Kolmogorov (1941) and Wiener (1942) are

Theorem 7.77 If $x(t)$ is regular, then $\sum_{s=0}^{\infty} b_s \xi^s$ has no zeros in the unit circle $|\xi| < 1$, where the b_s are the coefficients of the Wold decomposition of $x(t)$.

Theorem 7.78 $x(t)$ is regular if and only if all the following conditions are true for $-0{\cdot}5 \leqslant \nu \leqslant 0{\cdot}5$:

(1) $F_{xx}(\nu)$ is absolutely continuous with respect to Lebesgue measure.

(2) $F'_{xx}(\nu) > 0$, except for a set of Lebesgue measure zero.

(3) $\int_{-0\cdot5}^{0\cdot5} \log F'(\nu)\, d\nu$ is finite (i.e. $> -\infty$, since it is always $< \infty$).

Because condition (1) in Theorem (7.78) is equivalent to $S_1 \cup S_2 =$ null set, we have

Theorem 7.79 $x(t)$ is non-singular if and only if both conditions (2) and (3) are true.

Example 7.70 Consider a time-series $x(t)$ for which conditions (1) and (2) hold, and (3) does not. Then we see that $x(t)$ is singular, and moreover by Theorem 7.61, $x(t)$ is a running average of white noise, with spectrum S_x of Lebesgue measure 1.

7.8 Electric filters, prediction, and filtering

The stochastic process given by the finite linear operation

$$\epsilon(t) = a_0 x(t) + a_1 x(t-1) + \ldots + a_n x(t-n) \tag{7.80}$$

was introduced by Yule (1921, 1927). Equation (7.80) states that the convolution of the *input* $x(t)$ with the constant coefficients $a_0, \ldots, a_n$ yields the *output* $\epsilon(t)$. Because equation (7.80) involves only a finite number of constant coefficients, and because only inputs at or prior to the time t (and no future inputs) are required to compute the output at time t, the linear operation (7.80) can be realized physically by a linear *electric filter* with pulsed data (i.e. a pulsed filter), or by a (high-speed) digital computer. A pulsed filter is a transmission device which is supplied with input data at specified equally-spaced times and, in response, furnishes output data at the same moments. Hurewicz (1947) shows that a pulsed filter (7.80) is equivalent to the combination of a clamping device together with a linear filter of the continuous type.

The *transient-response function* of a filter is the response of the filter to a single unit input at $t = 0$ (i.e. $x(0) = 1, x(t) = 0$ for $t \neq 0$). Hence we see that the constant coefficients $a_0, a_1, \ldots, a_n$ are the transient-response function of the filter (7.80). A filter is essentially an instrument for the separation of different frequency ranges. To see this, let the input $x(t)$ to the filter (7.80) be the (complex) sinusoidal wave $x(t) = e^{2\pi i\nu t}$. Then we see that the output is necessarily a sinusoidal wave of the same frequency ν but generally will differ in amplitude and phase, and so may be written $\epsilon(t) = A(\nu)\, e^{2\pi i\nu t}$. The function $A(\nu)$, which is a complex function expressing the amplitude and phase changes, is called the *transfer function*, or the *filter characteristics*, or the *frequency-response function*, or the *gain*, of the filter. Substituting into equation (7.80), we have $A(\nu) = \sum_0^n a_s e^{-2\pi i\nu s}$. The absolute value $|A(\nu)|$ is called the *absolute gain*, and $|A(\nu)|^2$ the *power transfer function*.

If $x(t)$ is a stationary process, then $\epsilon(t)$ given by equation (7.80) (N.B. n is finite!) is called a *Yule process*. Using the spectral representation of $x(t)$, we have

$$\epsilon(t) = \sum_0^n a_s x(t-s) = \int_{-0\cdot5}^{0\cdot5} e^{2\pi i\nu t} A(\nu)\, dx(0)\, E_x(\nu) \tag{7.81}$$

where $A(\nu)$ is the transfer function. The Yule process $\epsilon(t)$ is stationary, and subordinated to $x(t)$. In the isomorphism between $\mathbf{X}$ and $\mathbf{L}^2(F_{xx})$, we have

$$\epsilon(0) = \sum_0^n a_s x(-s) \leftrightarrow A(\nu) = \sum_0^n a_s \mathrm{e}^{-2\pi i\nu s}$$

From equation (7.81) we obtain the autocovariance

$$R_{\epsilon\epsilon}(s) = \int_{-0\cdot5}^{0\cdot5} \mathrm{e}^{2\pi i\nu s} |A(\nu)|^2 \, dF_{xx}(\nu) \tag{7.82}$$

so that

$$F_{\epsilon\epsilon}(\nu) = \int_{-0\cdot5}^{\nu} |A(\nu)|^2 \, dF_{xx}(\nu) \tag{7.83}$$

These equations show the filtering action of the transfer function. In particular, we see that the input spectral power $dF_{xx}(\nu)$ is multiplied by the power transfer function $|A(\nu)|^2$ to yield the output spectral power $dF_{\epsilon\epsilon}(\nu)$.

Now, let $\epsilon(t)$ be the limit $(\Rightarrow)$ of a sequence of Yule processes $\sum_0^n a_s x(t-s)$ as $n \to \infty$ where $x(t)$ is a stationary time-series. Then $\epsilon(t)$ is stationary and subordinated to $x(t)$. By Theorem 7.40, we have $\epsilon(0) \leftrightarrow \Phi_\epsilon^{(x)}(\nu)$, which serves as the definition of the transfer function; i.e. $A(\nu) \equiv \Phi_\epsilon^{(x)}(\nu)$. Then, with this $A(\nu)$, $\epsilon(t)$, $R_{\epsilon\epsilon}(s)$, $F_{\epsilon\epsilon}(\nu)$ are equal respectively to the right-hand sides of equations (7.81), (7.82), (7.83).

Let $x(t)$ be a non-singular, stationary time-series. The problem of the linear, least-squares *prediction* of $x(t+\alpha)$ from the whole past $x(s), s \leqslant t$, is the problem of finding an element $\hat{x}(t+\alpha)$ in $\mathbf{X}(t)$ such that the distance $\|x(t+\alpha) - \hat{x}(t+\alpha)\|$ is a minimum. Clearly the solution is

$$\hat{x}(t+\alpha) = x(t+\alpha) P_{\mathbf{X}(t)}$$

which is, since $\theta(s) P_{\mathbf{X}(t)} = \theta(s)$ if $s \leqslant t$, and $= 0$ if $s > t$,

$$\begin{aligned} \hat{x}(t+\alpha) &= y(t+\alpha) P_{\mathbf{X}(t)} + \sum_{s=0}^{\infty} b_s \theta(t+\alpha-s) P_{\mathbf{X}(t)} \\ &= y(t+\alpha) + \sum_{s=\alpha}^{\infty} b_s \theta(t+\alpha-s) \end{aligned} \tag{7.84}$$

a result due to Wold (1938). This equation exhibits the least-squares prediction operator. We have, by equations (7.73) and (7.74),

$$\hat{x}(0+\alpha) = y(0) + \sum_{s=\alpha}^{\infty} b_s \theta(\alpha - s) \leftrightarrow \Lambda_y(\nu) + \frac{\Lambda_z(\nu)}{B(\nu)} \sum_{s=\alpha}^{\infty} b_s \mathrm{e}^{-2\pi i\nu(s-\alpha)}$$

Hence the expression on the right is the transfer function of the optimum linear predictor, a result due to Wiener (1942) and Kolmogorov (1941). That is, the transfer function is equal to one for $\nu \in S_y$ (i.e. waves of "pure" frequencies are passed without modulation). The transfer function is equal to

$$\frac{1}{B(\nu)} \sum_{t=0}^{\infty} b_{t+\alpha}\, \mathrm{e}^{-2\pi i \nu t} \tag{7.85}$$

for $\nu \in S_z$ (i.e. waves of "non-pure" frequencies are modulated, i.e. their spectral power density is multiplied by the factor (7.85)). The "non-pure" frequencies have their spectral power density spread over the whole spectral band,

i.e. $F'_{zz}(\nu) > 0$ for ν in $S_z = \{-0{\cdot}5 \leqslant \nu \leqslant 0{\cdot}5$, except for $S_y\}$

The "pure" frequencies have their power so concentrated that it is crowded in the set S_y of Lebesgue measure zero. Since

$$\|y(t+\alpha) - \hat{y}(t+\alpha)\| = 0$$

the past of $y(t)$ determines its future for an infinite time (i.e. for arbitrary *prediction lead* $\alpha > 0$), so the singular component $y(t)$ is *deterministic.* On the other hand, the future of the non-singular time-series $x(t)$ is not completely determined by a linear operation on its past. That is, we have the (non-null) *prediction error*

$$x(t+\alpha) - \hat{x}(t+\alpha) = \sum_{s=0}^{\alpha-1} b_s\, \theta(t+\alpha-s) \tag{7.86}$$

and the positive *mean-square prediction error*

$$\|x(t+\alpha) - \hat{x}(t+\alpha)\| = \sum_{s=0}^{\alpha-1} |b_s|^2 > 0$$

since $b_0 > 0$.

Now suppose that $x(t)$, $m(t)$, and $n(t)$ form a multiple stationary time-series, and $x(t)$ is additively composed of the message $m(t)$ and the noise $n(t)$; i.e. $x(t) = m(t) + n(t)$. We assume that these three time-series are regular. Hence they have absolutely continuous spectral distribution functions. The problem of linear least-squares *filtering* (using the whole past $x(s), s \leqslant t$) is the problem of finding an element $\hat{m}(t+\alpha)$ in $\mathbf{X}(t)$ such that the distance $\|m(t+\alpha) - \hat{m}(t+\alpha)\|$ is a minimum. To find the solution, we use the Wold decomposition

$$x(t) = \sum_{s=0}^{\infty} b_s\, \theta(t-s)$$

Since $\theta(t)$, $-\infty < t < \infty$, is a maximal orthonormal set in $\mathbf{X}$, the projection of the message $m(t)$ on the time-domain $\mathbf{X}$ has the Fourier series representation

$$m(t) P_{\mathbf{X}} = \sum_{s=-\infty}^{\infty} c_s\, \theta(t-s), \quad c_s = ((m(t), \theta(t-s))) \tag{7.87}$$

Hence the least-squares approximation to $m(t+\alpha)$ is

$$\hat{m}(t+\alpha) = m(t+\alpha)P_{\mathbf{X}(t)} = m(t+\alpha)P_{\mathbf{X}}P_{\mathbf{X}(t)}$$

since projection on $\mathbf{X}$ followed by projection of $\mathbf{X}(t)$ is the same as projection on $\mathbf{X}(t)$ directly. Using equation (7.87), we have

$$\hat{m}(t+\alpha) = \sum_{s=-\infty}^{\infty} c_s\,\theta(t+\alpha-s)P_{\mathbf{X}(t)} = \sum_{s=\alpha}^{\infty} c_s\,\theta(t+\alpha-s) \qquad (7.88)$$

The transfer function is the expression on the right of

$$\hat{m}(0+\alpha) \leftrightarrow \frac{1}{B(\nu)}\sum_{t=0}^{\infty} c_{t+\alpha}\,e^{-2\pi i\nu t} \qquad (7.89)$$

a result due to Wiener (1942). This is the optimum transfer function for separating message from message-plus-noise by a linear operation. If α is positive, it is called the *lead* of the filter; if α is negative, then $-\alpha$ is called the *lag*. Thc classic papers on the intuitive content of Wiener's work as applied to electric networks are Blackman, Bode, and Shannon (1946) and Bode and Shannon (1950).

Because the transfer functions (7.85) and (7.89) in general involve an infinite number of coefficients, in this sense they are not necessarily physically realizable by an electric filter, or high-speed digital computer. Nevertheless, because $\hat{x}(t+\alpha)$ and $\hat{m}(t+\alpha)$ are elements of the present-and-past time-domain $\mathbf{X}(t)$, and since every element in $\mathbf{X}(t)$ is a limit ($\Rightarrow$) of a sequence of Yule processes (7.81) (see Section 4.6), we can approximate $\hat{x}(t+\alpha)$ and $\hat{m}(t+\alpha)$ by Yule processes which are physically realizable.

7.9 Wavelet theory

The Wold decomposition reaches its full beauty in physical applications. This application to wavelet theory was developed to solve certain problems in the interpretation of geophysical recordings of seismic wave propagation through the earth (Robinson, 1954, 1957).

Suppose that a real-valued time-series $x(t)$ is additively composed of many overlapping wavelets which arrive as time progresses. We assume that each wavelet has the same shape c_s but that each is amplified by a random factor. In keeping with physical processes, we require the wavelets to be one-sided ($c_s = 0$ for $s < 0$) and to damp out

$$\left(\sum_{0}^{\infty} c_s^2 < \infty\right)$$

Let the amplification factor of the wavelet which arrives at time t be $\theta(t)$. Then we see that the wavelet with arrival time t has zero amplitude for times prior to t, has amplitude $c_0\,\theta(t)$ at time t, has amplitude $c_1\,\theta(t)$ at time $t+1$, has amplitude $c_2\,\theta(t)$ at time $t+2$, and so forth. We assume that amplification factors $\theta(t)$, $-\infty < t < \infty$, are uncorrelated variates with the same variance, so by centring and scaling, $\theta(t)$ becomes white noise.

Consider the time instant t. The wavelet which arrives at time t contributes $c_0\,\theta(t)$ to $x(t)$. The wavelet which arrived at time $t-1$ contributes $c_1\,\theta(t-1)$ to $x(t)$. The wavelet with arrival time $t-2$ contributes $c_2\theta(t-2)$ to $x(t)$, and so forth. Because $x(t)$ is the sum of all these contributions, we have

$$x(t) = c_0\,\theta(t) + c_1\,\theta(t-1) + c_2\,\theta(t-2) + \ldots = \sum_{s=0}^{\infty} c_s\,\theta(t-s) \qquad (7.90)$$

In this equation we see that the basic wavelet shape c_s represents the "dynamics", and the amplification factors, or impulses, $\theta(t)$, the "random" nature, of the time-series $x(t)$.

> *Wavelet problem* Given a fixed realization of the time-series $x(s)$, for $s \leqslant t$, find (1) the basic wavelet shape c_s, and (2) the amplification factor $\theta(s)$ for every $s \leqslant t$.

That is, we are given the observed past values of a time-series that is the resultant of many overlapping and interfering wavelets which arrive at various times with random amplifications. We are to pick out each wavelet as a distinct entity and identify it by its arrival time.

For the wavelet problem to have a unique solution, a stability condition must be imposed on the wavelet shape. Suppose that the wavelet shape damps to zero (i.e. $c_s = 0$ for $s > n$). Form the homogeneous difference equation

$$c_0\,\epsilon(t) + c_1\,\epsilon(t-1) + \ldots + c_n\,\epsilon(t-n) = 0$$

from the wavelet shape. For an arbitrary set of initial values $\epsilon(0), \epsilon(1), \ldots, \epsilon(n-1)$, the sequence $\epsilon(n), \epsilon(n+1), \ldots$ may be generated by recursive deductions from this difference equation, and this sequence will form a general solution. A necessary and sufficient condition that $\sum_{0}^{\infty} |\epsilon(t)|^2$ converges is that the magnitude of the roots of the characteristic equation

$$c_0 + c_1\,\xi + \ldots + c_n\,\xi^n = 0$$

be greater than one (i.e. that $\sum_{0}^{n} c_s\,\xi^s$ has no zeros for $|\xi| \leqslant 1$). In this case the general solution of the homogeneous difference equation is a damped oscillation, and the difference equation is said to be stable. In case there are any roots with magnitude equal to one, then the difference equation is semi-stable; whereas if there are any roots with magnitude less than one, the difference equation is unstable. Accordingly, if $c_s = 0$ for $s > n$, we shall impose the stability requirement that $\sum_{0}^{n} c_s\,\xi^s$ (which is analytic in the whole plane) have no zeros for $|\xi| < 1$.

In the general case, we shall impose the stability requirement that $\sum_0^\infty c_s \xi^s$ be analytic for $|\xi|<1$ and have no zeros for $|\xi|<1$. With this stability requirement, $x(t)$ is regular and representation (7.90) is its Wold decomposition (7.75) where $c_s = b_s$.

The solution of the wavelet problem proceeds according to the following steps.

(1) Given the observed time-series $x(s)$, $s \leqslant t$, the *first step* is to average out (i.e. destroy) the random, uncorrelated elements $\theta(s)$, $s \leqslant t$, and yet preserve as much information about the wavelet shape as possible. Phase information cannot be saved; but we can obtain the spectral energy information about b_s. To do this, we compute the autocovariance $R(u)$ from $x(t)$ by the time-average (7.34) in which $m = t-u$, and $n \to -\infty$. Because

$$R(u) = \left(\left(\sum_0^\infty b_s \theta(t+u-s), \sum_0^\infty b_s \theta(t-s)\right)\right) = \sum_{s=0}^\infty b_{s+u} b_s$$

we see that the autocovariance contains no information about individual values of $\theta(s)$. From $R(u)$ we obtain the spectral power function $F'(u)$ by Theorem 7.20. By equations (7.63) and (7.76) we have

$$|B(\nu)| = \left|\sum_{s=0}^\infty b_s \mathrm{e}^{-2\pi i \nu s}\right| = \sqrt{\{F'(\nu)\}}$$

so we may compute the absolute gain $|B(\nu)|$ of the transfer function $B(\nu)$ by taking the square root of the spectral power function. That is, the spectral energy function $|B(\nu)|^2$ for the wavelet shape is equal to the spectral power function $F'(\nu)$ of the time-series.

(2) The *second step* is to compute $B(\nu)$ from $|B(\nu)|$, and thus obtain the wavelet shape b_s. Equivalently, if we write

$$B(\nu) = |B(\nu)| \mathrm{e}^{i\phi(\nu)} \tag{7.91}$$

the second step is to find the phase function $\phi(\nu)$. We recognize this step to be the problem of the factorization of the spectral power function, which was first solved by Wold (1938) in the case of physically realizable transfer functions, and by Kolmogorov (1939, 1941) and Wiener (1942) in the transcendental case. Because $x(t)$ is regular, condition (3) of Theorem 7.78 is true. As a consequence, $\log\sqrt{\{F'(\nu)\}}$, which is an even real function of ν, may be expanded in a real, symmetric Fourier cosine series

$$\log\sqrt{\{F'(\nu)\}} = \tfrac{1}{2}\log F'(\nu) = \sum_{-\infty}^\infty \delta_s \cos 2\pi\nu s = \delta_0 + 2\sum_1^\infty \delta_s \cos 2\pi\nu s \tag{7.92}$$

where the Fourier coefficients δ_s are given by

$$\delta_s = \delta_{-s} = \int_0^{0\cdot 5} \cos 2\pi\nu s \log F'(\nu)\, d\nu \tag{7.93}$$

By taking the logarithm of each side of equation (7.91), and utilizing equation (7.92), we have

$$\log B(\nu) = \log\sqrt{\{F'(\nu)\}} + i\phi(\nu) = \delta_0 + 2\sum_1^\infty \delta_s \cos 2\pi\nu s + i\phi(\nu) \qquad (7.94)$$

Because $x(t)$ is regular, $\sum_0^\infty b_s \xi^s$ has no singularities or zeros within the unit circle. Hence $\log \sum_0^\infty b_s \xi^s$ is analytic within the unit circle, and consequently has the power series representation

$$\log \sum_0^\infty b_s \xi^s = \beta_0 + 2\sum_1^\infty \beta_s \xi^s \qquad (7.95)$$

for $|\xi| < 1$. As $|\xi|$ approaches 1, this power series has

$$\log B(\nu) = \beta_0 + 2\sum_1^\infty \beta_s e^{-2\pi i\nu s}$$

$$= \beta_0 + 2\sum_1^\infty \beta_s \cos 2\pi\nu s - 2i\sum_1^\infty \beta_t \sin 2\pi\nu s \qquad (7.96)$$

as its limit values on the unit circle $|\xi| = 1$ where $\xi = e^{-2\pi i\nu}$.

Comparing equations (7.94) and (7.96), we see that $\delta_s = \beta_s$ and

$$\phi(\nu) = -2\sum_1^\infty \delta_s \sin 2\pi\nu s = -2\sum_1^\infty \sin 2\pi\nu s \int_0^{0\cdot5} \cos 2\pi\nu s \log F'(\nu)\, d\nu$$

This equation shows that the phase function $\phi(\nu)$ is the Hilbert transform of $\log\sqrt{\{F'(\nu)\}}$, and thus $\phi(\nu)$ is uniquely determined by $x(s)$, $s \leqslant t$. Substituting $\phi(\nu)$ into equation (7.91), we obtain the desired transfer function $B(\nu)$, from which we obtain the wavelet shape by

$$b_s = \int_{-0\cdot5}^{0\cdot5} B(\nu)\, e^{2\pi i\nu s}\, d\nu$$

Alternatively, we may write equation (7.95) as

$$\sum_0^\infty b_s \xi^s = \exp\left(\delta_0 + 2\sum_1^\infty \delta_s \xi^s\right)$$

which gives the wavelet shape b_s in terms of the δ_s of equation (7.93). In particular, this equation gives $b_0 = e^{\delta_0}$ which shows the reason for condition (3) in Theorem 7.78.

(3) The *last step* consists of removing the wavelet shape b_s from $x(t)$, thereby leaving, as a residual, the random components $\theta(t)$. The wavelet shape b_s uniquely determines the least-squares prediction operator, as seen by equation (7.85). If we let the prediction lead $\alpha = 1$, then equation (7.86) shows that if we predict $x(s)$ from $\mathbf{X}(s-1)$ the prediction error is $b_0\,\theta(s)$. We may repeat this process for all $s \leqslant t$,

thereby obtaining the amplification factors $\theta(s)$ for $s \leqslant t$. Thus we have obtained the theoretical solution to the wavelet problem.

The practical solution of the wavelet problem in which we only have a finite section $x(1), x(2), \ldots, x(N)$ of an observed time-series involves statistical estimation. The method we proposed uses first the Gauss method of least squares to obtain the prediction operator (with $n < N$)

$$\hat{x}(t+1) = c + \sum_{s=0}^{n} k_s x(t-s) \tag{7.97}$$

where the constant c appears if the section of $x(t)$ does not have a zero average. Then, provided the difference equation is stable, the wavelet shape is computed by

$$b_{t+1} = \sum_{s=0}^{n} k_s b_{t-s}, \quad (t = 0, 1, 2, \ldots) \tag{7.98}$$

where we let $b_t = 0$ for $t < 0$, and b_0^2 equal to the variance of the prediction errors $w(t) = x(t) - \hat{x}(t)$. That is, the wavelet shape b_t for $t > 0$ is determined by successive step-by-step predictions from its own past values, where its initial values are $b_t = 0$ for $t < 0$ and b_0 for $t = 0$. The amplitude factors $\theta(t)$ are the normalized prediction errors: $\theta(t) = w(t)/b_0$.

The coefficients k_s are *autoregressive* coefficients of the time-series. The operator (7.97) cannot predict from past values of the time-series the arrival of a new wavelet, and thus a prediction error $b_0\,\theta(t)$ is introduced at the arrival time t of each wavelet. Nevertheless, for times subsequent to the arrival time of the wavelet, the prediction operator can perfectly predict the remaining values of the wavelet, as seen by equation (7.98). Hence the prediction error $b_0\,\theta(t)$ indicates the arrival of a new wavelet at time t with initial amplitude $b_0\,\theta(t)$. For this reason $\theta(t)$ is called the innovation at time t.

The wavelet method provides the estimate of the spectral power function given by

$$[\text{Estimate of } F'(\nu)] = \left| \sum_{s \geqslant 0} b_s \, e^{-2\pi i \nu s} \right|^2$$

where the Wold coefficients (i.e. wavelet shape) are computed from

$$\begin{aligned} b_0 &= \|w(t)\| \\ b_1 &= \|w(t)\| \, k_0 \\ b_2 &= \|w(t)\| \, (k_0^2 + k_1) \\ b_3 &= \|w(t)\| \, (k_0^3 + 2k_0 k_1 + k_2), \ldots \end{aligned}$$

and where the autoregressive coefficients $k_1, k_2, \ldots, k_n$ $(n < N)$ are computed from the observed time-series $x(0), x(1), \ldots, x(N)$ by the Gauss method of least squares.

So far we have confined ourselves to a section of a time-series which we assumed to be stationary. The optimum prediction operator transformed this section into the uncorrelated prediction errors $b_0 \theta(t)$, the mean-square value of which is a minimum. As we have seen, the prediction operator cannot predict from past values of the time-series the initial arrival of a new wavelet, and thus a prediction error, in the form of an impulse $b_0 \theta(t)$, is introduced at the arrival time of each wavelet. Nevertheless, for times subsequent to the arrival time of the wavelet, the prediction operator can perfectly predict the wavelet, and thereby yield zero error of prediction for this wavelet.

Suppose now we apply the prediction operator to another section of the time-series which has dynamics different from the original section. The prediction operator, applied to the new section, will encounter wavelets of a different shape, and consequently the prediction errors will no longer be impulses, but instead will be transient time-functions. As a result the mean-square prediction error will be much greater. This procedure provides a method for the detection of various types of wavelets comprising a non-stationary time-series. In the case of multiple time-series, multiple prediction operators can be used, which take into account coherency relationships. These multiple operators yielded highly successful results in the analysis of seismic records (Wadsworth, Robinson, Bryan, and Hurley, 1953; Robinson, 1953, 1953 a). The high-speed Whirlwind digital computer at the Massachusetts Institute of Technology was used.

EXERCISES

General

0.1 Complete proofs and give examples and counter-examples.

0.2 By use of the references, further develop any of the topics or related material, bringing the treatment up to date and including your own research.

0.3 Develop applications to your field. What rôle do digital computers play?

Chapter I

References: Cramér (1946), Halmos (1950), Kolmogorov and Fomin (1957), Munroe (1953), von Neumann (1933).

1.1 Verify the metric conditions for the distance functions in Examples 1.10, 1.11, and 1.12.

1.2 Show that the set of all n-tuples of real numbers, $x = (u_1, \ldots, u_n)$, $y = (v_1, \ldots, v_n)$, with distance function

$$d(x,y) = \left(\sum_{i=1}^{n} |v_i - u_i|^p\right)^{1/p}, \quad p \geqslant 1$$

forms a metric space.

1.3 Give examples to show that a limit point of a set **S** of a metric space **K** can either belong to **S** or not.

1.4 Find the definition of a topological space, and show that every metric space can be considered as a topological space. What is the Hausdorff axiom? Is every metric space a Hausdorff space?

1.5 What is Lebesgue measure? What is Lebesgue–Stieltjes measure with distribution function F, and how does it differ from Lebesgue measure? (Cramér, 1946, p. 22, 48–56.)

1.6 Write a paper on the Lebesgue integral.

1.7 Find the definition of measure preserving transformation. Differentiate between a one-to-one, and a single-valued, measure preserving transformation.

1.8 Prove the Radon–Nikodym Theorem 1.20, and show that $f(\omega) = d\nu/d\mu$ is non-negative. Show that Theorem 1.20 as stated is still true if ν is a signed-measure (i.e. can assume positive and negative values) where $f(\omega)$ will now also assume positive and negative values.

Chapter II

References: Cramér (1946), Doob (1953), Feller (1950), Gnedenko and Kolmogorov (1954), Halmos (1950), Kendall and Stuart (1958), Kolmogorov (1933), Loève (1955), Savage (1954), Wadsworth and Bryan (1958).

2.1 A die is rolled until an ace (one-spot) comes up. Describe a sample space for the experiment and assign probabilities to the various sample points.

2.2 The number of eggs that an insect lays is a Poisson variate with mean a. Each egg has a probability p of developing, and develops independently of the other eggs. Show that the number of eggs that develop is a Poisson variate with mean ap.

2.3 Three points are taken at random on a circle. Show that the probability that they lie on the same semicircle is 3/4. Assume that the limits of elementary intervals of arc are equally probable.

2.4 Two statistically independent observations, x_1 and x_2, are drawn from a population with a continuous distribution function. Show that $P(x_1 \leqslant x_2) = 1/2$ regardless of the form of the distribution function.

2.5 How many different sequences of three notes can be formed from an eight-note scale if no note is repeated?

2.6 Suppose that 5 men out of 25 are colour-blind and 5 women out of 100 are colour-blind. A colour-blind person is chosen at random. What is the probability of the person being a woman? (Assume men and women to be in equal numbers.)

2.7 A man wishes to unlock his door in the dark, and has a ring of m keys. The correct key can only be found by trying it in the lock. He can select successive keys at random, or by rotation key by key around the ring. Find the probability of success at the nth trial by each method, and so determine which is more efficient.

2.8 A random sample of n statistically independent observations is taken from a population with an unknown continuous distribution function. Prove that the probability that an additional $(n+1)$st observation is greater in magnitude than any of the n already drawn is $1/(n+1)$.

2.9 In the Gregorian calendar, is the probability of Christmas falling on a Sunday equal to 1/7? (Leap years are centesimal years divisible by 400, and non-centesimal years divisible by 4.)

2.10 Two urns identical in appearance contain respectively three white and two black balls; two white and five black balls. One urn is selected and a ball taken from it. What is the probability that this ball is white?

2.11 What is the probability that the birthdays of six people will fall in exactly two calendar months? (Assume equal probabilities for the 12 months.)

2.12 What is the probability of throwing nine heads twice in five throws of ten true coins?

2.13 Three urns contain respectively one white and two black balls; three white and one black ball; two white and three black balls. One ball is taken from each urn. What is the probability that among the balls drawn there are two white and one black?

2.14 Let the radius of the base of a circular cone be x, and the height of the cone be y. If all values of x from 0 to 1 are equally likely, and all values of y from 0 to 2 are equally likely, what is the probability that the volume of the cone will be less than $\pi/3$?

2.15 Find the probabilities of all the poker hands.

2.16 Let n statistically independent observations of a variate x with density function $F'(u)$ be taken at random. Denote the smallest of these n values by y. Find the distribution of y.

2.17 If three cards are drawn from a deck (or complete pack), and x and y denote the number of spades and hearts obtained respectively, find (a) the distribution function of x, and (b) the conditional distribution function of y given x.

2.18 In a sample of n statistically independent observations from a distribution with mean a and variance σ^2, let x_i, x_j denote any pair of observations, and $\bar{x}$ the sample mean. Show that

$$E[(x_i-\bar{x})(x_j-\bar{x})] = \frac{-\sigma^2}{n} \quad (i \neq j)$$

2.19 Review transformations of variates with continuous distributions. These transformations follow the ordinary laws for the transformation of differentials (Kendall and Stuart, 1958). If $F'(u)$ is the density function of the variate $x(\omega) = u$, find the density function of the variate $y(\omega)$ where:

(a) $F'(u) = \dfrac{1}{\pi}\dfrac{1}{1+u^2} \quad (-\infty < u < \infty), \quad y = x^2$

(b) $F'(u)$ as in (a), $\quad y = \tan^{-1} x$

(c) $F'(u) = \dfrac{1}{\sqrt{(2\pi)}} u^{-2} \mathrm{e}^{-1/2u^2} \quad (-\infty < u < \infty), \quad y = 1/x$

2.20 Let x_1 and x_2 be two independent observations of the variate $x(\omega) = u$ with density function $F'(u) = 1$ for $0 \leqslant u \leqslant 1$ and $= 0$ elsewhere. Find the distribution function $G(v)$ of the product $x_1(\omega)x_2(\omega)$. Answer: $G'(v) = v(1-\log v)$ for $0 \leqslant v \leqslant 1$ and $= 0$ elsewhere.

2.21 The variates $x(\omega) = u$ and $y(\omega) = v$ have the joint density function $F'(u, v) = 24v(1-u)$ for $0 \leqslant u \leqslant 1$, $0 \leqslant v \leqslant u$ and $F' = 0$ elsewhere. Find the density function $H'(t)$ of the quotient $y(\omega)/x(\omega)$. Answer: $H'(t) = 2t$ for $0 \leqslant t \leqslant 1$ and $= 0$ elsewhere.

2.22 If x is a variate with density function $F'(u)$, and $y = F(x)$, what is the distribution function of: (a) y; (b) z, where $z = y^2$?

2.23 For an observed value u_i of the test variate x with density function F', let the measure of significance S_i be the area under F' to the right of u_i. If n statistically independent tests yield significance measures $S_1, S_2, \ldots, S_n$, let the combined significance measure P be the probability of obtaining a set of n observed values for which $S_1 S_2 \ldots S_n \leqslant \lambda$. Show that

$$P = \lambda\left[1 - \log\lambda + \frac{(\log\lambda)^2}{2!} - \frac{(\log\lambda)^3}{3!} + \ldots + (-1)^{n-1}\frac{(\log\lambda)^{n-1}}{(n-1)!}\right]$$

If we let $A = -2\log\lambda$, show that P is equal to the area (from A to ∞) under the chi-square density function for $2n$ degrees of freedom.

2.24 In Exercise 2.23, let two tests be made. If $S_1 = 0{\cdot}1$, how small must S_2 be in order that the combined results be significant at the 5 per cent level? Use a chi-square table.

2.25 Find the density function $H'(u)$ of the quotient of two independent normal variates, each with zero mean and unit variance. Answer: $H'(u) = 1/\pi(1+u^2)$, $-\infty < u < \infty$.

2.26 Statistically independent variates x_1 and x_2 are distributed as chi-square with $2a$ and $2b$ degrees of freedom respectively. Find the joint density function of y and z, and the single density function of z where $y = x_1 + x_2$ and $z = x_1/(x_1+x_2)$.

2.27 Derive the Student t distribution and the F distribution.

2.28 Two statistically independent observations x_1 and x_2 are drawn from a Cauchy distribution. Find the distribution of their average $(x_1+x_2)/2$.

2.29 Summarize the main properties of beta and gamma functions and how they relate to the beta and gamma distributions.

2.30 If x_1 and x_2 are independent exponential variates with means a_1 and a_2 respectively, find the distribution of $x_1 - x_2$ if (1) $a_1 \neq a_2$ and (2) if $a_1 = a_2$.

2.31 Accidents in a certain industry are found to follow a Poisson distribution with a mean of four accidents per week. (a) Find the median time between accidents; (b) Find the density function of the time required for ten accidents to occur.

2.32 Show that the variance of a sum of statistically independent variates is equal to the sum of the variances of the individual variates.

2.33 Prove that in order to obtain the lower critical limit for an F distribution, we may take the reciprocal of the upper critical limit of the "F" with interchanged degrees of freedom.

2.34 Find the definitions of convergence in probability (or convergence in measure), convergence with probability 1 (or almost certain convergence), and mean-square convergence (or convergence in the mean). Give examples of each. Show that convergence with probability 1 and mean-square convergence each imply convergence in probability (but not conversely).

2.35 If a fair coin comes up heads, a red ball is placed in an urn, but if tails, a white ball. This is done seven times. Afterwards, three balls are drawn and they turn out to be white. What is the probability that there were three white and four red balls in the urn ?

2.36 If x is a variate with density function $F'(u)$, find the density function of: (a) $|x|$, (b) x^2.

2.37 If x is a non-negative variate with a skewed density function, then the density function of which of the following tends to be more nearly symmetrical than the density function of x ? (a) $x^{1/2}$; (b) x^3; (c) e^x.

2.38 Let an urn contain N balls, of which N_1 are white and N_2 are black ($N = N_1+N_2$). Draw a random sample (without replacements) of n balls, where n_1 denotes the number of white balls, and n_2 the number of black balls, in the sample ($n = n_1+n_2$). The balls may be drawn successively or simultaneously. Describe a suitable sample space, and by assuming all balls have an equal chance, assign probabilities to the sample points. If the variate x is the number of white balls in the sample, show that

$$P(x = n_1) = \frac{\binom{N_1}{n_1}\binom{N_2}{n_2}}{\binom{N}{n}}, \quad n_1 = 0, 1, 2, \ldots, \min(N_1, n)$$

This is the *hypergeometric distribution.*

2.39 Using Exercise 2.38, find the probability that a poker hand of five cards will contain two kings.

Chapter III

References: Bochner (1955), Doob (1953), Gnedenko and Kolmogorov (1954), Kendall and Stuart (1958), Khintchine (1933, 1949), Lévy (1937, 1948), Loève (1955).

3.1 What is the relationship between composition of distribution functions and multiplication of characteristic functions ?

3.2 Find the definition of a non-negative definite function. Prove the Herglotz–Bochner theorem: A function of a real variable is non-negative definite and continuous if and only if it is a characteristic function.

3.3 Derive the inversion formula (3.15) and prove its uniqueness.

3.4 What is the relationship between the characteristic function and the moments of the distribution?

3.5 Find the characteristic function of the binomial, Pascal, rectangular, exponential, gamma, Cauchy, and Laplace distributions.

3.6 What is the relationship between the characteristic function and the moment-generating function? Does the latter always exist?

3.7 Prove the Lévy continuity theorem 3.21.

3.8 Suppose that the characteristic function $f(t)$ is infinitely divisible. Show that the characteristic functions $f_{n\cdot}(t) \Rightarrow 1$ as $n \to \infty$, and so the common distribution function $F_{n\cdot}(u)$ converges weakly to the distribution degenerate at zero.

3.9 Show that the characteristic function of an infinitely divisible law never vanishes.

3.10 Show that the characteristic function

$$f(t) = (1-b)(1-b\,\mathrm{e}^{it})^{-1}, \quad 0<b<1$$

is infinitely divisible.

3.11 Verify the continuity equation (3.41).

3.12 Show that Kolmogorov's formula (3.40) is unique.

3.13 Show that a gamma variate (Example 2.44) is infinitely divisible, with Kolmogorov formula

$$\log f(t) = \frac{ikt}{a} + k\int_0^\infty (\mathrm{e}^{itu}-1-itu)\frac{\mathrm{e}^{-au}}{u}\,du$$

and partial variance

$$K(u) = k\int_0^u v\,\mathrm{e}^{-av}\,dv$$

3.14 In a large sample of size n from a binomial population with known probability p of success on an individual trial, denote the observed number of successes by O_1, the expected number of successes by E_1, the observed number of failures by O_2, and the expected number of failures by E_2. Show that the statistic

$$\sum_{i=1}^{2}\frac{(O_i-E_i)^2}{E_i}$$

is distributed approximately as chi-square with one degree of freedom. State the generalization of this theorem to the case of sampling from a known multinomial population, and discuss how it may be used to test *goodness of fit.*

3.15 Is the distribution of the sum of two statistically independent variates drawn from a rectangular distribution (a) gamma, (b) normal, (c) rectangular, (d) triangular?

3.16 State and prove Tchebycheff's inequality.

3.17 Is the sum of ten statistically independent variates all from the same exponential distribution (a) an exponential variate, (b) a Poisson variate, (c) a normal variate, (d) a gamma variate?

3.18 If the variate x has density function $f(u)$ and the statistically independent variate y has density function $g(v)$, then the sum $x+y$ has density function: (a) $f(u)+g(v)$, (b) $\frac{1}{2}[f(u)+g(v)]$, (c) $\int f(t-v)g(v)\,dv$, (d) $g(u+v)f(u+v)$?

3.19 Prove: If x is a normal variate with mean a and variance σ^2, then the sample mean $\bar{x}$ of n observations is a normal variate with mean a and variance σ^2/n.

3.20 Using the inequality (where $\delta > 0$)

$$\int_{|v|\geq \epsilon D_n} v^2\, dF_j(v+E(y_j)) \leq \frac{1}{\epsilon^\delta D_n^\delta}\int_{-\infty}^{\infty} |v|^{2+\delta}\, dF_j(v+E(y_j))$$

in Theorem 3.71 obtain Ljapounov's Central Limit Theorem.

Chapter IV

References: Besicovitch (1932), Birkhoff and MacLane (1948), Cooke (1950, 1953), Halmos (1951, 1958), Hamburger and Grimshaw (1956), Riesz and Sz.-Nagy (1952), Stone (1932), von Neumann (1929, 1932, 1933), Weyl (1928).

4.1 Give examples of linear spaces.

4.2 A linear space is *normed* if to each element x there corresponds a non-negative number $\|x\|$ such that (1) $\|x\| = 0$ if and only if $x = 0$, (2) $\|ax\| = |a|\|x\|$, (3) $\|x+y\| \leq \|x\|+\|y\|$. A complete normed space is called a *Banach space*. For what distance function is a Banach space a metric space? Show that von Neumann space is a Banach space. Give other examples of Banach space.

4.3 Prove the Schwarz inequality: If x and y are elements of von Neumann space $\mathbf{V}$, then $|((x,y))| \leq \|x\|\|y\|$.

4.4 Establish subadditivity: If x and y are elements of von Neumann space $\mathbf{V}$, then $\|x+y\| \leq \|x\|+\|y\|$.

4.5 Show that $((x,by)) = \bar{b}((x,y))$.

4.6 Show that a space satisfying Postulates P_1, P_2, P_4', necessarily satisfies Postulates P_3 and P_5, and so is a unitary space.

4.7 Show that Bohr's *almost periodic functions* as generalized by Besicovitch (1932), satisfy Postulates P_1, P_2, P_3, and P_4, but not P_5.

4.8 In Euclidean space (i.e. unitary space over the real numbers $a, b, \ldots$ instead of the complex numbers) is a point a linear manifold? Is a line? Is a plane? Must they pass through the origin (i.e. null element) to be linear manifolds?

4.9 Show that every linear manifold in finite-dimensional unitary space is closed, and hence we cannot obtain an intuitive-geometrical picture of a non-closed linear manifold.

4.10 Let $\theta_1, \theta_2, \ldots$ be an orthonormal set in Hilbert space. Then $a_1\theta_1 + \ldots + a_n\theta_n$ ($n = 1, 2, \ldots$; $a_1, \ldots, a_n$ arbitrary) form a linear manifold. It is not closed because

$$\sum_1^\infty \frac{1}{n}\theta_n$$

is a limit point, but not an element of the manifold. Exhibit other non-closed linear manifolds.

Chapter V

References: Same as Chapter IV.

5.1 There are two geometrical interpretations of a unitary operator. First, it may be interpreted as the replacement of every element by another of equal length in such a way that angles between elements are unchanged. This process may be described as a generalized rotation of the space about the origin (i.e. null element). Secondly, it may be interpreted as the substitution for the co-ordinates of every element in one orthonormal system in terms of new co-ordinates in a new system. This process may be described as a transformation of the co-ordinate system.

5.2 Show that $\mathbf{L}^2$ forms a complete metric space.

5.3 Find the definition of class $\mathbf{L}^1$ and show that it forms a complete metric space.

5.4 Write a paper on the connection between linear operators and infinite matrices in Hilbert space.

5.5 Develop the properties of the stochastic integral. Find applications in the theory of stochastic processes; for example, in the theory of Brownian motion.

5.6 If $E(\lambda)$ is a resolution of the identity, show that

$$E(\lambda)E(\mu) = E(\min\{\lambda, \mu\})$$

5.7 Prove that $\|x-y\|^2 + \|x+y\|^2 = 2(\|x\|^2 + \|y\|^2)$ in Hilbert space.

5.8 Show that the following are orthonormal sets in $\mathbf{L}^2$ (Lebesgue measure):

(a) $e^{2\pi i n t}$ with configuration space Ω the interval $-0{\cdot}5 \leqslant t \leqslant 0{\cdot}5$.

(b) $\sqrt{\frac{2n+1}{2}}\,P_n(t)$ with $\Omega = \{-1 \leqslant t \leqslant 1\}$,

where $P_n(t)$ is the Legendre polynomial of order n.

(c) $\dfrac{\sqrt{\{2t\}}\,J_p(r_n t)}{J_{p+1}(r_n)}$ with $\Omega = \{0 \leqslant t \leqslant 1\}$,

where r_n is the nth positive zero of the Bessel function J_p.

Chapter VI

References: Same as Chapter IV.

6.1 Derive the normal equations (Section 6.1).

6.2 Show that every eigenvalue of a Hermitian operator is real.

6.3 Show that solutions of the characteristic equation corresponding to distinct eigenvalues are orthogonal.

6.4 Derive von Neumann's celebrated result showing the basic identity of the Heisenberg–Born–Jordan matrix-mechanics form of quantum mechanics and the Schroedinger wave-mechanics form of quantum mechanics by use of abstract Hilbert space (von Neumann, 1932).

6.5 Let the Hilbert space be the $\mathbf{L}^2$ (Lebesgue measure) space of functions $x(t)$ defined on the real line, $-\infty < t < \infty$. The *Hermite* operator H is defined by $xH \equiv -x'' + t^2 x$ on $-\infty < t < \infty$, where the prime indicates differentiation with respect to t. Show that the eigensolutions are, except for constant factors, $e^{-\frac{1}{2}t^2} H_n(t)$, where the $H_n(t)$ are the *Hermite polynomials*. Show that the eigenvalues are odd positive integers. What relationship do these polynomials have to the normal distribution ?

6.6 Let the Hilbert space be the $\mathbf{L}^2$ space of Exercise 6.5. Show that the operator defined by $xH \equiv -x''$ on $-\infty < t < \infty$ has the set $0 \leqslant \lambda < \infty$ as spectrum. (Note: the spectrum is the set of non-constancy points of the resolution of the identity $K(\lambda)$ belonging to H.)

Chapter VII

References: See REFERENCES.

7.1 Define the strictly stationary stochastic process. How does it differ from a second-order stationary process ?

7.2 Specialize the equations of Chapter VII for the cases: when $x(t)$ is real; when $x(t)$ has an absolutely continuous spectral distribution function.

7.3 State and prove Theorems 6.1 and 6.2 of Stone (1932).

7.4 Prove the ergodic theorem for strictly stationary processes (Doob, 1953).

7.5 If equation (7.20) is the counterpart of equation (7.22), find the equation for $x(0)E(\nu)$ that is the counterpart of $F(\nu) = W(\nu) + C$.

7.6 Show that $|\Delta F_{12}|^2 \leqslant \Delta F_{11}\Delta F_{22}$. Discuss its intuitive meaning when $<$ holds and when $=$ holds, and relate to the coherency between two time-series (Wiener, 1930; Cramér, 1940).

7.7 If $x(t)$ is (1) a running average of white noise $\theta(t)$, and is also (2) equivalent to $\theta(t)$, then does S_x have Lebesgue measure 1 ?

7.8 If $x(t)$ is (1) a running average of white noise $\theta(t)$, and also (2) S_x has Lebesgue measure 1, then is $x(t)$ equivalent to $\theta(t)$?

7.9 If $F_{xx}(\nu)$ is absolutely continuous for all ν $(-0{\cdot}5 \leqslant \nu \leqslant 0{\cdot}5)$ and S_x has Lebesgue measure 0·5, then is $x(t)$ singular, non-singular, regular, or non-regular ?

7.10 If $F'_{xx}(\nu) = 0$ only on a set of Lebesgue measure zero, then is $x(t)$ singular, non-singular, regular, or non-regular, or can no statement be made ?

7.11 If $F'(\nu) = 0$ on a ν-set of positive Lebesgue measure, is $x(t)$ singular, non-singular, regular, or non-regular ?

7.12 What is the transfer function of the optimum linear predictor for the time-series of Exercise 7.9; of Exercise 7.10; of Exercise 7.11 ? Give physical interpretations of these time-series.

7.13 Let $F_{xx}(\nu)$ be absolutely continuous with respect to Lebesgue measure and let S_x have Lebesgue measure 1. Show that $\theta(t)$, subordinated to $x(t)$, is white noise if and only if, for ν in S_x,

$$|\Phi_\theta^{(x)}(\nu)|^2 = \frac{1}{F'_{xx}(\nu)}$$

7.14 In the Wold decomposition show that $\mathbf{Y}(t) = \mathbf{X}(-\infty)$ and $\mathbf{Z}(t) = \mathbf{\Theta}(t)$, and so establish Theorem 7.71.

7.15 In the running average representation (7.60), verify and interpret $b_s = R_{x\theta}(s)$.

7.16 Why is $F'(\nu)$, $\nu \epsilon S_3$, called the spectral *power* function ? (Wiener, 1930, 1942.)

7.17 Show that the integral in condition (3) of Theorem 7.78 is $<\infty$ for any stationary time-series $x(t)$. (Hint: For $F' > 1$, $\log F' \leqslant F'$.)

7.18 Does the spectral power $dF(\nu)$ have to be concentrated in a ν-set of Lebesgue measure zero in order for $x(t)$ to be singular? If not, then why does the singular component of the Wold decomposition have its spectral power concentrated on a ν-set of Lebesgue measure zero?

7.19 Find the optimum linear predictor for the time-series $x(t)$ of Example 7.70. Describe the physical meaning of condition (3) of Theorem 7.78, and hence give intuitive reasons why $x(t)$ of Example 7.70 is singular.

7.20 In terms of electric filter theory, why cannot a function and its Fourier transform be simultaneously very small at infinity? (Wiener, 1942.)

7.21 Let $x(t)$ be a regular time-series. Interpret its Wold decomposition as a filter with input $\theta(t)$, transfer function $B(\nu)$, and output $x(t)$. If we call $\log|B(\nu)|$ the attenuation of the filter, then condition (3) of Theorem 7.78 is a condition on the attenuation, and is called the *Paley–Wiener criterion.* Show that the filter cannot have infinite attenuation in any finite frequency band.

7.22 Let equation (7.60) represent a filter. Show that if any $b_r \neq 0$ for $r<0$, then its output $x(t)$ at time t is affected by a future (i.e. subsequent to time t) input $\theta(t-r)$, which is physically impossible. Show that this difficulty may be overcome by introducing a sufficiently large time delay into the network; i.e. if b_{-m} where m is positive is the first non-vanishing coefficient, then equation (7.60) can be computed at the delayed time $t+m$, but no sooner. What application does this have to numerical smoothing methods?

7.23 If the Yule process

$$\epsilon(t) = \sum_{0}^{n} a_s x(t-s)$$

is white noise, then $x(t)$ is called an *autoregressive process of order n.* In particular, an autoregressive process of order 1 is a *Markov process.* For an autoregressive process, find the autocovariance, spectral distribution function, optimum linear predictor, and mean-square prediction error, in terms of the constant coefficients $a_0, a_1, \ldots, a_n$. Is an autoregressive process regular? What stability restriction must be made on the constant coefficients of an autoregressive process? (Wold, 1938.)

7.24 Let $x(t)$ have the Wold decomposition $x(t) = y(t)+z(t)$ where $y(t)$ is singular with spectrum S_y the single point $\nu = 0$, and where $z(t)$ is a Markov time-series with $a_0 = 1, |a_1| < 1$. By use of Gibbs' ergodic theorem, show that, as $n \to \infty$, the Yule process

$$-a_1 x(t) + \frac{1+a_1}{n}[x(t)+x(t-1)+\ldots+x(t+n-1)] \Rightarrow \hat{x}(t+1)$$

(Hint: Since 1 is an eigenvalue of U, we have the characteristic equation $y(t+1) = y(t)U = 1.y(t)$, so $y(t)$ is a constant.)

7.25 Let $x(t)$ have the Wold decomposition $x(t) = y(t) + z(t)$ where S_y has a finite number of points and $z(t)$ is an autoregressive process of order n. Find the Yule process in $\mathbf{X}(t)$ that converges ($\Rightarrow$) to $\hat{x}(t+1)$; to $\hat{x}(t+\alpha)$.

7.26 Find the counterparts of equations (7.88) and (7.89) if $x(t)$ is non-singular.

7.27 If $A_1(\nu)$ and $A_2(\nu)$ are the transfer functions of two Yule processes, describe the process with transfer function $A_1(\nu)A_2(\nu)$. Extend your results to processes that are limits ($\Rightarrow$) of Yule processes as $n \to \infty$.

7.28 Show that the transfer function (7.85) for the optimum predictor ($\nu \epsilon S_z$) is

$$\frac{1}{B(\nu)} \sum_{s=0}^{\infty} e^{-2\pi i \nu s} \int_{-0\cdot 5}^{0\cdot 5} B(\sigma)\, e^{2\pi i \sigma (s+\alpha)}\, d\sigma$$

Show that this transfer function has minimum phase characteristic in the sense of Bode. (Bode and Shannon, 1950.)

7.29 Show that the transfer function (7.89) for the optimum (message from message plus noise) filter is

$$\frac{1}{B(\nu)} \sum_{s=0}^{\infty} e^{-2\pi i \nu s} \int_{-0\cdot 5}^{0\cdot 5} \frac{F'_{mm}(\sigma) + F'_{nm}(\sigma)}{B(-\sigma)}\, e^{2\pi i \sigma (s+\alpha)}\, d\sigma$$

Show that this transfer function has minimum phase characteristic in the sense of Bode.

7.30 Corresponding to equation (7.86) find the filtering error $m(t+\alpha) - \hat{m}(t+\alpha)$. Divide the mean-square filtering error into two components, one depending on α and the other not depending on α (Wiener, 1948).

7.31 Let $x(t)$, $-\infty < t < \infty$, be a stationary stochastic process for which the time t is continuous. The continuity condition $\|x(t) - x(0)\| \to 0$ as $t \to 0$ is imposed. Because of this continuity condition, the closed linear manifold generated by $x(s)$, $r < s \leqslant t$ is the closed linear manifold generated by $x(s_1), x(s_2), \ldots$ where $s_1, s_2, \ldots$ is any sequence everywhere dense in the interval $r < s \leqslant t$. Hence the time-domain of $x(t)$ is a Hilbert space, and the prediction and filtering theory of stationary time-series with continuous time t reduces to that of stationary time-series with discrete time t (Wiener, 1942; Doob, 1953). In practice this means that any physical time-series (in which time is continuous) may be transformed into a discrete sequence (i.e. a time-series with integer times) without losing essential information, provided we make the time-spacing fine enough. Discuss applications to high-speed digital computers.

7.32 Relate Exercise 7.31 to *Shannon's sampling theorem* (Shannon, 1949). Discuss the rôle of stochastic processes in *Shannon's information theory* (Shannon, 1948; McMillan, 1953; Khintchine, 1957).

7.33 Let $x(t) = \sum_0^\infty c_s\,\theta(t-s)$ be a real time-series where

$$\sum_0^\infty c_s^2 < \infty, \quad c_0 > 0$$

and $\theta(t)$ is white noise. Assume that $\sum_0^\infty c_s \xi^s$ has zeros for $|\xi| < 1$. Show that $x(t)$ is regular, and hence has a Wold decomposition, with coefficients $b_0, b_1, \ldots$. Show that $b_0 > c_0$ (Doob, 1953). Give applications to wavelet theory (Robinson, 1957).

7.34 Show that the Yule process $\epsilon(t)$ is equal to $x(t)A$ where the linear operator A is $\sum_0^n a_s U^{-s}$.

7.35 Let $x(t)$ be a regular time-series. Use $F'_{xx}(\nu) = |B(\nu)|^2$ to show that the elements $e^{2\pi i\nu t}/B(\nu)$, $-\infty < t < \infty$, form an orthonormal set in $\mathbf{L}^2(F_{xx})$.

7.36 Expand upon the following. The elements $e^{2\pi i\nu t}$, $-\infty < t < \infty$, form an orthonormal set in function space $\mathbf{L}^2(0{\cdot}5+\nu)$ with Lebesgue measure $d\nu$. Let $\sum_0^\infty b_s \xi^s$ be a power series that is analytic for $|\xi| < 1$ and has no zeros for $|\xi| < 1$. Substituting $\xi = r\,e^{-2\pi i\nu}$ (where $r < 1$), the power series becomes $\sum_0^\infty b_s r\,e^{-2\pi i\nu s}$. Since the $e^{2\pi i\nu t}$ are orthonormal, we have

$$\int_{-0\cdot5}^{0\cdot5} \left(\sum_0^\infty b_s r\,e^{-2\pi i\nu s}\right) e^{2\pi i\nu t}\,d\nu = b_t r$$

Since $b_t r \to b_t$ as $r \to 1$ for each t, we have

$$\sum_0^\infty b_t r\,e^{-2\pi i\nu t} \to \sum_0^\infty b_t\,e^{-2\pi i\nu t} \equiv B(\nu)$$

That is, the power series has the limiting value $B(\nu)$ on the unit circle.

7.37 Expand upon the following. The elements $e^{2\pi i\nu t}$, $-\infty < t < \infty$, do *not* form an orthonormal set in $\mathbf{L}^2(F_{xx})$ (see Exercise 7.35). Let $\sum_0^\infty a_s \xi^s$ be a power series that is analytic for $|\xi| < 1$ and has no zeros for $|\xi| < 1$. The power series may be written

$\sum_{0}^{\infty} a_s r \, e^{-2\pi i \nu s}$ (where $r < 1$). Since the $e^{2\pi i \nu t}$ are not orthonormal, we have

$$\int_{-0.5}^{0.5} \left(\sum_{0}^{\infty} a_s r \, e^{-2\pi i \nu s} \right) e^{2\pi i \nu t} \, dF_{xx}(\nu) \neq a_t r$$

Hence the limiting value of the power series in general cannot and should not be written $\sum_{0}^{\infty} a_s \, e^{-2\pi i \nu s}$.

7.38 Let $x(t)$ be a regular time-series with Wold coefficients $b_0, b_1, \ldots$. Then the power series $\sum_{0}^{\infty} b_s \xi^s$ has no zeros or singularities for $|\xi| < 1$. Hence the function $1 \big/ \sum_{0}^{\infty} b_s \xi^s$ has no zeros or singularities for $|\xi| < 1$, and therefore can be expanded in a power series $\sum_{0}^{\infty} a_s \xi^s$ for $|\xi| < 1$. Show that the a_s and b_s are related by

$$a_0 b_0 = 1; \quad \sum_{0}^{n} a_{n-s} b_s = 0, \quad n = 1, 2, 3, \ldots$$

7.39 Let $x(t)$ be a regular time-series with Wold coefficients $b_0, b_1, \ldots$. Using the Exercise 7.37, show that the next sentence is *not* generally correct, and hence the conclusion in the following sentence is *not* valid. The power series $\sum_{0}^{\infty} a_s \xi^s$ of Exercise 7.38 has the limiting value $\sum_{0}^{\infty} a_s e^{-2\pi i \nu s}$ on the unit circle. Since $x(t) \leftrightarrow e^{2\pi i \nu t}$ in the isomorphism between $\mathbf{X}$ and $\mathbf{L}^2(F_{xx})$, we conclude that $\sum_{0}^{\infty} a_s x(-s) \leftrightarrow \sum_{0}^{\infty} a_s e^{-2\pi i \nu s}$, and so $x(t)$ has the infinite autoregressive representation $\theta(t) = \sum_{0}^{\infty} a_s x(t-s)$.

7.40 Let $x(t)$ be a regular time-series with Wold coefficients $b_0, b_1, \ldots$. From the Wold coefficients, define the coefficients $a_0, a_1, \ldots$ by means of the equations at the end of Exercise 7.38. Show that the following argument is *not* valid. We have

$$\begin{aligned} x(t) &= x(t) + 0x(t-1) + 0x(t-2) + \ldots \\ &= \sum_{n=0}^{\infty} \left(\sum_{s=0}^{n} a_{n-s} b_s \right) x(t-n) = \sum_{s=0}^{\infty} b_s \sum_{r=0}^{\infty} a_r x(t-s-r) \end{aligned}$$

which comparing with the Wold decomposition (7.75), we see that

$$\theta(t) = \sum_{r=0}^{\infty} a_r x(t-r)$$

so $x(t)$ has an infinite autoregressive representation. (Hint: In $\mathbf{X}$ the $\theta(t)$ are orthonormal elements. If $\Sigma b_s \theta(s) = \Sigma c_s \theta(s)$, then we have $((\Sigma b_s \theta(s), \theta(t))) = ((\Sigma c_s \theta(s), \theta(t)))$, which is $b_t = c_t$. But the $x(t)$ are *not* orthonormal elements in $\mathbf{X}$.)

7.41 For finite sections of time-series, Bartlett (1946, 1950), Daniell (1946, in Discussion, pp. 85–97, *J. Roy. Statist. Soc.*, **B, 8**), Kendall (1946), Tukey (1949), and Tukey and Hamming (1949) obtained results on the effects of sampling and the choice of computational procedure on the behaviour of the estimated spectral power. Discuss estimates of spectral power, and their applicability to physical time-series. Include more recent developments (Bartlett, 1955; Grenander and Rosenblatt, 1957; Blackman and Tukey, 1958; and references given in these works).

7.42 The applicability of time-series methods to physical problems depends chiefly upon having adequate statistical estimation and testing procedures. Study the contributions to this field (Bartlett, 1946, 1955; Champernowne, 1948; Grenander, 1950; Kendall, 1945, 1946; Kendall, 1955—see under Kendall and Stuart; Quenouille, 1947, 1957; Whittle, 1951, 1953, 1954; Wold, 1953; and references given in these works).

REFERENCES

Bartlett, M. S. (1946). On the theoretical specification and sampling properties of autocorrelated time-series. *J. Roy. Statist. Soc.*, **B**, **8**, 27.

Bartlett, M. S. (1950). Periodogram analysis and continuous spectra. *Biometrika*, **37**, 1.

Bartlett, M. S. (1955). *Stochastic Processes*. Cambridge Univ. Press.

Bawly, G. M. (1936). Über einige Verallgemeinerungen der Grenzwertsätze der Wahrscheinlichkeitsrechnung. *Rec. Math.* (*Mat. Sbornik*) *N.S.*1, **43**, 917.

Besicovitch, A. S. (1932). *Almost Periodic Functions*. Cambridge Univ. Press.

Birkoff, G. and MacLane, S. (1948). *Modern Algebra, N.Y.*

Blackman, R. B., Bode, H. W., and Shannon, C. E. (1946). *Monograph on Data Smoothing and Prediction in Fire Control Systems*. NDRC Report.

Blackman, R. B. and Tukey, J. W. (1958). The measurement of power spectra from the point of view of communications engineering. *Bell Sys. Tech. J.*, **37**, 185.

Bochner, S. (1955). *Harmonic Analysis and the Theory of Probability*. U. of Cal. Press, Berkeley.

Bode, H. W. and Shannon, C. E. (1950). A simplified derivation of linear least square smoothing and prediction theory. *Proc. Inst. Rad. Eng.*, **38**, 417.

Champernowne, D. G. (1948). Sampling theory applied to autoregressive schemes. *J. Roy. Statist. Soc.*, **B**, **10**, 204.

Cooke, R. G. (1950). *Infinite Matrices and Sequence Spaces*. Macmillan, London.

Cooke, R. G. (1953). *Linear Operators*. Macmillan, London.

Cramér, H. (1940). On the theory of stationary random processes. *Ann. Math.*, **41**, 215.

Cramér, H. (1942). On harmonic analysis in certain functional spaces. *Ark. Mat. Astr. Fys.*, **28B**, No. 12.

Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton Univ. Press.

Doob, J. L. (1953). *Stochastic Processes*. Wiley, N.Y.

Feller, W. (1950). *Probability Theory*. Wiley, N.Y.

Fischer, E. (1907). Sur la convergence en moyenne. *Comp. Rend. Acad. Sci. Paris*, **144**, 1022 and 1148.

Fréchet, M. (1937). *Généralités sur les Probabilités. Variables Aléatoires*. Paris.

Gibbs, W. (1902). *Elementary Principles of Statistical Mechanics*. Yale Univ. Press.

Gnedenko, B. V. (1939). On the theory of limit theorems for sums of independent random variables (Russian). *Izvestiya Akad. Nauk. S.S.S.R., Ser. Mat.*, 181 and 643.

Gnedenko, B. V. (1944). Limit theorems for sums of independent random variables. (Russian). *Uspekhi Mat. Nauk.*, **10**, 115. (English translation, No. 45, Am. Math. Soc., N.Y.)

Gnedenko, B. V. and Kolmogorov, A. N. (1954). *Limit Distributions for Sums of Independent Random Variables* (Russian). English trans., Addison-Wesley, Cambridge, Mass.

Grenander, U. (1950). Stochastic processes and statistical inference. *Ark. Mat.*, **1**, 195.

Grenander, U. and Rosenblatt, M. (1957). *Statistical Analysis of Stationary Time Series*. Wiley, N.Y.

Halmos, P. R. (1950). *Measure Theory*. Van Nostrand, Princeton.

Halmos, P. R. (1951). *Hilbert Space*. Chelsea, N.Y.

Halmos, P. R. (1958). *Finite-Dimensional Vector Spaces*. Van Nostrand, Princeton.

Hamburger, H. L. and Grimshaw, M. E. (1956). *Linear Transformations*. Cambridge Univ. Press.

Hilbert, D. (1912). *Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen*. Leipzig.

Hopf, E. (1937). *Ergodentheorie*. Springer, Berlin (also Chelsea, N.Y.).

Hurewicz, W. (1947). Filters and servo systems with pulsed data. *M.I.T. Rad. Lab. Ser.*, **25**, 231.

Kendall, M. G. (1945). On the analysis of oscillatory time-series. *J. Roy. Statist. Soc.*, **108**, 93.

Kendall, M. G. (1946). *Contributions to the Study of Oscillatory Time-series*. Cambridge Univ. Press.

KENDALL, M. G. (1957). *A Course in Multivariate Analysis.* Griffin's Statistical Monographs and Courses, **2.** Charles Griffin and Co., London.

KENDALL, M. G. and STUART, A. (1958). *The Advanced Theory of Statistics*, vol. I. Vol. II (M. G. KENDALL), 1955 ; revision by Kendall and Stuart in preparation. Charles Griffin and Co., London.

KHINTCHINE, A. JA. (1933). *Asymptotische Gesetze der Wahrscheinlichkeitsrechnung.* Springer, Berlin (also Chelsea, N.Y.).

KHINTCHINE, A. JA. (1934). Korrelationstheorie der stationären stochastischen Prozesse. *Math. Ann.*, **109**, 604.

KHINTCHINE, A. JA. (1937). Zur Theorie der unbeschränkt teilbaren Verteilungsgesetze. *Rec. Math.* (*Mat. Sbornik*) *N.S.*2, **44**, 79.

KHINTCHINE, A. JA. (1948). *Mathematical Foundations of Statistical Mechanics* (Russian). English trans., Dover, N.Y.

KHINTCHINE, A. JA. (1957). *Mathematical Foundations of Information Theory* (Russian). English trans., Dover, N.Y.

KOLMOGOROV, A. N. (1932). Sulla forma generale di un processo stocastico omogeneo. *R.C. Acad. Lincei.*, (**6**), **15**, 805 and 866.

KOLMOGOROV, A. N. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung.* Springer, Berlin. (English trans., Chelsea, N.Y.)

KOLMOGOROV, A. N. (1939). Sur L'interpolation et extrapolation des suites stationnaires. *C.R. Acad. Sci. Paris*, **208**, 2043.

KOLMOGOROV, A. N. (1941). Stationary sequences in Hilbert space. (Russian). *Bull. Moscow State Univ., Math.*, **2.** (Spanish trans., *Trab. Estad.*, **4**, 55 and 243.)

KOLMOGOROV, A. N. (1941a). Interpolation und Extrapolation von stationären zufälligen Folgen (Russian with German summary). *Bull. Acad. Sci. U.R.S.S., Ser. Math.*, **5**, 3.

KOLMOGOROV, A. N. and FOMIN, S. V. (1957). *Functional Analysis* (Russian). English trans., Graylock, Rochester, N.Y.

KOOPMAN, B. O. (1931). Hamiltonian systems and linear transformations in Hilbert space. *Proc. Nat. Acad. U.S.*, **17**, 315.

LÉVY, P. (1937). *Théorie de L'Addition des Variables Aléatoires.* Gauthier-Villars, Paris.

LÉVY, P. (1948). *Processus Stochastiques et Mouvement Brownien.* Paris.

LOÈVE, M. (1955). *Probability Theory.* Van Nostrand, Princeton, N.J.

MCMILLAN, B. (1953). The basic theorems of information theory. *Ann. Math. Statist.*, **24**, 196.

MOYAL, J. E. (1949). Stochastic processes and statistical physics. *J. Roy. Statist. Soc.*, **B**, **11**, 150.

MUNROE, M. E. (1953). *Measure and Integration.* Addison-Wesley, Cambridge, Mass.

QUENOUILLE, M. H. (1947). A large-sample test for the goodness-of-fit of autoregressive schemes. *J. Roy. Statist. Soc.*, **110**, 123.

QUENOUILLE, M. H. (1957). *The Analysis of Multiple Time-Series.* Griffin's Statistical Monographs and Courses, **1.** Charles Griffin and Co., London.

QUENOUILLE, M. H. (1958). *Fundamentals of Statistical Reasoning.* Griffin's Statistical Monographs and Courses, **3.** Charles Griffin and Co., London.

RIESZ, F. (1907). Sur les systèmes orthogonaux de fonctions. *C.R. Acad. Sci. Paris*, **144**, 615, 734.

RIESZ, F. and SZ.-NAGY, B. (1952). *Leçons d'Analyse Fonctionelle.* English trans., F. Ungar, N.Y.

ROBINSON, E. A. (1953). *Linear Operator Study of a Seismic Profile.* M.I.T. Geophysical Analysis Group, **4**, Cambridge, Mass.

ROBINSON, E. A. (1953a). *On the Theory and Practice of Linear Operators in Seismic Analysis.* M.I.T. Geophys. Anal. Group, **5**, Cambridge, Mass.

ROBINSON, E. A. (1954). *Predictive Decomposition of Time Series with Applications to Seismic Exploration.* M.I.T. Geophys. Anal. Group, **7**, Cambridge, Mass.

ROBINSON, E. A. (1957). Predictive decomposition of seismic traces. *Geophysics*, **22**, 767.

ROBINSON, E. A., SIMPSON, S. M., and SMITH, M. K. (1954). *Further Research on Linear Operators in Seismic Analysis.* M.I.T. Geophys. Anal. Group, **6**, Cambridge, Mass.

SAVAGE, L. J. (1954). *The Foundations of Statistics.* Wiley, N.Y.

SHANNON, C. E. (1948). A mathematical theory of communication. *Bell Sys. Tech. J.*, **27**, 379 and 623.

SHANNON, C. E. (1949). Communication in the presence of noise. *Proc. Inst. Rad. Eng.*, **37**, 10.

SIMPSON, S. M. (1955). *Linear Operators and Seismic Noise*. M.I.T. Geophys. Anal. Group, **9**, Cambridge, Mass.

SLUTSKY, E. (1927). The summation of random causes as the source of cyclic processes (Russian with English summary). *Prob. Ec. Cond.*, Inst. Ec. Conj. Moscow.

STONE, M. H. (1929 ; 1930). Linear transformations in Hilbert space. *Proc. Nat. Acad. U.S.*, **15**, 198 and 423 ; **16**, 172.

STONE, M. H. (1932). *Linear Transformations in Hilbert Space*. Am. Math. Soc. Colloquium Publ., **15**, N.Y.

TUKEY, J. W. (1949). The sampling theory of power spectrum estimates. *Sym. Appl. Autocor. Anal. Phys. Prob.*, Oceanographic Inst., Woods Hole, Mass.

TUKEY, J. W. and HAMMING, R. W. (1949). *Measuring Noise Color*. Bell Tel. Lab., Murray Hill, N.J.

VON NEUMANN, J. (1929). Eigenwerttheorie Hermitescher Funktionaloperatoren. *Math. Ann.*, **102**, 49.

VON NEUMANN, J. (1931). Uber Funktionen von Funktionaloperatoren. *Ann. Math.* (2), **32**, 191.

VON NEUMANN, J. (1932). *Mathematische Grundlagen der Quantenmechanik*. Springer, Berlin. (English trans., Princeton Univ. Press.)

VON NEUMANN, J. (1932a). Proof of the quasiergodic hypothesis. *Proc. Nat. Acad. U.S.*, **18**, 70.

VON NEUMANN, J. (1933). *Functional Operators*. Inst. Adv. Study, Princeton, N.J. (Reprinted, Princeton Univ. Press, *Ann. Math. Studies*, **21**, **22**.)

WADSWORTH, G. P. and BRYAN, J. G. (1958). *Elementary Statistics*. M.I.T. Math. Dept., Cambridge, Mass.

WADSWORTH, G. P., ROBINSON, E. A., BRYAN, J. G., and HURLEY, P. M. (1953). Detection of reflections on seismic records by linear operators. *Geophysics*, **18**, 539.

WEYL, H. (1909). Über die Konvergenz von Reihen. *Math. Ann.*, **67**, 225.

WEYL, H. (1928). *Gruppentheorie und Quantenmechanik*. English trans., Dover, N.Y.

WHITTLE, P. (1951). *Hypothesis Testing in Time-series Analysis*. Almqvist and Wiksells, Uppsala.

WHITTLE, P. (1953). The analysis of multiple stationary time series. *J. Roy. Statist. Soc.*, **B**, **15**, 125.

WHITTLE, P. (1954). Some recent contributions to the theory of stationary processes. Appendix 2 of 1954 edition of WOLD (1938).

WIENER, N. (1923). Differential space, *M.I.T. J. Math. Phys.*, **2**, 131.

WIENER, N. (1930). Generalized harmonic analysis. *Acta Math.*, **55**, 117.

WIENER, N. (1942). *Stationary Time Series*. M.I.T. N.D.R.C. Report, Cambridge, Mass. (Reprinted, Wiley, N.Y.)

WIENER, N. (1948). *Cybernetics*. M.I.T., Cambridge, Mass., and Wiley, N.Y.

WINTNER, A. (1929). Zur Theorie der beschränkten Bilinearformen. *Math. Zeit.*, **30**, 228.

WOLD, H. (1938). *Stationary Time Series*. Almqvist and Wiksells, Uppsala. (2nd edn, 1954.)

WOLD, H., in association with JURÉEN, L. (1953). *Demand Analysis*. Wiley, N.Y.

YULE, G. U. (1921). On the time correlation problem. *J. Roy. Statist. Soc.*, **84**, 497.

YULE, G. U. (1927). On a model of investigating periodicities in disturbed series with special reference to Wolfer's sunspot numbers. *Phil. Trans.*, **A**, **226**, 267.

YULE, G. U. and KENDALL, M. G. (1937, 14th edn, 1958). *An Introduction to the Theory of Statistics*. Charles Griffin and Co., London.

INDEX