U. S. AIR FORCE
PROJECT RAND

TRANSLATION

ON LINEAR METHODS IN PROBABILITY THEORY

Kari Karhunen

Translated from the German by I. Selin

T-131

August 11, 1960

SOURCE OF TRANSLATION: Annales Academiae Scientiarum
Fennicae, A.I. 37,
Helsinki, 1947

FOREWORD

TABLE OF CONTENTS

INTRODUCTION

Probability theory can be considered from the standpoint of pure

mathematics as a branch of abstract measure and integration theory,

characterized by a special terminology and special ways of posing problems.

One axiomatic foundation of probability theory from this point of view is

due to Kolmogorov [1] (for an extended treatment of probability theory on

this basis, see Cramér [1] and Frechet [1]). In the latter work elementary

events are interpreted as elements of a set E, on which a measure P, called

the probability, is defined. Random events--i.e., events for which a

probability is defined--are then P-measurable subsets of E, and random

variables are P-measurable real functions defined on E.

The Kolmogorov theory makes possible a rigorous treatment of questions

concerning arbitrary infinite sets of random variables. To this class

belongs in particular the theory of stochastic processes. These are defined

in several ways. According to Khinchin [1], a stochastic process is a one-

parameter family of random variables. When such a system is given, to every

elementary event there corresponds a uniquely determined real function of

the parameter. Hence one may also treat the stochastic process as a set

of real functions, on which a probability measure is given. When Ω is the

set of all real functions, the two points of view are essentially equivalent.

In this case certain difficulties arise. For instance, in order to study

the structure of the process more closely, it is necessary either to allow

the set Ω to consist only of functions with certain given properties, such

as continuity or measurability, or else to consider the process as an abstract

function and assign appropriate definitions to the properties being investi-

gated. From the former point of view, which corresponds to a certain degree

to that of the ergodic theories (cf. Hopf [1], Doob [1] has proceeded to construct a general theory of stochastic processes.

The investigations of Khinchin (1) and Cramer (2) on the correlation theory of stationary stochastic processes essentially follow the latter line, in which an abstract function of the parameter is treated and which is related to the problems posed in statistical ergodic theory. These investigations contain no general theory, but rather are applicable mainly to special problems.

In the works cited above, an attempt was made to develop a systematic theory in which the stochastic processes would be interpreted as functions with values in an abstract space. There one makes the restrictive assumption, which is nevertheless fulfilled in the most important applications, that all the random variables under consideration have finite variance. Thus they are quadritically integrable functions defined on E and thus (cf. Sz. Nagy (1), pg. 6) can be considered as elements of a Hilbert space (in the general sense: the dimension may not be enumerable), provided that random variables which differ only on a set of measure zero are not considered as distinct (cf. Kolmogorov (2)). Then the general methods of Hilbert space, which are particularly well adapted to the treatment of linear problems, may be applied. One is led to a theory which is a generalization of the classical correlation theory, as the Hilbert space is a generalization of a finite-dimensional vector space. The spectral representations of a stochastic process are of primary importance. One such representation, which decomposes the process into pair-wise uncorrelated infinitesimal components, reflects the structural properties of the process and is an important tool in the investigation of its analytical properties (cf. Karhunen (2)).

The parameter of an ordinary stochastic process is a real number which is taken to be time in the physical and statistical interpretations. Looking ahead to later applications we shall generalize the statement of the problem so that the parameter element is an arbitrary set (i.e., a point in a multidimensional space). It no longer seems appropriate to use the term "process," which denotes a one-dimensional model. Following Wiener and some French authors, we shall use the simple term "random function," which is the natural one from our point of view. As did Cramér and Wiener, we shall mainly study complex-valued random functions. The results hold for the most part with little or no modification in the real case also.

In Section I we summarize the most important properties of infinite sets of random variables, mainly following Kolmogorov [1].

In Section II we construct the Hilbert space corresponding to a given set of random variables with finite dispersions. We reproduce several results which are fundamental for the succeeding theory.

In Section III we treat certain vital questions about random functions and their simplest correlation properties. In Section IV we present a new definition of the integral of a random function. The necessary and sufficient conditions for the existence of these integrals are given and their properties are studied.

In Section V the spectral representation of a random function is defined. A new concept of integral is there introduced, which can be interpreted as an integral of a complex function with respect to a random measure function. The concept is a generalization of Doob's definition of the integral of a real function with respect to a differential stochastic process (Doob [1] pg. 133). We give necessary and sufficient conditions that a spectral repre-

sentation of a given type be possible for a given random function.

In Section VI we treat applications of stationary random functions. The earlier results of Khinchin [1], Slutsky [2] and Cramer [3] are completed and stationary random functions with absolutely continuous spectra are investigated more deeply. Most of the results are analogous to those which Kolmogorov [2] proved for stationary time series. Later we hope to be able to announce further applications of the theory developed here (cf. Karhunen [1], Nr. 4, and [2]; several results are stated in the former note which are proved in the present work.

I. GENERALITIES ON PROBABILITY FIELDS AND RANDOM VARIABLES

1. Following Kolmogorov [1], the concept of a probability field is defined in the following manner. Let E be a set of elements ξ and ψ a set of subsets of E; the elements of the set E are called the elementary events; while those of ψ are the random events. We postulate that ψ satisfies the following axioms:

I. E is a finitely additive set, i.e., sums, intersections and differences of two sets of the class of sets ψ are also sets of the same class. In particular, ψ contains the null set 0.

II. ψ contains E.

III. To each set A in ψ there corresponds a non-negative real number P(A). This number P(A) is called the probability of the event A.

IV. P(E) = 1.

V. Let A_1, A_2 ...., A_n... be a denumerable number of disjoint sets of ψ, and let their sum $\sum A_n$ also belong to ψ. Then $P(\sum A_n) = \sum P(A_n)$. Therefore P(A) is a non-negative, completely additive set-function defined on ψ.

The finitely additive set ψ along with the set function P(A) is called a probability field.

The finitely additive set ψ is a Borel set if all denumerable sums of disjoint sets in ψ belong to ψ.[*] Then ψ contains all countable unions and intersections of its member sets. If ψ is a Borel set, then the corresponding probability field is called a Borel probability field. There is a theorem that for any given probability field there corresponds a uniquely defined smallest Borel extension. In the sequel we shall always consider that this extension has been realized, and hence we can limit ourselves to Borel probability fields in this work (see Kolmogorov [1], pp. 15-16).

---

[*] Tr. note: This definition of a Borel set seems incomplete. One must further postulate that all countable intersections and subtractions belong to ψ. cf. Cramér, Mathematical Methods of Statistics, Princeton (1946), pp.13;14.

It is often possible and expedient to transform the originally given probability field to a newer and simpler one in the following way: The set E is uniquely reflected on another set E'. To the different elements of E' correspond the different subsets of E. Each subset A' of E' has as its image in E the set of all the elements of E which transform into elements of A'. Let ψ be the system of all the subsets A' of E' whose images belong to ψ. Then ψ' is also a Borel set, and we set $P'(A') = P(A)$, where A is the image of A', so that $P'(A')$ is a probability function on ψ'. Hence ψ' and $P'(A')$ define a Borel probability field (Kolmogorov [1], pp. 19–20).

2. A unique real function $x(\xi)$ defined on the base set E is called a random variable, if for each real number a, the set of all ξ such that $x(\xi) < a$ belongs to ψ. In other words, a random variable is a real function uniquely defined on E which is measurable with respect to $P(A)$. The function $F(a;x) = P(x < a)$ is called the distribution function of the random variable x. It is obviously non-decreasing and continuous to the left. Furthermore we have the formulae:

$$\lim_{a \to -\infty} F(a;x) = F(-\infty;x) = 0, \qquad (1.1)$$

$$\lim_{a \to +\infty} F(a;x) = F(+\infty;x) = 1. \qquad (1.2)$$

Consider a finite set of random variables $x_1, x_2 \ldots x_n$. The set $\left[x_i < a_i; i = 1, 2, \ldots n \right]$ of all ξ for which the inequality $x_1 < a_1$, $x_2 < a_2, \ldots, x_n < a_n$ holds, belongs to ψ for any choice of the real numbers $a_1, a_2, \ldots a_n$. This is because the set is the intersection of the sets $\left[x_1 < a_1 \right], \left[x_2 < a_2 \right], \ldots \left[x_n < a_n \right]$. The function $F(a_1, a_2, \ldots, a_n; x_1, x_2, \ldots, x_n) = P\left[x < a_i; i = 1, 2, \ldots, n \right]$, which is defined for all values of $a_1, a_2, \ldots, a_n$, is called the n-dimensional distribution

function of the random variables x_1, x_2,..., x_n. It is non-decreasing and continuous to the left for each variable. In analogy with the formulae (1.1) and (1.2), we now have

$$\lim_{a_i \to -\infty} F(a_1,..., a_i,..., a_n; x_1,..., x_i,..., x_n)$$

$$= F(a_1,..., a_{i-1}, -\infty, a_{i+1}..., a_n; x_1,..., x_i,..., x_n) = 0,$$

$$(1.3)$$

$$\lim_{a_1 \to +\infty, a_2 \to +, ,..., a_n \to +\infty} F(a_1, a_2,..., a_n; x_1, x_2,..., x_n)$$

$$= F(+\infty, +\infty,..., +\infty; x_1, x_2,..., x_n) = 1. \qquad (1.4)$$

It is clear that the finite sums and products of random variables are themselves random variables. In particular, all linear combinations of random variables are random variables. The distribution functions of the generated random variables are obviously uniquely determined by those of the original variables.

We now consider an arbitrary set of random variables x_μ, where the index μ ranges over a set M of power m. These random variables form the base set E in the m-dimensional Euclidean space R^M. If m is finite, then R^M is a finite dimensional space; if m is denumerable, then R^M is the space of all real number sequences, if m is the power of the continuum, R^M is the space of all real functions of a real variable. In the sequel we shall principally consider sets M which are at the most of the power of the continuum.

Consider the transformation of the set R^M to a new basic set E'; we wish to specify the corresponding probability field. If m is finite and equal to n,

it is easily seen that ψ consists of all Borel sets of $R^M \cdot R^n$, and that

the corresponding probability function $P(A')$ is uniquely specified by know-

ledge of the n-dimensional distribution function $F(a_1,\ldots,a_n;x_1,\ldots,x_n)$.

However, when m is infinite, the problem becomes more complicated. Next we

consider subsets of R^M which are defined by a finite number of inequalities

$a_i \leq x_{\mu_i} < b_i (i=1,2,\ldots,n)$, where n and the indices μ_i are arbitrary, and

a_i and b_i denote arbitrary real numbers, $a_i < b_i$. The sets which are formed

by countable sums and intersections of such subsets are called Borel sets in

the space R^M. All of these Borel sets form a Borel system of sets, which

can be transformed into the system ψ'. Then the following fundamental theorem

holds (Kolmogorov [1] p.27):

The knowledge of all the finite-dimensional distribution functions

$F(a_1,a_2,\ldots,a_n; x_1,x_2,\ldots,x_n)$ uniquely defines the probability function $P(A')$

for all sets of ψ'. Every system of such distribution functions which are

symmetrical in the pairs (a_k, x_{μ_k}) and which satisfy the conditions

$$F(a_1,a_2,\ldots,a_i, +\infty, +\infty,\ldots, +\infty : x_{\mu_1},x_{\mu_2},\ldots,x_{\mu_i},\ldots,x_{\mu_n})$$

$$= F(a_1,a_2,\ldots,a_i; x_{\mu_1},x_{\mu_2},\ldots,x_{\mu_i})$$

(1.5)

for $i < n$, defines a probability function $P(A')$ on ψ' which satisfies axioms

I-V.

3. If x and y are two random variables, then obviously the probability

$P[x=y] = P[x-y = 0]$ is defined, because x—y is a random variable. If

$P[x=y] = 1$, then x and y are called equivalent. Two equivalent random

variables have the same distribution function: $F(a;x) = F(a;y)$. In the

sequel—when not explicitly stated to the contrary—we shall consider equiva-

lent random variables as being identical and therefore simply write x=y.

This is permitted because equivalence is a commutative and transitive property.

Two sets of random variables $\left[x_\mu\right]$ and $\left[y_\mu\right]$ are considered equivalent if their elements are pairwise equivalent: $x_\mu = y_\mu$. If the sets under consideration are countable, it follows from their equivalence that $P\left[x_\mu = y_\mu \text{ for all } \mu\right] = 1$. This is because the subset of E defined by the property $\left[x_\mu = y_\mu \text{ for all } \mu\right]$ is the intersection of the countable sequence of sets $\left[x_1 = y_1\right]$, $\left[x_2 = y_2\right]$, ..., and for each of the latter sets, $P\left[x_1 = y_1\right] = 1$. If on the other hand the given sets are non–denumerable, then $P\left[x_\mu = y_\mu \text{ for all } \mu\right]$ need not even be defined, or be equal to unity. However in most cases a non–denumerable set of random variables can be replaced by an equivalent set. This is due to the fact that such sets in most applications can be considered as idealizations of an arbitrary number of finite sets, and for the latter sets the equality always holds. The passage to an equivalent set means only that in this case one substitutes one idealized scheme for another.

In order to illustrate the problem, let us consider the following example: Let y be a given random variable. We construct a set of generated random variables, in which we choose the set of all real numbers for the index–set M, and for each μ

$$x_\mu = \begin{cases} 0, & \text{if } y \neq \mu \\ \\ \mu, & \text{if } y = \mu \end{cases}$$

When the distribution function of y is continuous, for each μ, $P(y=\mu) = 0$ and hence $P\left[x_\mu = 0\right] = 1$, $P\left[x_\mu = \mu\right] = 0$. The set $\left[x_\mu\right]$ is equivalent to the set $\left[x'_\mu\right]$, where $x'_\mu = 0$. Hence $P\left[x_\mu = x'_\mu \text{ for all } \mu\right] = P\left[y \neq \mu \text{ for all } \mu\right] = 0$. If the basic variable y is the observable, and the relation between y and the set $\left[x_\mu\right]$ is that of this problem, then the transformation to the equivalent

set $\begin{bmatrix} x'_\mu \end{bmatrix}$ is obviously not permitted. However, if we consider the variables x_μ as our observables, then each observable will almost certainly—i.e., with probability 1—be equal to zero. Hence if the relation between these generated variables and the basic variable y is not known to us, we must obviously view the scheme $\begin{bmatrix} x'_\mu \end{bmatrix}$ as the only natural one.

The problems with which we shall concern ourselves in the sequel are always such that the given sets of random variables can be considered basic; i.e., it will be assumed that the knowledge of the distribution functions is all the information which we possess about the structure of the set. An important question is then to decide if there exists a set of random variables which is equivalent to the original set and whose elements have a certain property β. If the answer is affirmative, we say that the original set almost certainly has the property β. If for instance each element of the set is almost certainly bounded, then we say that the entire set is almost certainly bounded, because obviously there exists an equivalent set whose elements are simultaneously bounded. (It is evident that one can also construct equivalent sets for which this does not hold.)

4. Given a sequence of random variables $x_1, x_2, \ldots, x_n, \ldots$. It can be proved (cf. Kolmogorov [1], p.30, for instance) that the set A of elementary events for which the sequence converges always belongs to the system of sets ψ. Thus it is meaningful to speak of the probability of the convergence of this sequence. In particular, if this probability is unity, we say that the sequence almost certainly converges. Then there also exists one (up to equivalence) uniquely defined random variable x, such that $P\begin{bmatrix} x = \lim_{n \to \infty} x_n \end{bmatrix}$ is equal to unity. We then simply write $x = \lim_{n \to \infty} x_n$ and say that the sequence almost certainly converges to x.

When there exists a random variable x which has the property that for arbitrarily small $\epsilon > 0$, the probability $P\left[|x_n - x| > \epsilon| \right]$ goes to zero as $n \longrightarrow \infty$, one says that the sequence $x_1, x_2, \ldots, x_n, \ldots$ converges to x in probability. x' is uniquely determined (up to equivalence). Any almost certainly convergent series also converges in probability to the same limit. However, a sequence may converge in probability without converging with probability 1; it may even diverge with probability one. For convergence in probability it is necessary and sufficient that for arbitrary positive δ and ϵ there exist an $n_{\delta, \epsilon}$, such that

$$P\left[|x_m - x_n| > \epsilon \right] < \delta$$

so long as both m and n $> n_{\delta_\epsilon}$.

If the sequence $x_1, x_2, \ldots, x_n, \ldots$ converges to x in probability, then the series of the corresponding distribution functions $f(a; x_n)$ converges to the distribution function $F(a, x)$ of x at every point of continuity of $F(a; x)$.

A sub-sequence which converges with probability one can be chosen from any sequence which converges in probability. The limit of the sub-sequence is the same as that of the original sequence (for various concepts of convergence of random variables, see Frechet [1] pp. 158 ff).

5. If the abstract integral (cf. for instance, Saks [1])

$$\int_E x(\xi) dP(\xi) = \int_E x dP(x)$$

exists, it is said that the random variable x has the expectation

$$E(x) = \int_E x dP(x) \tag{1.6}$$

If $F(a;x)$ is the distribution function of x, we also have

$$E(x) = \int_{-\infty}^{\infty} a d_a F(a;x),$$

where the latter integral is understood in the Stieltjes sense.

x^n and $|x|^n$ are also random variables for integer n_0. If the integral

$$\int_E |x|^n dP(x)$$

is finite, then $E(x^m)$ and $E(|x|^m)$ are defined and finite for $0 < m \leq n$. $E(|x|^n) = 0$ if and only if $x = 0$.

6. If x' and x'' are two random variables, then $x = x' + ix''$ defines a complex random variable. The conjugate random variable $x' - ix''$ will be denoted by $\bar{x}$, as is usual. The expectation $E(x)$ is understood $E(x') + iE(x'')$.

A sequence of complex random variables $x_1 = x_1' + ix_1''$, $x_2' + ix_2''$, ..., $x_n = x_n' + ix_n''$ is said to converge with probability one (or in probability) if both sequences x_1', x_2', x_3', ..., x_n', ... and x_1'', x_2'', ..., x_n'', converge with probability one(or in probability). The same criterion for convergence in probability holds as was the case above for real functions, since

$$\left. \begin{array}{c} |x_m' - x_n'| \\ |x_m'' - x_n'| \end{array} \right\} \leq |x_m - x_n| \leq |x_m' - x_n'| + |x_m'' - x_n''|.$$

If $E(|x|^2)$ is finite, we have

$$E(|x|^2) = E(x'^2) + E(x''2)$$

and hence $E(x'^2)$, $E(x''^2)$, $E(x')$ and $E(x''^2)$ are finite, and so $E(x)$ exists.

For arbitrary random variables x and y and complex numbers a and b, it is obvious that

$$E\left[|ax + by|^2\right] = |a|^2 E(|x|^2) + a\bar{b}\, E(x\bar{y}) + \bar{a}b\, E(\bar{x}y) + |b|^2 E(|y|\,) \geq 0.$$

If $E(|x|^2)$ and $E(|y|^2)$ are finite, it follows that $E(x\bar{y})$ is also finite. Now taking $a = E(|y|^2)$, $b = -E(x\bar{y})$, by a simple computation [directly for $E(|y|^2) > 0$, but also for general $E(|y|^2)$]

$$E(|x|^2)E(|y|^2) - |E(x\bar{y})|^2 \geq 0.$$

The Schwarz inequality also holds

$$|E(x\bar{y})| \leq \sqrt{E(|x|^2)E(|y|^2)}. \tag{1.7}$$

The equality holds if and only if there exist numbers a and b such that $ax + by = 0$.

We also note that, from the obvious inequality

$$E(|x|^2) = \int_E |x|^2 dP \geq \int_{|x| \geq a} |x|^2 dP \geq a^2 P\left\{|x| \geq a\right\} \quad (a > 0)$$

the Chebychev inequality follows for complex x also:

$$P\left\{|x| \geq a\right\} \leq \frac{E(|x|^2)}{a^2} \qquad (a > 0) \tag{1.8}$$

II. LINEAR SETS OF RANDOM VARIABLES

7. Let $\left[x_\mu\right]$ be a given set of real or complex random variables. The set of all finite linear combinations

$$\sum_{k=1}^{n} c_k x_{\mu k}$$

of the x_μ with real (complex) coefficients c_k is called the real (complex) linear hull L of the set$\left[x_\mu\right]$.

In the sequel we primarily treat complex sets, but the resulting theorems generally hold for real sets also.

L is obviously a linear set, i.e., if x and y belong to L, then the linear combination ax + by is also an element of L.

We now assume that $E(|x_\mu|^2)$ is finite for every μ. Then $E(x_\mu)$ is also defined and finite. We can take $E(x_\mu) = 0$ without loss of generality. Then, for each element z of L, $E(|z|^2)$ is finite and $E(z) = 0$.

The non-negative number $+\sqrt{E(|z|^2)}$ is called the norm of z and is denoted by $\| z \|$. $\|z\| = 0$ if and only if z = 0, and it then follows from $\|z\| = 0$ that $P\left\{z = 0\right\} = 1$ so that z is equivalent to 1.

For arbitrary y and z in L, $E(y\bar{z})$ is defined and finite, by the Schwarz inequality, and

$$|E(y\bar{z})| \leq \|y\| \cdot \|z\| .$$

We define the distance between the elements y and z by the expression $\|y - z\|$. This is permissible since $\|y - z\| = 0$ if and only if y = z, and the triangular inequality holds

$$\|y - z\| \leq \|y - w\| + \|w - z\| .$$

In fact

$$\|y - z\|^2 = E\left[|(y - w) + (w - z)|^2\right]$$

$$= E\left[|y - w|^2\right] + 2E\left[R(y - w)(w - z)\right] + E\left[|w - z|^2\right]$$

$$= \left[\|y - w\| + \|w - z\|\right]^2 - 2\left\{\|y - w\|\ \|w - z\| - E\left[R(y - w)(w - z)\right]\right\},$$

where R is the real part of the appropriate random variable. By the Schwarz inequality, the expression in the brackets $\langle\ \rangle$ is non-negative, whence the triangular inequality follows.

We say that a sequence $x_1, x_2, \ldots, x_n, \ldots$ of random variables converges in the mean to the random variable x, written $x = \underset{n \to \infty}{(\text{l.i.m. } x_n)}$, if $\lim_{n \to \infty} \|x_n - x\| = 0$. The following Lemma holds (Cramér [2], Fréchet [1], Levy [1])

Lemma 1: In order that a sequence of random variables $x_1, x_2, \ldots, x_n, \ldots$ with finite norms converge in the mean, it is necessary and sufficient that for every ξ there exist an n_ξ such that $\|x_m - x_n\| < \xi$ for all $m, n > n_\xi$ (the Cauchy convergence condition). If $\underset{n \to \infty}{\text{l.i.m. } x_n}$ is denoted by x, then $\|x\|$ is finite and $\|x\| = \lim_{n \to \infty} \|x_n\|$, $E(x) = \lim E(x_n)$. If two sequences $x_1, x_2, \ldots, x_n, \ldots$ and $y_1, y_2, \ldots, y_n, \ldots$ converge in the mean to the random variables x and y respectively, then $E(x\bar{y}) = \lim E(x_n\bar{y}_n)$.

Proof: If $x = \text{l.i.m. } x_n$, then for sufficiently large n_ξ and $n, m > n_\xi$, $\|x_m - x\| < \frac{1}{2}\xi$ and $\|x_n - x\| < \frac{1}{2}\xi$. By the triangular inequality, $\|x_m - x_n\| < \xi$. The Cauchy condition is also necessary. Assume the converse, i.e., that the Cauchy condition is fulfilled. By the Chebychev inequality

$$P\left\{|x_m - x_n| > \eta\right\} < \frac{\xi^2}{\eta^2} \qquad \text{for } m, n > n_\xi, \eta > 0.$$

If $\varepsilon = \eta \cdot \delta$ and $n_{\delta,\eta} = n_\varepsilon$, we have

$$P\left\{ |x_m - x_n| > \eta \right\} < \delta \qquad\qquad \text{for } m,n > n_{\delta,\eta}.$$

The given sequence thus converges in probability to a unique random variable x. We assert that $\|x\|$ is finite and that $x = \text{l.i.m. } x_n$. From the given sequence we can choose the sub-sequence x_{n_1}, x_{n_2}, ..., x_{n_k}, ... which converges to x almost certainly, i.e., at each point of a sub-set Q of the whole space E, with $P(Q) = 1$. For $N > n_\varepsilon$, $n_k > n_\varepsilon$, from the Cauchy condition

$$\int_Q |x_{n_k} - x_N|^2 dP = \int_E |x_{n_k} - x_N|^2 dP = \|x_{n_k} - x_N\|^2 < \varepsilon^2 ,$$

and in particular

$$\liminf_{k \to \infty} \int_Q |x_{n_k} - x_N|^2 dP < \varepsilon^2 .$$

Now it follows from Fatou's lemma (cf. Saks [1], p.29) that

$$\int_Q \lim_{k \to \infty} |x_{n_k} - x_N|^2 dP \leq \liminf_{k \to \infty} \int_Q |x_{n_k} - x_N|^2 dP < \varepsilon^2 .$$

In addition, since $\lim_{\substack{k \to \infty \\ Q}} |x_{n_k} - x_N|^2 = |x - x_N|^2$,

$$\int_Q |x - x_N|^2 dP < \varepsilon^2 .$$

Because $P(Q) = 1$, $P(E - Q) = 0$, and we finally obtain

$$\|x - x_N\|^2 \ = \ \int_E |x - x_N|^2 dP \ = \ \int_Q |x - x_N|^2 dP + \int_{E-Q} |x - x_N|^2 dP$$

$$= \ \int_Q |x - x_N|^2 dP \ < \ \varepsilon^2$$

or $\|x - x_N\| \leq \varepsilon$ for $N > n_\varepsilon$. This means that $x = \underset{n \to \infty}{\text{l.i.m.}}\ x_n$.

Furthermore $\|x\| \leq \|x_N\| + \|x - x_N\| \leq \|x_N\| + \varepsilon$. $\|x_N\|$ is finite, and therefore so is $\|x\|$. By the triangular inequality, $\|x\| - \|x_N\|$

$\leq \|x - x_N\| < \varepsilon$, and so $\|x\| = \underset{n \to \infty}{\lim} \|x_n\|$. If $x = \underset{n \to \infty}{\text{l.i.m.}}\ x_n$,

$y = \underset{n \to \infty}{\text{l.i.m.}}\ y_n$, obviously $x + y = \underset{n \to \infty}{\text{l.i.m.}} (x_n + y_n)$,

$x + iy = \underset{n \to \infty}{\text{l.i.m.}} (x_n + iy_n)$ and therefore

$$RE(x\bar{y}) \ = \ \frac{1}{2}\left[\|x+y\|^2 - \|x\|^2 - \|y\|^2 \right] \ = \ \underset{n \to \infty}{\lim} \frac{1}{2}\left[\|x_n+y_n\|^2 - \|x_n\|^2 - \|y_n\|^2 \right]$$

$$= \ \underset{n \to \infty}{\lim} RE(x_n\bar{y}_n) \ ,$$

$$IE(x\bar{y}) \ = \ \frac{1}{2}\left[\|x+iy\|^2 - \|x\|^2 - \|y\|^2 \right]$$

$$= \ \underset{n \to \infty}{\lim} \frac{1}{2}\left[\|x_n+iy_n\|^2 - \|x_n\|^2 - \|y_n\|^2 \right] \ = \ \underset{n \to \infty}{\lim} IE(x_n\bar{y}_n) \ .$$

With this the lemma is proved.

8. To the hull L we now add all the random variables which are limits in the mean of convergent sequences in L. The set formed in this way is called the closed linear hull of $\left\langle x_\mu \right\rangle$ and is denoted by L_2.

The set L_2 forms a Euclidean space, i.e., the following conditions hold:

A. L_2 is a linear space: a linear combination with complex (or real) coefficients of two elements of L_2 is still an element of L_2.

B. Scalar multiplication is defined in L_2, when we view $E(x\bar{y})$ as the scalar product of x and y. The following rules hold: $E(ax \cdot y) = aE(x\bar{y})$, $E[(x_1+x_2)\bar{y}] = E(x_1\bar{y}) + E(x_2\bar{y})$, $E(x\bar{y}) = \overline{E(y\bar{x})}$. Furthermore $E(x\bar{x}) = E(|x|^2) > 0$ for $x \neq 0$, and $E(x\bar{x}) = E(|x|^2) = 0$ for $x = 0$. The norm $\|x\|$ is defined as the square root of the product $E(x\bar{x})$.

$\|x-y\|$ is understood as the distance between two elements x and y.

C. L_2 is complete (with respect to convergence in the mean), i.e., the Cauchy convergence condition holds.

Only condition C must be proved. Let x_1, x_2, ..., x_n, ... be a sequence in L_2 which converges to x in the mean. We must show that x is an element of L_2. Since x_n is an element in L_2, it can be represented in the form $x_n = \underset{k \to \infty}{\text{l.i.m.}} \, x_n^{(k)}$, where $x_n^{(1)}$, $x_n^{(2)}$, ..., $x_n^{(k)}$, ... are elements of L. Then it is obvious that $x = \underset{n \to \infty}{\text{l.i.m.}} \, x_n^{(n)}$, where $x_1^{(1)}$, $x_2^{(2)}$, $x_n^{(n)}$, ... are elements of L. Hence by definition, x is an element of L_2.

The elements z_1, z_2, ..., z_n of L_2 are called linearly dependent if there exist constants c_1, c_2, ..., c_n, not all equal to zero, such that

$$\sum_{k=1}^{n} c_k z_k = 0.$$

Otherwise z_1, z_2, ..., z_n are said to be linearly independent.

Two elements x and y are called orthogonal, in symbols $x \perp y$, if $E(x\bar{y}) = 0$. Correspondingly, two subsets S_1 and S_2 of L_2 are called orthogonal, in symbols $S_1 \perp S_2$, if each element of S_1 is orthogonal to each element of S_2.

From this discussion it is seen that the general theorem for Euclidean spaces can be developed for L_2 (cf. for instance Sz. Nagy [1]).

9. A subset S of L_2 is called a base for L_2 if the finite linear combinations $\sum c_k y_k$ ($y_k \in$ S) are dense in L_2, i.e., for each x in L_2 and each $\varepsilon > 0$, there exist a finite set of elements $y_1, y_2, \ldots, y_n$ of S and corresponding constants $c_1, c_2, \ldots, c_n$ such that

$$\left\| x - \sum_{k=1}^{n} c_k y_k \right\| < \varepsilon.$$ Since L is dense in L_2, the set $\left\{ x_\mu \right\}$ is a base of L_2.

The smallest possible power of a base of L_2 is called the dimension of L_2 (Dim L_2). Dim L_2 cannot in any case be greater than the power of the set $\left\{ x_\mu \right\}$, or what is the same thing, the power m of the index set M. If m is finite and the x_μ are linear independent, then Dim L_2 = m. L_2 then has the same structure as the usual m—dimensional Euclidean space R^m. If Dim L_2 is countably infinite, L_2 is a Hilbert space (cf. von Neumann [1], Stone [1]). In both cases L_2 is separable, i.e., L_2 contains a countably infinite dense subset (this set is obtained by forming all the countably infinite linear combinations with rational coefficients of the elements of a denumerable base). If Dim L_2 is non—denumerable, L_2 is naturally non—separable.

Lemma 2: If $y \perp z$ for each z in a base, y = 0. For each $\varepsilon > 0$ there is an element y_ε in L_2, such that $\|y - y_\varepsilon\| < \varepsilon$ and $y_\varepsilon = \sum c_k z_k$, where each z_k is an element of the base. Then $E(y \bar{y}_\varepsilon) = 0$ and hence

$$\|y\|^2 \leq \|y\|^2 + \|y_\varepsilon\|^2 = \|y\|^2 + \|y_\varepsilon\|^2 - 2RE(y\bar{y}_\varepsilon) = \|y - y_\varepsilon\|^2 < \varepsilon^2,$$

whence $\|y\| < \varepsilon$ for each $\varepsilon > 0$ and hence y = 0.

Corollary: If $E(y_1 \bar{z}) = E(y_2 \bar{z})$ for every z in a base, then $y_1 = y_2$. Since $E\left[(y_2 - y_1)z\right] = E(y_2 z) - E(y_1 z) = 0$, $y_2 - y_1 = 0$.

A system D of elements is called orthonormal if for two elements y and z of D,

$$E(y\bar{z}) \;=\; \delta_{yz} \;=\; \begin{cases} 0, \; y \neq z \\ 1, \; y = z \end{cases}$$

If D is a base of L_2, then it is called complete.

Let $\left\{ z_x \right\}$ be an orthonormal system, which need not be countable, and y be an arbitrary element of L_2. If z_{x_1}, z_{x_2},...z_{x_n} are distinct, then the Bessel inequality holds:

$$\|y\|^2 - \sum_{i=1}^{n} |E(y\bar{z}_{x_i})|^2 \;=\; \|y - \sum_{i=1}^{n} \cdot E(y\bar{z}_{x_i})z_{x_i}\|^2 \;\geq\; 0.$$

Hence it follows that $\sum_{i} |E(yz_{x_i})|^2$ always converges in the sense that at most a countable number of terms are different from zero and by an arbitrary ordering of the terms the limit of the partial sums exists. Furthermore, $\sum E(y\bar{z}_x)z_x$ converges in the mean in the same sense as above. In fact, since

$$\|\sum_{i=1}^{n} E(y\bar{z}_{x_i})z_{x_i}\| \;\leq\; \|y\| \quad \text{for arbitrary } z_{x_i}$$

, at most a countable number

of the terms of the series in question can have a positive norm and therefore be different from zero. Let z_{x_1}, z_{x_2},...,z_{x_n},... be the corresponding elements of $\left\{ z_x \right\}$ taken in arbitrary order. The convergence of the series

$$\sum_{i=1}^{\infty} |E(y\bar{z}_{x_i})|^2 \quad \text{implies that} \quad \|\sum_{i=m}^{n} E(y\bar{z}_{x_i})z_{x_i}\|^2 \;=\; \sum_{i=m}^{n} |E(y\bar{z}_{x_i})|^2 \;<\; \varepsilon$$

if $n > m > n_\varepsilon$ and n_ε is taken large enough.

Clearly

$$y - \sum_x E(y\bar{z}_x)z_x \perp \left\{z_x\right\} .$$

If the system $\left\{z_x\right\}$ is complete, then $y = \sum_x (y\bar{z}_x)z_x$. This series is

then called the expansion of the random variable y in terms of the ortho-

normal system $\left\{z_x\right\}$.

One can show that L_2 always contains a complete orthonormal system

and that its power equals Dim L_2. If Dim L_2 is finite, the proof is trivial.

If Dim L_2 is countably infinite, then one can construct the system from a

countable base by means of the orthogonalization process of E. Schmidt. If

L_2 is not denumerable, it is necessary to apply the well-ordering principle

and define the orthogonal system by transfinite induction (Löwig [1]).

We wish to avoid the well-ordering principle in the sequel and therefore

will not enter more deeply into this question.

10. Let S be some subset of L_2. The linear hull of its elements,

which is also a subset of L_2, will be written (S). Precisely as was done

above with L, (S) can be augmented to be a Euclidean space. Such a subset

of L_2 which forms a Euclidean space is called a subspace in L_2. In particular,

[S] is the subspace spanned by S. Obviously S is a base of [S]. If, $z \perp S$,

where z is an element of L_2, then clearly $z \perp (S)$ and $z \perp [S]$.

The following lemma (F. Riesz [1]) is fundamental to our further work.

As in the case with Lemma 4 below, it holds without restriction on the

dimension of L_2, and can be proved without use of the well-ordering principle.

Lemma 3. If S is not a base of L_2, there exists an element z in L_2

different from zero, such that $z \perp [S]$.

Due to the importance of the theorem, we sketch the proof due to

F. Riesz: Let x be an element not contained in $[S]$ and let d be the distance

between x and $[S]$: $d = \min_{y \in [S]} ||x-y||$. Let y_1, y_2, ..., y_n, ... be a

sequence in $[S]$ such that $||x-y_n|| \rightarrow d$. This series converges in the mean

to an element y^* of $[S]$, and

$$\left|\left| \frac{y_m-y_n}{2} \right|\right|^2 = \left|\left| \frac{x-y_n}{2} - \frac{x-y_m}{2} \right|\right|^2 = \frac{1}{2}||x-y_n||^2 + \frac{1}{2}||x-y_m||^2$$

$$- \left|\left| x - \frac{y_m+y_n}{2} \right|\right|^2 \leq \frac{1}{2}||x-y_n||^2 + \frac{1}{2}||x-y_m||^2 - d^2 ,$$

because $\frac{y_m+y_n}{2}$ is an element of $[S]$ and therefore $\left|\left| x - \frac{y_m+y_n}{2} \right|\right| \geq d$.

Since $||x-y_n|| \rightarrow d$ we get $\left|\left| \frac{y_m-y_n}{2} \right|\right| \rightarrow 0$. Obviously $||x-y^*|| = d$.

Let y be an arbitrary element of $[S]$. For each complex c, $y^* + cy$ belongs

to $[S]$, and $||x - (y^* + cy)|| \geq d$, i.e.,

$$0 \leq ||x-(y^*+cy)||^2 - ||x-y^*||^2 = -\bar{c}E\left[(x-y^*)\bar{y}\right] - cE\left[\overline{y(x-y^*)}\right]$$
$$+ c\bar{c}||y||^2.$$

It is easily seen that for each c this is possible, because $E\left[\overline{y(x-y^*)}\right] = 0$.

For arbitrary y in $[S]$, $y \perp x-y^*$, i.e., $x-y^* \perp [S]$, so that $x-y^*$ is the

desired element.

Let S_1, S_2, ..., S_n, ... be pairwise orthogonal subspaces. The subspace

spanned by the system $\left\{S_1, S_2, ..., S_n, ...\right\}$ is denoted by $S_1(+)S_2(+)...(+)$

$S_n(+)$. $S_1(+)S_2(+)...(+)S_n(+)...$ therefore consists of those elements of L_2

which can be represented in the form $\sum_n y_n$ $(y_n \in S_n)$ with convergent $\sum_n ||y||^2$.

If M and N are subspaces and N is a subspace of M, then the set of the elements of M which are orthogonal to N is called the orthogonal complement of N in M; it is written $M(-)N$. $M(-)N$ is obviously a subspace. $M = N(+)[M(-)N]$. To prove this let $N(+)[M(-)N]$ be denoted by S; S is then a subspace of M. If S were not dense in M, then there would be an element z in M different from zero such that $z \perp S$. Then z would be orthogonal to N, and z would therefore be an element of $M(-)N$ and also an element of S, which is impossible because $z \perp S$. S is therefore a base of M, and because it is a linear closed subspace it must coincide with M.

From this discussion it follows that every element y of M can be represented in the form $z = z_1 + z_2$, with z_1 from N and z_2 from $M(-)N$. This representation is unique, since if $z_1 + z_2 = z_1' + z_2'$ with z_1 and z_1' from N, z_2 and z_2' from $M(-)N$, it follows that $z_1 - z_1' = z_2 - z_2'$. $z_1 - z_1'$ is simultaneously an element of N and $M(-)N$ and therefore $z_1 - z_1' = 0$; i.e., $z_1 = z_1'$, $z_2 = z_2'$.

The uniquely defined element z_1 is called the projection of y on N and is written $P_N y$. $P_N y$ depends only upon y and N and is independent of the special choice of the subspace M. $P_N(x+y) = P_N(x) + P_N(y)$, which follows almost directly from the definition of P_N. Since $P_N y \perp P_{L_2(-)N} y$,

$$E(y\overline{P_N y}) = E(P_N y \cdot \overline{P_N y}) = ||P_N y||^2.$$

If M is a subspace and N is a subspace of M, $M = N + [M(-)N]$, and $P_M y = P_N(y) + P_{M(-)N} y$. Therefore $P_N(P_M y) = P_N(P_N y) + P_N(P_{M(-)N} y)$, and since $N \perp M(-)N$,

$$P_N(P_M y) = P_N(y).$$

11. A complex-valued function $\lambda(x)$ defined on a subset S of L_2 is called a linear operation if for arbitrary x and y in S and complex numbers a and b, $\lambda(ax+by) = a\lambda(x)+b\lambda(y)$, so long as ax+by is an element of S. λ can be uniquely extended to (S) if for z = ax+by (x and y in S), we define $\lambda(z) = a\lambda(x)+b\lambda(y)$. The definition is unique, since it follows that if $z = a_1x_1+b_1y_1 = a_2x_2+b_2y_2$, and if $b_2 \neq 0$,

$$y_2 = \frac{a_1}{b_2}x_1 + \frac{b_1}{b_2}y_1 - \frac{a_2}{b_2}x_2 \quad \text{and} \quad \lambda(y_2) = \frac{a_1}{b_2}\lambda(x_1) + \frac{b_1}{b_2}\lambda(y_1) - \frac{a_2}{b_2}\lambda(x_2) \; ,$$

i.e., $a_1\lambda(x_1) + b_1\lambda(y_1) = a_2\lambda(x_2) + b_2\lambda(y_2)$.

$\lambda(x)$ is said to be bounded if there exists a number k such that for every x in (S), $|\lambda(x)| \leq k||x||$. If $\lambda(x)$ is bounded and continuous, then $|\lambda(x_n) - \lambda(x)| = |\lambda(x_n-x)| \leq k||x_n-x||$; i.e., if $x = \text{l.i.m.}x_n$ and $\lambda(x)$ is defined, then $\lambda(x) = \lim\lambda(x_n)$.

Proposition 1. If λ is defined on (S) and is bounded, it can be uniquely extended to [S] and remains bounded on [S].

Proof: Let $x = \text{l.i.m.}x_n$ with $x_n \in (S)$. Since $|\lambda(x_m) - \lambda(x_n)|$ $= |\lambda(x_m-x_n)| \leq k||x_m - x_n||$, then $\lim\lambda(x_n)$ exists. We also see that from the equality $\text{l.i.m.}x_n = \text{l.i.m.}x_n'$, it follows that $\lim\lambda(y_n)$ $= \lim\lambda(x_n')$, we can define $\lambda(x) = \lim\lambda(x_n)$. $\lambda(x+y) = \lim\lambda(x_n+y_n)$ $= \lim\lambda(x_n) + \lim\lambda(y_n) = \lambda(x) + \lambda(y)$ and $|\lambda(x)| = \lim |\lambda(x_n)| \leq \lim k||x_n||$ $= k||x||$, and therefore $\lambda(x)$ is linear on [S] and is bounded.

For arbitrary $z \in L_2$ it is clear that $\lambda(x) = E(x\bar{z})$ is a bounded linear operation defined on L_2, and z is uniquely defined by $\lambda(x)$. Conversely, the following is true (F. Riesz [1]):

Lemma 4: If $\lambda(x)$ is a bounded linear operation defined on a subspace S, then S contains one and only one element z^* such that $\lambda(x) = E(xz^*)$ for each $x \in S$.

Proof (after F. Riesz): The uniqueness of z^* is clear. It suffices to prove its existence. The set of x for which $\lambda(x) = 0$ is obviously a subspace M of S. M is either the whole space, in which case $\lambda(x) = E(x.0)$, or else S contains an element $z \neq 0$ which is orthogonal to M. Then we set

$z^* = \dfrac{\overline{\lambda(z)}}{\|z\|^2} \cdot z$; the equation $\lambda(x) = E(x\overline{z^*})$ holds as well for $x = z$ as for

any element $x \in M$. If x is arbitrary in S, then $x' = x - cz$ also belongs to M,

for $c = \dfrac{\lambda(x)}{\lambda(z)}$, because $\lambda(x') = \lambda(x) - c\lambda(z) = 0$. Then we have $\lambda(x) = \lambda(x' + cz) =$

$\lambda(x') + c\lambda(z) = E(x'\overline{z^*}) + E(cz.\overline{z^*}) = E(x\overline{z^*})$.

12. A sequence x_1, x_2, ..., x_n, ... of elements of L_2 is said to be weakly convergent to x if for each $z \in L_2$, $E(x\overline{z}) = \lim\limits_{n \to \infty} E(x_n\overline{z})$. The limit x is unique. For convergence in this sense it suffices to consider a base of L_2 instead of arbitrary $z \in L_2$.

If x_n converges to x in the mean, then by Lemma 1 it converges weakly to x.

III. RANDOM FUNCTIONS

13. Let T be an arbitrary set. To each element t of T let there corres-
pond a uniquely defined real (complex) random function $x(t)$. $x(t)$ is then
called a real (complex) random function of argument t. T is called the domain
of definition and the set of random variables $\{x(t)\}$ is the range of $x(t)$.

Two random functions $x(t)$ and $y(t)$ are said to be equivalent, if they
both have the same domain and if $x(t) = y(t)$ for every t. It is obvious
that $x(t)$ and $y(t)$ have the same range.

To each elementary event ζ in the probability field on which $x(t)$ is
defined, there corresponds a real or complex function $x(t;\zeta)$. Each such
function is called a realization of the random function $x(t)$. If the reali-
zations themselves are elementary events, we say that the random function is
directly defined. The set E of the elementary events is a subset of the
space R^T (c.f. para. 3). If $E = R^T$, i.e., if every real function defined on
T is a realization of $x(t)$, then $x(t)$ is called a general random function.

If $x(t)$ is general and directly defined, by paragraph 3 the probability
measure P on E is defined by the system of all of the finite-dimensional
distribution functions $F(a_1, a_2, \ldots, a_n; x(t_1), x(t_2), \ldots, x(t_n))$. How-
ever, if the realizations of $x(t)$ all possess some restrictive property $\mathfrak{C}$,
such as measurability, continuity, etc., i.e., if $x(t)$ is not general, so
that E is a proper subset of R^T, then the definition of the probability
measure is somewhat more difficult. Following Doob, one can begin with a
measure P* defined on R^T. Let S* be a P* measurable subspace of R^T. Then
one can construct a measure P on E, considering subsets S of E which can
be expressed in the form $S = E \cdot S*$. Doob has shown (Doob [1] pp. 109-110),
that one can set $P(S) = P*(S*)$, if every P* measurable subset of R^T containing

E has P* measure 1, i.e. if E is of outer P* measure 1. The set E, along with the probability measure E defines a non-general random function $x_\epsilon(t)$, which is obviously equivalent to the corresponding general random variable.

Conversely, let x(t) be a directly or indirectly defined random function, all of whose realizations possess the property ϵ. $x_\epsilon(t)$ defines a system of finite dimensional distribution functions $F(a_1, a_2, \ldots, a_n; x_\epsilon(t_1), x_\epsilon(t_2), \ldots, x_\epsilon(t_n))$. Hence $x_\epsilon(t)$ defines a measure P* on R^T. It is easily shown (Doob and Ambrose [1]) that the set E of the realizations of $x_\epsilon(t)$ has outer P*-measure 1. Let x(t) be a general random function and P* be the corresponding probability measure on R^T. We assume that there exists a random function $x_\epsilon(t)$, all the realizations of which possess the property ϵ, and which is equivalent to x(t). Since the same probability measure is obviously defined by equivalent random functions, then by the discussion above the set of realizations $x_\epsilon(t)$ has outer P*-measure 1. We say that a general random function x(t) possesses a property with probability one, if there exists a subset E of R^T with outer P* measure 1 all of whose elements possess the property E. It follows from the discussion above that this is the case if and only if there exists a random function, equivalent to x(t), all of whose realizations possess the property ϵ (cf paragraph 3).

We shall not distinguish between equivalent functions in the sequel. Therefore, we cannot directly investigate particular properties of the realizations. If we wish to do so, we must first show that there exists a random function, equivalent to the given function, whose realizations possess the desired property. Then we investigate the random function specified in this way.

Note that our definition of the random variable coincides essentially with Wiener's. The concept of a general directly defined random function is

the same as Khinchin's concept of a stochastic process. Doob was the first to treat non-general directly defined random variables. Doob and Ambrose [1] have exhaustively studied the connection between these different concepts.

14. As we did before, we shall assume throughout that $\{x(t)\}$ is defined and finite for every t in T. The closed linear hull of the range $\{x(t)\}$ of $x(t)$ is called the linear space corresponding to the random function $x(t)$ and is written $L_2(x)$, or simply L_2 if no confusion results.

The real (or complex) function

$$r(s,t) = E\left\{x(s)\overline{x(t)}\right\}, \tag{3.1}$$

defined on T x T$^{(1)}$ is called the correlation function of $x(t)$. By the Schwarz inequality, it is finite for all s and t in T. Obviously,

$$r(s,t) = \overline{r(t,s)} \tag{3.2}$$

and in particular

$$r(t,t) = \left\|x(t)\right\|^2 \geq 0. \tag{3.3}$$

Since

$$\sum_{i,j=1}^{n} r(t_i,t_j)a_i a_j = \left\|\sum_{i=1}^{n} a_i x(t_i)\right\|^2 \geq 0 \tag{3.4}$$

for arbitrary t_i and arbitrary complex a_i, $r(s,t)$ is a positive definite (or semidefinite) Hermite function. Conversely, it can be shown that any such function can be taken to be the correlation function of some random function (cf. Khinchin [1]).

15. The random function $x(t)$ is called separable if $L_2(x)$ is separable. If $x(t)$ is separable, then there exists a finite or denumerable orthonormal system z_1, z_2, ..., z_n ... in $L_2(x)$ such that for every t

$$x(t) = \sum_{k} z_k f_k(t) \quad \text{with } f_k(t) = E\left\{\overline{z_k}x(t)\right\},$$

_____

(1) T x T denotes the set of couples $\{s,t\}$ where $s,t \in T$.

where the sum, if infinite, converges in the mean. Furthermore, from

Lemma 1 we have

$$E\left\{x(s)\overline{x(t)}\right\} = E\left\{\sum_k z_k f_k(s) \overline{\sum_k z_k f_k(t)}\right\} = \sum_k f_k(s)\overline{f_k(t)} \ ,$$

so that the bilinear representation

$$r(s,t) = \sum_k f_k(s)\overline{f_k(t)} \tag{3.6}$$

holds for the correlation function. The series converges in the usual

sense for all s and t. Conversely, if $f_1(t), f_2(t), \ldots, f_k(t), \ldots$ are

finite functions defined on T, and the series $\sum \left|f_k(t)\right|^2$ converges through-

out T, it is easily seen that (3.6) defines a positive-definite Hermite

function on T x T. If $z_1, z_2, \ldots, z_k \ldots$ is an orthonormal system of random

variables, (3.5) defines a separable random function. Therefore, r(s,t) is

the correlation function of a separable random function if and only if it

can be represented in the form (3.6) and the series $\sum \left|f_k(t)\right|^2$ converges

for every t in T.

16. Let T be a topological space. A random function defined on T is

continuous in the mean, (continuous, for short) at the point t if for every

$\epsilon > 0$ there exists a neighborhood $D_\epsilon(t)$ of t, such that for every s in

$D_\epsilon(t)$ $x(s) - x(t) < \epsilon$. If x(t) is continuous at every point of the set

S, it is said to be continuous on S.

Theorem 1. The random function x(t) is continuous on the set S if and

only if the correlation function r(s,t) is continuous at every diagonal

point (t,t) of the set S x S; a posteriori r(s,t) is continuous at every

point of the set S x S.

Proof: We have

$$\left\| x(s) - x(t) \right\|^2 = r(s,s) - r(s,t) - r(t,s) + r(t,t);$$

from the continuity of $r(s,t)$ at the point (t,t), the continuity of $x(t)$ at the point t follows. Conversely, let $x(t)$ be continuous on S. If t and v are two arbitrary points of S, then for arbitrary s and u in T,

$$\left| r(s,u) - r(t,v) \right| = \left| E\left\{ x(s)\overline{x(u)} \right\} - E\left\{ x(t)\overline{x(v)} \right\} \right|$$

$$\leq \left| E\left\{ x(s)\overline{x(u)} \right\} - E\left\{ x(s)\overline{x(v)} \right\} \right| + \left| E\left\{ x(s)\overline{x(v)} \right\} - E\left\{ x(t)\overline{x(v)} \right\} \right|$$

$$= \left| E\left\{ x(s) \overline{[x(u) - x(v)]} \right\} \right| + \left| E\left\{ [x(s) - x(t)] \overline{x(v)} \right\} \right|$$

$$\leq \left\| x(s) \right\| \cdot \left\| x(u) - x(v) \right\| + \left\| x(s) - x(t) \right\| \cdot \left\| x(v) \right\|.$$

Choosing s and u such that $\left\| x(s) - x(t) \right\| < \epsilon$ and $\left\| x(u) - x(v) \right\| < \epsilon$, we have $\left\| x(s) \right\| \leq \left\| x(t) \right\| + \left\| x(s) - x(t) \right\| \leq \left| x(t) \right\| + \epsilon$. Therefore $\left| r(s,u) - r(t,v) \right| \leq \left[\left\| x(t) \right\| + \epsilon \right] \epsilon + \left\| x(v) \right\| \epsilon = \left[\left\| x(t) \right\| + \left\| x(v) \right\| \right] \cdot \epsilon + \epsilon^2.$ Hence, $r(s,t)$ is continuous at the point (t,v). Since this is an arbitrary point in $S \times S$, $r(s,t)$ is continuous at every point of $S \times S$.

Theorem 2. Let the domain T of the random function $x(t)$ be separable --i.e., T contains a dense, enumerable set of points--and $x(t)$ be continuous in T. Then $x(t)$ is separable. For the set $t_1, t_2, \ldots, t_n, \ldots$ is dense in T, and $x(t_1), x(t_2), \ldots x(t_n)$ forms a base of $L_2(x)$ since $s(t)$ is continuous.

17. Consider a set of random variables $x_1(t), x_2(t), \ldots, x_n(t)$ with the same domain T. We denote the closed linear hull of the sum of their domains by $L_2(x_1, x_2, \ldots, x_n, \ldots)$.

The function

$$r_{mn}(s,t) = E\left\{ x_m(s)\overline{x_n(t)} \right\} \tag{3.7}$$

is called the cross-correlation function of $x_m(t)$ and $x_n(t)$. Obviously,

$$r_{mn}(s,t) = \overline{r_{nm}(t,s)}. \tag{3.8}$$

If the random function $x_n(t)$ is continuous at the point t_o, then the function $r_{mn}(s,t)$ $(m = 1,2,\ldots)$ is continuous at every point (s,t_o), because

$$E\left\{x_m(s)\ \overline{x_n(t)}\right\} - E\left\{x_m(s)\overline{x_n(t_o)}\right\} \leq \|x_m(s)\| \cdot \|x_n(t) - x_n(t_o)\|$$

IV.   THE INTEGRAL OF A RANDOM FUNCTION

18.   Let the space T be the domain of the random function x(t), and define a measure τ such that T is the sum of a finite or denumerable number of τ-measurable subsets of finite measure.

We say that the random function x(t) is τ-measurable -- or simply measurable if no confusion is possible -- if the function $E\left\{\overline{zx(t)}\right\}$ is τ-measurable for every z in $L_2(x)$ (cf. Saks [1] pp. 12-18, where the measurability property of real functions is treated. A complex function is naturally considered measurable if the real and imaginary parts are measurable.)

Theorem 3.   Let x(t) and y(t) be measurable random functions, a and b complex constants and f(t) a measurable function. Furthermore, let $x_1(t), x_2(t), \ldots, x_n(t), \ldots$ be a sequence of measurable random functions which converges in the mean to the random function $x_\infty(t)$ for all t. Then the random functions ax(t) + by(t), f(t)x(t) and $x_\infty(t)$ are measurable.

Proof.   Let z be an arbitrary element of $L_2(x,y)$. We can write $z = z_1 + z_2$ where $z_1 \in L_2(x)$, and $z_2 \in L_2(x,y)$ (−) $L_2(x)$. Then $E\left\{\overline{zx(t)}\right\} = E\left\{\overline{z_1 x(t)}\right\} + E\left\{\overline{z_2 x(t)}\right\} = E\left\{\overline{z_1 x(t)}\right\}$, so that $E\left\{\overline{zx(t)}\right\}$ is a measurable function. Similarly, $E\left\{\overline{zy(t)}\right\}$ and finally $\bar{a}E\left\{\overline{zx(t)}\right\} + \bar{b}E\left\{\overline{zy(t)}\right\} = E\left\{\overline{z\left[ax(t) + by(t)\right]}\right\}$ is measurable. Since every element of $L_2(ax + by)$ is obviously an element of $L_2(x,y)$, by the definition ax(t) + by(t) is measurable.

If $E\left\{\overline{zx(t)}\right\}$ is measurable, then so is $\overline{f(t)} \cdot E\left\{\overline{zx(t)}\right\} = E\left\{\overline{zf(t)x(t)}\right\}$, from which the measurability of f(t)x(t) follows.

As is shown above, $E\left\{\overline{zx(t)}\right\}$ is measurable for every z in $L_2(x_1, x_2, \ldots, x_n, \ldots)$ and each n. From Lemma 1, $E\left\{\overline{zx_\infty(t)}\right\} = \lim_{n \to \infty} E\left\{\overline{zx_n(t)}\right\}$.

From a well-known theorem in the theory of functions of real variables (cf. Saks $\begin{bmatrix} 1 \end{bmatrix}$ p. 15), the measurability of $E\{z\overline{x_\infty(t)}\}$, and hence of $x_\infty(t)$, follows. (It is easily seen that each $z\epsilon L_2(x_\infty)$ is an element of $L_2(x_1, x_2,...,x_n,...)$.)

Theorem 4. $x(t)$ is measurable if and only if the correlation function $r(s,t)$ is a measurable function of t for every fixed s.

Proof. The necessity of the condition follows directly from the definition $r(s,t) = E\left\{x(s)\overline{x(t)}\right\}$. The sufficiency is seen in this way: First let z be an element of the linear hull $L(x)$ of $\left\{x(t)\right\}$; $z = \sum_{k=1}^{n} a_k x(t_k)$. Then

$$E\left\{z\overline{x(t)}\right\} = \sum_{k=1}^{n} a_k E\left\{x(t_k)\overline{x(t)}\right\} = \sum_{k=1}^{n} a_k r(t_k,t) \qquad (4.0)$$

is measurable, since it is a linear combination of the measurable functions $r(t_1,t)$, $r(t_2,t)$, ..., $r(t_n,t)$. Any $z\epsilon L_2(x)$ can be expressed in the form $z = \underset{n\to\infty}{\text{l.i.m.}} z_n$, where $z_1,z_2, ..., z_n, ...$ belong to $L(x)$. Then $E\left\{z\overline{x(t)}\right\} = \underset{n\to\infty}{\lim} E\left\{z_n\overline{x(t)}\right\}$. Since every $E\left\{z_n\overline{x(t)}\right\}$ is measurable, so is $E\left\{z\overline{x(t)}\right\}$. By the definition $x(t)$ is seen to be measurable.

19. Now we wish to define the definite integral of a random function on a τ-measurable set S. Following Slutsky $\begin{bmatrix} 1 \end{bmatrix}$ and Doob $\begin{bmatrix} 1 \end{bmatrix}$ (cf. also Doob and Ambrose $\begin{bmatrix} 1 \end{bmatrix}$), one can investigate the measurability of the realizations. If almost all the realizations are measurable and their integrals on S constitute a random variable, the integral of the random function can be defined by this random variable. Note that it is not sufficient to assume that $x(t)$ is measurable with probability one. In this case the integrals of almost all the realizations exist, but they need not constitute a random variable in the probability field (Slutsky's definition is deficient in this respect). This definition leads to a fairly complicated development. This can be avoided

by following Cramér's procedure. Instead of considering the individual realizations, one treats the limit in the mean of the random function defined by the Riemann sums. Cramér's definition requires the continuity of the random function; it has no meaning for a discontinuous random function--especially if T is not a topological space. We shall present a very general, and yet simple, definition which requires only the measurability of the random function in the sense given above.

Theorem 5. Let $x(t)$ be a measurable random function and S a τ-measurable subset of T. If $E\left\{z\overline{x(t)}\right\}$ possesses a finite definite integral on S for each $z \in L_2$, and if the expression

$$\frac{1}{\|z\|}\left|\int_S E\left\{z\overline{x(t)}\right\} d\tau(t)\right| \tag{4.1}$$

is bounded, L_2 contains a uniquely defined $X(S)$, such that for each $z \in L_2$,

$$E\left\{z\overline{X(S)}\right\} = \int_S E\left\{z\overline{x(t)}\right\} d\tau(t) \tag{4.2}$$

Proof: For convenience we write

$$I_S(z) = \int_S E\left\{z\overline{x(t)}\right\} d\tau(t). \tag{4.3}$$

Then obviously

$$I_S(az_1 + bz_2) = aI_S(z_1) + bI_S(z_2)$$

and, if m is the upper bound of the expression (4.1)

$$\left|I_S(z)\right| \leq m\|z\|.$$

$I_S(z)$ is a bounded linear operation on L_2. By Lemma 4 there exists a

uniquely defined element $X(S) \in L_2$ such that $E\left\{z\overline{X(S)}\right\} = I_S(z)$.

Let S_1 and S_2 be disjoint sets, so that for each z, if $X(S_1)$ and $X(S_2)$ exist

$$E\left\{z\overline{X(S_1 + S_2)}\right\} = I_{S_1+S_2}(z) = I_{S_1}(z) + I_{S_2}(z) = E\left\{z\overline{X(S_1)}\right\} + E\left\{z\overline{X(S_2)}\right\} .$$

From the corollary to Lemma 2

$$X(S_1 + S_2) = X(S_1) + X(S_2)$$

Therefore $X(S)$ is an additive random set function defined on a system of τ-measurable subsets of T. We call $X(S)$ the definite integral of $x(t)$ on S, and denote it by

$$X(S) = \int_S x(t)d\tau(t) \tag{4.4}$$

It is easily verified that the definite integral has the usual properties

$$\begin{cases} \displaystyle\int_{S_1+S_2} x(t)d\tau(t) = \int_{S_1} x(t)d\tau(t) + \int_{S_2} x(t)d\tau(t), \\[2em] \displaystyle\int_S [x_1(t) + x_2(t)]d\tau(t) = \int_S x_1(t)d\tau(t) + \int_S x_2(t)d\tau(t), \\[2em] \displaystyle\int_S ax(t)d\tau(t) = a\int_S x(t)d\tau(t), \\[2em] \displaystyle\int_S 0 \cdot d\tau(t) = 0 \end{cases} \tag{4.5}$$

provided that the integrals on the right exist.

We have

$$E\left\{X(S_1)\overline{X(S_2)}\right\} = I_{S_2}\left[X(S_1)\right] = \int_{S_2} E\left\{X(S_1)\overline{x(t)}\right\} d\tau(t)$$

$$= \overline{\int_{S_2} E\left\{x(t)\overline{X(S_1)}\right\} d\tau(t)} = \overline{\int_{S_2} I_{S_1}\left[x(t)\right] d\tau(t)} = \overline{\int_{S_2} d\tau(t) \int_{S_1} E\left\{x(t)\overline{x(s)}\right\} d\tau(s)}$$

$$= \overline{\int_{S_2} d\tau(t) \int_{S_1} r(t,s)d\tau(s)} = \int_{S_2} d\tau(t) \int_{S_1} r(s,t)d\tau(s) = \iint_{S_1 S_2} r(s,t)d\tau(s)d\tau(t),$$

and therefore

$$E\left\{\int_{S_1} x(t)d\tau(t) \overline{\int_{S_2} x(t)d\tau(t)}\right\} = \iint_{S_1 S_2} r(s,t)d\tau(s)d\tau(t) \qquad (4.6)$$

and in particular

$$\left\|\int_S x(t)d\tau(t)\right\| = \sqrt{\iint_{S\,S} r(s,t)d\tau(s)d\tau(t)} \ . \qquad (4.7)$$

The finiteness of the integral

$$\iint_{S\,S} r(s,t)d\tau(s)d\tau(t) \qquad (4.8)$$

is necessary for the existence of the integral (4.4). We shall show that this condition is also sufficient. It remains to be shown that the expression (4.1) is bounded. Next we note that the integral (4.8) must be, in every case, real and positive. This is clearly seen when the integral is approximated by the appropriate sums, and it is noted that $r(s,t)$ is a positive definite function. Now let z be an arbitrary element of L_2. We write

$$f_z(t) = E\left\{\bar{z}x(t)\right\}, \quad x_z(t) = -\frac{z}{\|z\|^2} f_z(t).$$

Clearly, $z \perp x_z(t)$, and so

$$r(s,t) = E\left\{x(s)\overline{x(t)}\right\} = E\left\{x_z(s)\overline{x_z(t)}\right\} + \frac{f_z(s)\overline{f_z(t)}}{\|z\|^2} .$$

from which follows

$$\int_S \int_S r(s,t)d\tau(s)d\tau(t) = \int_S \int_S E\left\{x_z(s)\overline{x_z(t)}\right\}d\tau(s)d\tau(t)$$

$$+ \frac{1}{\|z\|^2} \int_S \int_S f_z(s)\overline{f_z(t)}d\tau(s)d\tau(t) \geq \frac{1}{\|z\|^2}\left|\int_S f_z(t)d\tau(t)\right|^2 ,$$

because the integral of $E\left\{x_z(s)\overline{x_z(t)}\right\}$ is positive (as is integral (4.8)).
We also obtain

$$\frac{1}{\|z\|}\left|\int_S E(x\overline{z(t)})d\tau(t)\right| \leq \sqrt{\int_S \int_S r(s,t)d\tau(s)d\tau(t)} , \qquad (4.9)$$

and the assertion is proved.

If the integral (4.4) exists, we say that x(t) is integrable on S. From the above, we have

Theorem 6. The random function x(t) is integrable on the set S if and only if its correlation function r(s,t) is finitely integrable on the set S x S.

20. Let f(t) be a real, τ-measurable function, such that $\left\|x(t)\right\| \leq f(t)$ for almost all t, i.e., for all t except a set of τ-measure zero. We have $\left|r(s,t)\right| \leq \left\|x(s)\right\| \left\|x(t)\right\| \leq f(s)f(t)$ for almost all s and t. Then it follows that

$$\int_S \int_S r(s,t)d\tau(s)d\tau(t) \leq \int_S \int_S f(s)f(t)d\tau(s)d\tau(t) = \left[\int_S f(t)d\tau(t)\right]^2$$

and by (4.7),

$$\left\| \int_S x(t)d\tau(t) \right\| \leq \int_S f(t)d\tau(t) . \tag{4.10}$$

In particular, if $\| x(t) \|$ itself is measurable, we can set $f(t) = \| x(t) \|$, and thus we obtain

$$\left\| \int_S x(t)d\tau(t) \right\| \leq \int_S \| x(t) \| d\tau(t). \tag{4.11}$$

__Theorem 7__. Let $x_1(t)$, $x_2(t)$, $\ldots$, $x_n(t)$, $\ldots$ be a sequence of measurable random functions which converges in the mean for all t. Assume furthermore, that there exists a real function $f(t)$, integrable on the set S, such that for every n and almost every t $\| x_n(t) \| \leq f(t)$. Then $x(t) = \underset{n \to \infty}{\text{l.i.m.}} x_n(t)$ is integrable on S, and the sequence

$$\int_S x_1(t)d\tau(t), \int_S x_2(t)d\tau(t), \ldots, \int_S x_n(t)d\tau(t), \ldots \tag{4.12}$$

is weakly convergent to $\int_S x(t)d\tau(t)$. If the convergence of the sequence $x_1(t), x_2(t), \ldots, x_n(t), \ldots$ is uniform in S, then

$$\int_S x(t)d\tau(t) = \underset{n \to \infty}{\text{l.i.m.}} \int_S x_n(t)d\tau(t).$$

Proof: from Lemma 1, $\| x(t) \| = \lim \| x_n(t) \| \leq f(t)$ for almost all t. By Theorem 3 and (4.10) $x(t)$ is integrable on S. Furthermore, Lemma 1 holds for all z in L_2 and almost all t.

$$\left| E\left\{ z\overline{x(t)} \right\} \right| = \lim_{n \to \infty} \left| E\left\{ z\overline{x_n(t)} \right\} \right| \leq \| z \| \cdot f(t).$$

holds for all z in L_2 and almost all t. By a well-known theorem of Lebesgue (cf. Saks [1], p. 29), it now follows $E\left\{ z\overline{x(t)} \right\}$ is integrable on S and that

$$\int_S E\left\{z\overline{x(t)}\right\} d\tau(t) = \lim_{n\to\infty} \int_S E\left\{z\overline{x_n(t)}\right\} d\tau(t)$$

By our definition of the integral of a random function it follows for each z in L_2

$$E\left\{z \overline{\int_S x(t)d\tau(t)}\right\} = \lim_{n\to\infty} E\left\{z \overline{\int_S x_n(t)d\tau(t)}\right\},$$

i.e., the sequence (4.12) converges weakly to $\int_S x(t)d\tau(t)$. Now let the sequence of random functions converge uniformly. Then there exists a number n_ϵ such that for every t, $||x_m(t) - x_n(t)|| < \epsilon$, so long as m, n $> n_\epsilon$. We choose a set S_ϵ of finite mass $\tau(S_\epsilon)$, such that

$$\int_{S-S_\epsilon} f(t)d\tau(t) < \epsilon$$

This is possible, since T is the sum of a countable number of subsets of finite measure, and therefore we can write

$$\int_S f(t)d\tau(t) = \sum_{n=1}^\infty \int_{S_n} f(t)d\tau(t)$$

where each S_n has finite measure. Then we can set $S_\epsilon = S_1 + S_2 + \ldots + S_n$, if n is taken sufficiently large. Then by (4.10) we have

$$\left\| \int_{S_\epsilon} x_m(t)d\tau(t) - \int_{S_\epsilon} x_n(t)d\tau(t) \right\| = \left\| \int_{S_\epsilon} (x_m(t) - x_n(t))d\tau(t) \right\| < \epsilon \cdot \tau(S_\epsilon)$$

and

$$\left\| \int_{S-S_\epsilon} x_m(t)d\tau(t) - \int_{S-S_\epsilon} x_n(t)d\tau(t) \right\| \leq \left\| \int_{S-S_\epsilon} x_m(t)d\tau(t) \right\| + \left\| \int_{S-S_\epsilon} x_n(t)d\tau(t) \right\|$$

$$\leq 2 \int_{S-S_\epsilon} f(t)d\tau(t) < 2\epsilon.$$

Then it follows that

$$\left\| \int_S x_m(t)d\tau(t) - \int_S x_n(t)d\tau(t) \right\| \leq \left\| \int_{S_\epsilon} x_m(t)d\tau(t) - \int_{S_\epsilon} x_n(t)d\tau(t) \right\|$$

$$+ \left\| \int_{S-S_\epsilon} x_m(t)d\tau(t) - \int_{S-S_\epsilon} x_n(t)d\tau(t) \right\| \leq (\tau(S_\epsilon) + 2) \cdot \epsilon$$

The last assertion follows from Cauchy's convergence condition.

21. A measurable random function $x(t)$ is of class zero, if for each measurable subset S of T

$$\int_S x(t)d\tau(t) = 0$$

From the definition of the integral it follows that $E\left\{z\overline{x(t)}\right\}$ vanishes almost everywhere for all z in L_2. It is easily seen that $x(t)$ is of class zero if and only if $r(s,t)$ vanishes for almost all s in T. A typical example is a random function whose range consists of mutually orthogonal elements, so that $r(s,t) = 0$ if $s \neq t$.

Theorem 8. A separable class-zero random function vanishes almost everywhere.

Proof: Separability implies that L_2 contains a countable base z_1, z_2, ...,z_n,.... Let S_n be the set of t for which $E\left\{z_n\overline{x(t)}\right\}$ does not vanish. Since each S_n is empty, so is the union $S = \sum S_n$. For every t not contained in S, $E\left\{z_n\overline{x(t)}\right\} = 0$ for all n, and so $x(t) = 0$.

Let T be a topological space and τ a measure defined so that the neighborhoods of T be empty. Then a class-zero function defined on T is either discontinuous at every point, or equal to zero. In fact, let $x(t_0) \neq 0$.

If $x(t)$ were continuous at $t = t_o$, there would be a neighborhood D_ϵ of t_o such that $\|x(t) - x(t_o)\| < \epsilon$ for every $t \in D_\epsilon$, and then

$$\left| E\left\{ x(t_o)\overline{x(t)} \right\} \right| > \left\| \tfrac{1}{2} x(t_o) \right\|^2 - \tfrac{1}{2} \epsilon^2.$$

Choosing $\epsilon < \|x(t_o)\|$, $\left| E\left\{ x(t_o)x(t) \right\} \right| > 0$ for every $t \in D_\epsilon$. This is impossible, since $E\left\{ x(t_o)\overline{x(t)} \right\}$ vanishes almost everywhere.

V. SPECTRAL REPRESENTATION OF RANDOM FUNCTIONS

22. Let R be an arbitrary set of elements a. Define a measure σ

on R such that R is the union of a countable number of subsets of finite

measure. Define a random set function Z(S) on the class of all the

measurable subsets of R so that (a) Z(S) is additive, i.e., for two

disjoint sets S_1 and S_2

$$Z(S_1 + S_2) = Z(S_1) + Z(S_2) \tag{5.1}$$

and (b) for any two measurable sets S_1 and S_2

$$E\left\{ Z(S_1)\overline{Z(S_2)} \right\} = E\left\{ Z(S_2)\overline{Z(S_1)} \right\} = \sigma(S_1 \cdot S_2). \tag{5.2}$$

A random set function which satisfies these two conditions is called

a random spectral function. From condition (b), for two distinct sets

S_1 and S_2 $Z(S_1) \perp Z(S_2)$. Set $S_1 = S_2 = S$ in (5.2), resulting in

$$\|Z(S)\|^2 = \sigma(S). \tag{5.3}$$

Let $S_1 \subset S_2 \subset \ldots \subset S_n \subset \ldots$ and $S = \lim_{n \to \infty} S_n$ be

measurable sets of finite measure. Then

$$Z(S) = \underset{n \to \infty}{\text{l.i.m.}} Z(S_n). \tag{5.4}$$

Since $S = \lim_{n \to \infty} S_n$ it follows (cf. Saks [1] p. 19) that $\sigma(S) = \lim_{n \to \infty} \sigma(S_n)$ and

$$\|Z(S) - Z(S_n)\|^2 = \|Z(S - S_n)\|^2 = \sigma(S - S_n) \to 0.$$

One can show that a spectral function can be constructed for any set R
of the above type, so that our definition is never vacuous.

23. We now wish to define the integral with respect to the spectral
function $Z(S)$ of a complex function $f(a)$ defined on R;

$$\int_R f(a)dZ(a) \qquad (5.5)$$

We now assume that $f(a)$ is bounded and $\sigma(R)$ is finite. It is well known
(cf. Saks [1], pp 14,30) that $f(a)$ may be uniformly approximated by a se-
quence $f_1(a)$, $f_2(a)$,..., $f_n(a)$,... of measurable functions which assume
only a finite number of different values on R. Let the values of $f_n(a)$
be $v_1^{(n)}$, $v_2^{(n)}$, ..., $v_{N_n}^{(n)}$ and the corresponding sets, on which these values
are assumed, $S_1^{(n)}$, $S_2^{(n)}$,..., $S_{N_n}^{(n)}$. We form the finite sums

$$I(f_n) = \sum_{k=1}^{N_n} v_k^{(n)} Z(S_k^{n}). \qquad (5.6)$$

By (5.2)

$$\| I(f_n) \|^2 = \sum_{k=1}^{N_n} | v_k^{(n)} |^2 \sigma(S_k^{(n)}). \qquad (5.7)$$

We wish to prove that the sums (5.6) converge to a definite limit in $L_2(Z)$
with increasing n.

For arbitrary n and m, obviously

$$S_k^{(n)} = S_k^{(n)} \cdot R = S_k^{(n)} \sum_{i=1}^{N_m} S_i^{(m)} = \sum_{i=1}^{N_m} S_k^{(n)} \cdot S_i^{(n)}.$$

and therefore

$$\| I(f_m) - I(f_n) \|^2 = \| \sum_{i=1}^{N_m} \nu_i^{(m)} Z(S_i^{(m)}) - \sum_{j=1}^{N_m} \nu_j^{(n)} Z(S_j^{(n)}) \|^2$$

$$= \| \sum_{i=1}^{N_m} \sum_{j=1}^{N_n} \nu_i^{(m)} Z(S_i^{(m)} \cdot S_j^{(n)}) - \sum_{i=1}^{N_m} \sum_{j=1}^{N_n} \nu_j^{(n)} Z(S_j^{(n)} \cdot S_i^{(m)}) \|^2$$

$$= \| \sum_{i=1}^{N_m} \sum_{j=1}^{N_n} (\nu_i^{(m)} - \nu_j^{(n)}) Z(S_i^{(m)} \cdot S_j^{(n)}) \| = \| I(f_m - f_n) \|^2$$

$$= \sum_{i=1}^{N_m} \sum_{j-1}^{N_n} |\nu_1^{(m)} - \nu_j^{(n)}|^2 \sigma(S_i^{(m)} \cdot S_j^{(n)})$$

$$= \sum_{i=1}^{N_m} \sum_{j=1}^{N_n} | f_m(a_{ij}) - f_n(a_{ij}) |^2 \sigma(S_i^{(m)} \cdot S_j^{(n)}),$$

where a_{ij} is any element of the set $S_i^{(m)} \cdot S_j^{(m)}$. By the uniform convergence of the sequence $f_1(a)$, $f_2(a)$, ..., $f_n(a)$, ..., we have $|f_m(a) - f_n(a)| < \epsilon$ for every a, so long as m, n $> n_\epsilon$. Then

$$\| I(f_m) - I(f_n) \|^2 \leq \sum_{i=1}^{N_m} \sum_{j=1}^{N_n} \epsilon^2 \sigma(S_i^{(m)} \cdot S_j^{(n)}) = \epsilon^2 \sigma(R).$$

Since $\sigma(R)$ is finite, the sequence $I(f_1)$, $I(f_2)$, ... $I(f_n)$, ... converges to a definite limit $I(f)$ due to the Cauchy convergence condition.

In the special case where $f(a) = \lim f_n(a) = 0$, from (5.6) and Lemma 1, $\left\| I(f) \right\| = \lim \left\| I(f_n) \right\| = 0$ and $\left\| I(f) \right\| = 0$. It follows directly that $I(f)$ is independent of the particular choice of the approximating functions. In fact, if $\lim f_n(a) = \lim f_n{}'(a) = f(a)$, we have $\left\| \text{l.i.m. } I(f_n) - \text{l.i.m. } I(f_n{}') \right\| = \lim \left\| I(f_n) - I(f_n{}') \right\| = \lim \left\| I(f_n - f_n{}') \right\| = 0$, and l.i.m. $I(f_n) = $ l.i.m. $I(f_n{}')$. Then we can define

$$\int_R f(a) \, d Z (a) = \underset{n \to \infty}{\text{l.i.m. }} I(f_n) \tag{5.8}$$

(Cf. Doob $\begin{bmatrix}1\end{bmatrix}$, pp 131-133).

It follows directly from (5.6) that the expression (5.8) possesses the usual properties of an integral

$$\int_R \left[f(a) + g(a) \right] dZ (a) = \int_R f(a) dZ(a) + \int_R g(a) dZ(a), \tag{5.9}$$

$$\int_{R_1 + R_2} f(a) dZ(a) = \int_{R_1} f(a) dZ(a) + \int_{R_2} f(a) dZ(a), \tag{5.10}$$

$$\int_R 0 \, dZ(a) = 0, \quad \int_R 1 \, dZ(a) = Z(R); \tag{5.11}$$

R_1 and R_2 in (5.10) are measurable, disjoint subsets of R.

By the definition of the usual Lebesgue integral, (5.7) leads to

$$\left\| I(f) \right\|^2 = \underset{n \to \infty}{\lim} \left\| I(f_n) \right\|^2 = \underset{n \to \infty}{\lim} \sum_{k=1}^{N_n} \left| v_k{}^{(n)} \right|^2 \sigma(S_k^{(n)}) = \int_R \left| f(a) \right|^2 d\sigma(a),$$

and hence

$$\left\| \int_R f(a) dZ(a) \right\|^2 = \int_R \left| f(a) \right|^2 d\sigma(a). \tag{5.12}$$

Applying (5.12) and the formula

$$E\left\{I_1 \cdot \bar{I}_2\right\} = \frac{1}{4}\left\{\left\|I_1 + I_2\right\|^2 - \left\|I_1 - I_2\right\|^2 + i\left\|I_1 + iI_2\right\|^2 - i\left\|I_1 - iI_2\right\|^2\right\}$$

by a simple computation we obtain

$$E\left\{\int_R f(a)dZ(a) \; \overline{\int_R g(a)dZ(a)}\right\} = \int_R f(a)\overline{g(a)}d\sigma(a) \; . \qquad (5.13)$$

In particular, if $g(a)$ is the characteristic function of a measurable subset S of R (i.e. $g(a) = 1$ if $a \in S, g(a) = 0$ if $a \in R - S$), then (5.11) and (5.13) result in

$$E\left\{Z(S) \cdot \int_R \overline{f(a)dZ(a)}\right\} = \int_S \overline{f(a)}d\sigma(a) \; . \qquad (5.14)$$

If S_1 and S_2 are two measurable subsets of R_1 it is easily seen that

$$E\left\{\int_{S_1} f(a)dZ(a) \; \overline{\int_{S_2} g(a)dZ(a)}\right\} = \int_{S_1 \cdot S_2} \overline{f(a)}g(a)d\sigma(a)$$

In particular,

$$\int_{S_1} f(a)dZ(a) \perp \int_{S_2} g(a)dZ(a), \text{ if } S_1 \cdot S_2 = 0 \; . \qquad (5.15)$$

24. We now extend the definition of the integral (5.5) so that we assume only that the integral (5.16) is finite, even if $\left|f(a)\right|$ is not bounded or $\sigma(R)$ is not finite:

$$\int_R \left|f(a)\right|^2 d\sigma(a) \; . \qquad (5.16)$$

First let $\left|f(a)\right|$ be bounded. Let $R = \sum_{n=1}^{\infty} R_n$, where the sets R_n are disjoint and all are of finite measure. We now write simply

$$\int_R f(a)dZ(a) = \sum_{n=1}^{\infty} \int_{R_n} f(a)dZ(a) \; .$$

The right-hand integrals are defined in the paragraph above. By (5.15), they are mutually orthogonal and by

$$\sum_{n=1}^{\infty} \left\| \int_{R_n} f(a)dZ(a) \right\|^2 = \sum_{n=1}^{\infty} \int_{R_n} \left| f(a) \right|^2 d\sigma(a) = \int_{R} \left| f(a) \right|^2 d\sigma(a)$$

the series converges in the mean.

Now let $\left| f(a) \right|$ be unbounded, and consider the sets R_1', R_2', ..., R_n', ... defined by the inequalities

$$\left| f(a) \right| < 1, \; 1 \leqq \left| f(a) \right| < 2, \; ..., \; n - 1 \leqq \left| f(a) \right| < n, \; ...$$

These sets are obviously disjoint and, due to the measurability of $f(a)$, measurable. For each R_n',

$$\int_{R_n'} f(a)dZ(a)$$

is defined. Then we can write

$$\int_{R} f(a)dZ(a) = \sum_{n=1}^{\infty} \int_{R_n'} f(a)dZ(a)$$

where the series converges in the mean.

It is easily verified that the limits in the above definitions are independent of the special choice of the approximating sequence of sets R_1, R_2, ..., R_n, ... or R_1', R_2', ... R_n', ... and so the definitions are unique. The properties of the integral stated previously remain true for the extended definition.

Equation (5.12) shows that it is essential that the integral (5.16) be finite.

25. We consider the set $\Lambda_2(R)$ of all complex (real) functions defined on R which are quadratically integrable with respect to σ. Two functions

$f(a)$ and $g(a)$ of this set will be considered identical if they coincide

almost everywhere with respect to σ, i.e. if

$$\int_R \left| f(a) - g(a) \right|^2 \, d\sigma(a) = 0 \qquad \rbrack$$

It is well known (cf. Sz. Nagy $\lbrack 1 \rbrack$, p 6, for instance) that $\Lambda_2(R)$ is

a Euclidian space if the scalar product is defined by

$$\int_R f(a)\overline{g(a)}d\sigma(a)$$

and the distance by

$$\left\| f - g \right\| = \sqrt{\int_R \left| f(a) - g(a) \right|^2 \, d\sigma(a)}. \qquad (5.17)$$

The theorems of §2 for L_2 hold intact for $\Lambda_2(R)$.

According to (5.5), to each element of $\Lambda_2(R)$ there corresponds a

uniquely determined element of $L_2(Z)$. We wish to show conversely that to

each element of $L_2(Z)$ there corresponds a unique element of $\Lambda_2(R)$. If z

is a member of $L(Z)$, this is trivial, since for $z = \sum_{k=1}^{k} c_k \, Z(S_k)$,

$$f(a) = c_k \quad \text{for } a \in S_k \qquad (5.18)$$

is the desired function. If z is an arbitrary element of $L_2(Z)$, there

exists a sequence of elements $z_1, z_2, \ldots, z_n, \ldots$ in $L(Z)$ such that

$z = \underset{n \to \infty}{\text{l.i.m.}} \, z_n$. To each z_n there corresponds a function $f_n(a)$ defined by

(5.18), such that

$$z_n = \int_R f_n(a)dZ(a) .$$

By (5.9) and (5.12),

$$\left\| z_m - z_n \right\|^2 = \int_R \left| f_m(a) - f_n(a) \right|^2 d\sigma(a) = \left\| f_m - f_n \right\|^2. \qquad (5.19)$$

The sequence $\{f_n\}$ convergences in the mean if $\{z_n\}$ does. By the Riesz-Fischer theorem (cf. Jessen $\begin{bmatrix}1\end{bmatrix}$, p 80), there exists a quadratically integrable function f(a) (with respect to σ) such that $\|f_n - f\| \longrightarrow 0$. It is easily seen that

$$z = \int_R f(a)dZ(a) \tag{5.20}$$

From Eq. (5.19), it is obvious that f is uniquely determined by z for arbitrary z_m and z_n in $L_2(Z)$. The same equation shows that the transformation defined between $\Lambda_2(R)$ and $L_2(Z)$ by (5.20) is isometric.* Thus we have

Theorem 9. Equation (5.20) defines a one-to-one isometric transformation between the linear spaces $L_2(Z)$ and $\Lambda_2(R)$.

The representation (5.20) of the set $L_2(Z)$ is called a spectral representation. If in particular R is the set of all positive integers and σ (S) is the number of elements in S, (5.20) results in the special expansion of the elements of $L_2(Z)$ into a complete orthonormal system Z(1), Z(2), ..., Z(n), ... (cf. paragraph 9).

26. Let f(t,a) be a complex function defined on TxR and quadratically integrable on R with respect to σ for every $t \in T$. Then

$$x(t) = \int_R f(t, a)dZ(a) \tag{5.21}$$

is a function defined on T whose range lies in $L_2(Z)$. It follows that $L_2(x) \leq L_2(Z)$. In order that $L_2(x) = L_2(Z)$, it is necessary and sufficient that the set $\{f(t, a)\}_T$ of functions f(t, a) in R, obtained by allowing t to assume all values in T, form a base for $\Lambda_2(R)$. This is so because then and only then is $\{x(t)\}_T$ a base for $L_2(Z)$. $\{f(t, a)\}_T$ is a base for $L_2(R)$ if and only if it is orthogonal to no element of $\Lambda_2(R)$.

---

*i.e., measure-preserving

This means that there exists no function φ (a) quadratically integrable

with respect to σ , such that $\| \varphi \| = 1$ and

$$\int_R f(t, a) \varphi(a) d\sigma(a) = 0 \text{ for every t in T} \tag{5.22}$$

By (5.13), the correlation function of the random function (5.21)

is

$$r(s, t) = \int_R f(s, a)\overline{f(t, a)} d\sigma(a) . \tag{5.23}$$

27. Now we wish to investigate under what conditions a spectral function

in R can be found corresponding to a given random function x(t), such that

(5.21) hold. We already know that a necessary condition is that (5.23)

hold for the correlation function r(s, t) of x(t). We shall show that this

condition is also sufficient.

Theorem 10. If the representation (5.23) of the correlation function

r(s, t) of x(t) is valid, then there exists a spectral function Z(S) de-

fined on R for which (5.21) holds. $L_2(x) = L_2(Z)$ if and only if there

exists no function φ (a) quadratically integrable with respect to σ such

that

$$\int_R |\varphi(a)|^2 d\sigma(a) = 1 \tag{5.23^1}$$

which satisfies condition (5.22).

Proof: First assume that there exists no function ρ (a) possessing

the stated properties.

Let S be an arbitrary measurable subset of R of finite measure $\sigma(S)$.

We define on operation $\mathscr{A}_S$ on L(x) by

$$\mathscr{A}_S(z) = \int_S \sum_{k=1}^n c_k f(t_k, a) d\sigma(a), \text{ if } z = \sum_{k=1}^n c_k x(t_k). \tag{5.24}$$

and show that $\mathcal{L}_S$ is unique, linear, and bounded.

(1) $\mathcal{L}_S$ is unique - Let

$$z = \sum_{k=1}^{n} c_k x(t_k) = \sum_{k=1}^{n'} c'_k x(t_k') \ .$$

Then for every t, by (5.23),

$$E\{z\overline{x(t)}\} = \sum_{k=1}^{n} c_k E\{\overline{x(t)}x(t_k)\} = \sum_{k=1}^{n} c_k \int_R \overline{f(t,a)}f(t_k,a)d\sigma(a)$$

$$= \sum_{k=1}^{n'} c'_k E\{\overline{x(t)}x(t'_k)\} = \sum_{k=1}^{n'} c'_k \int_R \overline{f(t,a)}f(t'_k,a)d\sigma(a)$$

and consequently

$$\int_R f(t,a)\left\{\sum_{k=1}^{n} c_k f(t_k,a) - \sum_{k=1}^{n'} c'_k f(t'_k,a)\right\}d\sigma(a) = 0$$

Denoting the expression in brackets $\{\}$ by $\varphi(a)$, we have our assumption

$$\int_R |\varphi(a)|^2 d\sigma(a) = 0$$

By the Schwarz inequality,

$$\left|\int_S \varphi(a)d\sigma(a)\right|^2 \leqq \int_S d\sigma(a) \int_S |\varphi(a)|^2 d\sigma(a) = 0,$$

and

$$\int_S \varphi(a)d\sigma(a) = \int_S \sum_{k=1}^{n} c_k f(t_k,a)d\sigma(a) - \int_S \sum_{k=1}^{n'} c'_k f(t'_k,a)d\sigma(a) = 0$$

from which the uniqueness of $\mathcal{L}_S$ follows.

(2) $\mathcal{L}_S$ is linear. Let

$$z_1 = \sum_{k=1}^{n_1} c_k^{(1)} x(t_k^{(1)}), \quad z_2 = \sum_{k=1}^{n_2} c_k^{(2)} x(t_k^{(2)}).$$

And so

$$\mathcal{L}_S(b_1 z_1 + b_2 z_2) = \int_S \left\{ b_1 \sum_{k=1}^{n_1} c_k^{(1)} f(t_k^{(1)}, a) + b_2 \sum_{k=1}^{n_2} c_k^{(2)} f(t_k^{(2)}, a) \right\} d\sigma(a)$$

$$= b_1 \int_S \sum_{k=1}^{n_1} c_k^{(1)} f(t_k^{(1)}, a) d\sigma(a) + b_2 \int_S \sum_{k=1}^{n_2} c_k^{(2)} f(t_k^{(2)}, a) d\sigma(a)$$

$$= b_1 \mathcal{L}_S(z_1) + b_2 \mathcal{L}_S(z_2).$$

(3) $\mathcal{L}_S$ is bounded. It follows from the Schwarz inequality that

$$\left| \mathcal{L}_S(z) \right|^2 = \left| \int_S \sum_{k=1}^{n} c_k f(t_k, a) d\sigma(a) \right|^2 \leq \sigma(S) \int_S \left| \sum_{k=1}^{n} c_k f(t_k, a) \right|^2 d\sigma(a) .$$

By (5.23),

$$\int_S \left| \sum_{k=1}^{n} c_k f(t_k, a) \right|^2 d\sigma(a) \leq \int_R \left| \sum_{k=1}^{n} c_k f(t_k, a) \right|^2 d\sigma(a)$$

$$= \int_R \sum_{i,j=1}^{n} c_i \bar{c}_j f(t_i, a) \overline{f(t_j, a)} d\sigma(a) = \sum_{i,j=1}^{n} c_i \bar{c}_j \int_R f(t_i, a) \overline{f(t_j, a)} d\sigma(a)$$

$$= \sum_{i,j=1}^{n} c_i \bar{c}_j r(t_i, t_j) = E \left\{ \sum_{i,j=1}^{n} c_i \bar{c}_j x(t_i) \overline{x(t_j)} \right\}$$

$$= E \left\{ \left| \sum_{k=1}^{n} c_k x(t_k) \right|^2 \right\} = \| z \|^2$$

and consequently

$$\left| \mathcal{L}_S(z) \right| \leq \sqrt{\sigma(S)} \| z \| .$$

Since $\sigma(S)$ is finite, $\mathcal{L}_S$ is bounded.

$\mathcal{L}_S$ can be extended to be a bounded linear operation on $L_2(x)$. By Lemma 4, $L_2(x)$ contains a unique element $Z(S)$ such that for every $z \in L_2(x)$,

$$E\left\{ z\overline{Z(S)} \right\} = \mathcal{L}_S(z) \tag{5.25}$$

Thus we obtain a random set function which is defined for every measurable S of finite measure. $Z(S)$ is additive, clearly $\alpha_{S_1 + S_2} = \alpha_{S_1} + \alpha_{S_2}$ if S_1 and S_2 are disjoint measurable sets, and consequently for all $z \in L_z(x)$,

$$E\left\{ z\overline{Z(S_1 + S_2)} \right\} = \alpha_{S_1 + S_2}(z) = \alpha_{S_1}(z) + \alpha_{S_2}(z)$$

$$= E\left\{ z\overline{Z(S_1)} \right\} + E\left\{ z\overline{Z(S_2)} \right\} = E\left\{ z\overline{[Z(S_1) + Z(S_2)]} \right\},$$

whence it follows that $Z(S_1 + S_2) = Z(S_1) + Z(S_2)$.

We now write

$$f_z(a) = \sum_{k=1}^{n} c_k f(t_k, a), \text{ if } z = \sum_{k=1}^{n} c_k x(t_k). \tag{5.26}$$

By (1), $f_z(a)$ is unique almost everywhere in R. By (5.24), for all $z \in L(x)$

$$\alpha_S(z) = \int_S f_z(a) d\sigma(a) = \int_R f_z(a) e_s(a) d\sigma(a), \tag{5.27}$$

where $e_s(a)$ is the characteristic function of the set S. If $y = \sum_{k=1}^{m} b_k x(s_k)$ is a second element in $L(x)$, we have

$$E(yz) = E\left\{ \sum_{k=1}^{m} b_k x(s_k) \sum_{k=1}^{n} \overline{c_k x(t_k)} \right\} = \sum_{i=1}^{m} \sum_{j=1}^{n} b_i \overline{c_j} r(s_i, t_j)$$

$$= \int_R \sum_{i=1}^{m} \sum_{j=1}^{n} b_i \overline{c_j} f(s_i, a) \overline{f(t_j, a)} d\sigma(a)$$

$$= \int_R \sum_{i=1}^{m} b_i f(s_i, a) \sum_{j=1}^{n} \overline{c_j f(t_j, a)} d\sigma(a)$$

and therefore

$$E(\overline{yz}) = \int_R f_y(a) \overline{f_z(a)} d\sigma(a). \tag{5.28}$$

In particular, for $y = z$ we obtain

$$\|z\|^2 = \int_R |f_z(a)|^2 d\sigma(a). \tag{5.29}$$

If z is any element in $L_2(x)$, then $z = \text{l.i.m.}_{n \to \infty} z_n$ where $z_n \in L(x)$. From (5.29), it follows that

$$\left\| z_m - z_n \right\|^2 = \int_R \left| f_{z_m}(a) - f_{z_n}(a) \right|^2 d\sigma(a).$$

In paragraph 25 it was shown that the sequence $f_{z_1}(a)$, $f_{z_2}(a)$, ... $f_{z_n}(a)$,... tends toward a limit in the mean $f_z(a)$. Since

$$\mathscr{A}_S(z) = \lim_{n \to \infty} \mathscr{A}_S(z_n) = \lim_{n \to \infty} \int_S f_{z_n}(a) d\sigma(a) = \int_S f_z(a) d\sigma(a)$$

and, if $y = \text{l.i.m.} \, y_n$, $y_n \in L(x)$,

$$E(y\bar{z}) = \lim_{n \to \infty} E(y_n \bar{z}_n) = \lim_{n \to \infty} \int_R f_{y_n}(a) \overline{f_{z_n}(a)} d\sigma(a)$$

$$= \int_R f_y(a) \overline{f_z(a)} d\sigma(a)$$

and the Eqs. (5.27), (5.28) and (5.29) hold for arbitrary y and z in $L_2(x)$.

Consider the expression $\mathscr{A}_S(x(t))$. By (5.27), on the other hand

$$\mathscr{A}_S(x(t)) = \int_R f(t,a) e_S(a) d\sigma(a)$$

and from (5.25) and (5.28), on the other hand

$$\mathscr{A}_S(x(t)) = E\left\{ x(t)\overline{Z(S)} \right\} = \int_R f(t,a) \overline{f_{z(S)}(a)} d\sigma(a).$$

For all t,

$$\int_R f(t,a) \left[\overline{f_{z(S)}(a)} - e_S(a) \right] d\sigma(a) = 0$$

and furthermore, almost everywhere on R

$$\overline{f_{z(S)}(a)} = f'_{z(S)}(a) = e_S(a).$$

and finally

$$E\{Z(S_1)\overline{Z(S_2)}\} = \int_R e_{s_1}(a)e_{s_2}(a)d\sigma(a) = \sigma(S_1 \cdot S_2).$$

Thus $Z(S)$ is a spectral function.

The range of $Z(S)$ is a base of $L_2(x)$, for otherwise there would exist an element $z \in L_2(x)$, different from zero, such that $z \perp Z(S)$ for every S of finite measure. Then for all S,

$$E\{z\overline{Z(S)}\} = \mathscr{L}_S(z) = \int_S f_z(a)d\sigma(a) = 0 ,$$

and consequently $f_z(a) = 0$ almost everywhere in R and thus by (5.29) $\| z \| = 0$. Then it must also be true that $L_2(x) = L_2(Z)$.

We must still prove that (5.21) holds. By (5.25) and (5.27) we have for all S and t

$$E\{x(t)\overline{Z(S)}\} = \mathscr{L}_S(x(t)) = \int_S f(t,a)d\sigma(a)$$

and by (5.14)

$$E\{\int_R f(t,a)dZ(a) \cdot \overline{Z(S)}\} = \int_R f(t,a)d\sigma(a) .$$

Thus for all S,

$$E\{x(t)\overline{Z(S)}\} = E\{\int_R f(t,a)dZ(a)\cdot\overline{Z(S)}\}$$

Since the range of $Z(S)$ is a base of $L_2(x)$, the corollary to Lemma 2 results in Eq. (5.21).

We next consider the case in which there exists at least one function $\varphi(a)$ which satisfies condition (5.22) and which does not vanish almost everywhere. We denote the closed linear hull of the set $\{f(t,a)\}_T$

by $\Lambda_2(f)$. Then $\varphi \perp \Lambda_2(f)$, so that the set $\Lambda_2(R)(-) \Lambda_2(f)$ is not empty (our notation will correspond to that used in $\S_2$). Let $\{g(t,a)\}_T$ be a base of $\Lambda_2(R)(-)\Lambda_2(f)$, where the arbitrarily chosen parameter t assumes all values in a certain set T'. T' is so chosen that it contains no element in common with the set T.

$$q(s,t) = \int_R g(s,a)\overline{g(t,a)}d\,\sigma(a)$$

defines a positive-definite Hermite function on T' x T'. One can then construct a random function y(t) on T' with the correlation function q(s,t) (cf. para. 14). Obviously, the corresponding probability field F' can be chosen in such a way that it has no elementary event in comma with the probability field corresponding to x(t).

We now join the probability fields F and F' together and denote the resulting field by F x F'. The random events of F x F' belong to the set A x A', with A in F and A' in F', and the corresponding probability measure is defined by $P(A \times A') = P(A)P(A')$. Both the random functions x(t) and y(t) belong to the field F x F'. Thus on F x F' we can define a random function w(t) with domain T + T', if we write

$$w(t) = \begin{cases} x(t) \text{ for } t \in T \\ y(t) \text{ for } t \in T' \end{cases}$$

Furthermore, defining f(t,a) = g(t,a) for $t \in$ T', the correlation function of w(t) is

$$p(s,t) = \int_R f(s,a)\overline{f(t,a)}d\,\sigma(a)\ .$$

In fact, if $s,t \in T$, $p(s,t) = r(s,t)$; $s,t \in T'$, $p(s,t) = q(s,t)$. If $s \in T$, $t \in T'$, $p(s,t) = 0$ since $f(s,a) \perp g(t,a)$ and

$$E\{w(s)\overline{w(t)}\} = E\{x(s)\overline{y(t)}\} = \int_{E \times E'} x(s)\overline{y(t)}dP = \int_E x(s)dP \int_{E'} \overline{y(t)}dP = 0 ,$$

where $E(E')$ is the set of elementary events in $F(F')$. In all cases, $p(s,t) = E\{w(s)\overline{w(t)}\}$.

Now $\{f(t,a)\}_{T+T'}$ is a base of $\Lambda_2(R)$, because $\{f(t,a)\}_T$ is a base of $\Lambda_2(f)$ and $\{f(t,a)\}_{T'} = \{g(t,a)\}_{T'}$ is a base of $\Lambda_2(R)(-)\Lambda_2(f)$; $\Lambda_2(R) = \Lambda_2(f)(+)(\Lambda_2(R)(-)\Lambda_2(f))$. Hence $\Lambda_2(R)$ contains a function $\varphi(a)$ with the properties

$$\int_R f(t,a)\varphi(a)d\sigma(a) = 0 \quad \text{for all } t \text{ in } T + T'$$

and

$$\int_R |\varphi(a)|^2 d\sigma(a) = 1 .$$

By the part of the theorem proved above, there exists a spectral function $Z(S)$ with $L_2(Z) = L_2(w) = L_2(x) (+) L_2(y)$ such that

$$w(t) = \int_R f(t,a)dZ(a) .$$

If $t \in T$ we have the representation (5.21). It cannot be true that $L_2(x) = L_2(Z)$, since by paragraph 26, $\{f(t,a)\}_T$ would not be a base of $\Lambda_2(R)$ - hence the theorem is proved.

28. The measures σ and ϵ together define a measure on the product set $T \times R$ (cf. Saks $\lceil 1 \rceil$, pp 82-87, and Jessen $\lceil 1 \rceil$, pp 42-56). If the function $f(t,a)$ is integrable on $T \times R$, then according to Fubini's theorem the order of integration with respect to τ and σ may be

interchanged. Then, there results

Theorem 11: If the function $f(t,a)$ is measurable on $T \times R$, then so is the random function $x(t)$ defined by (5.21). $x(t)$ is integrable on T if and only if

$$\int_R \left| \int_T f(t,a)d\tau(t) \right|^2 d\sigma(a)$$

is finite, in which case

$$\int_T x(t)d\tau(t) = \int_R \left[\int_T f(t,a)d\tau(t) \right] dZ(a) \,. \tag{5.30}$$

Proof: By (5.28), for every $z \in L_2(x)$,

$$E\left\{ z\overline{x(t)} \right\} = \int_R f_z(a)\overline{f(t,a)}d\sigma(a) \,.$$

By Fubini's theorem, $E\left\{ z\overline{x(t)} \right\}$ is measurable for all z. By (4.2) it follows further that for all S of finite measure,

$$E\left\{ Z(S) \overline{\int_T x(t)d\tau(t)} \right\} = \int_T E\left\{ Z(S) \overline{\int_R f(t,a)dZ(a)} \right\} d\tau(t) \,.$$

$$= \int_T \left[\int_S \overline{f(t,a)}d\sigma(a) \right] d\tau(t) \,.$$

Similarly,

$$E\left\{ Z(S) \overline{\int_R \left[\int_T f(t,a)d\tau(t) \right] dZ(a)} \right\} = \int_S \left[\overline{\int_T f(t,a)d\tau(t)} \right] d\sigma(a)$$

$$= \int_S \left[\int_T \overline{f(t,a)}d\tau(t) \right] d\sigma(a) \,.$$

Now, by the Schwarz inequality,

$$\left| \int_S \left[\int_T \overline{f(t,a)}d\tau(t) \right] d\sigma(a) \right| \leq \sqrt{\sigma(S) \int_R \left| \int_T \overline{f(t,a)}d\tau(t) \right|^2 d\sigma(a)} \,.$$

The left-hand integral is finite, and by Fubini's theorem

$$\int_T \left[\int_S \overline{f(t,a)}d\sigma(a) \right] d\, \tau(t) = \int_S \left[\int_T \overline{f(t,a)}d\, \tau(t) \right] d\sigma(a).$$

Consequently

$$E\left\{ Z(S) \overline{\int_T x(t)d\, \tau(t)} \right\} = E\left\{ Z(S) \int_R \overline{\left[\int_T f(t,a)d\, \tau(t) \right]} dZ(a) \right\},$$

if $\sigma(S)$ is finite. Since the range of $Z(S)$ is a base of $L_2(x)$, Eq. (5.30) follows from the corollary to Lemma 2.

29. Let $g(a)$ be a function which is quadratically integrable on every measurable subset S of R of finite measure $\sigma(S)$.

Then it is obvious that

$$Z'(S) = \int_S g(a)dZ(a) \tag{5.31}$$

is a spectral function, corresponding to the measure

$$\sigma'(S) = \int_S \left| g(a) \right|^2 d\sigma(a). \tag{5.32}$$

If the function $f(a)$ is quadratically integrable on R with respect to σ', then

$$\int_R f(a)dZ'(a) = \int_R f(a)g(a)dZ(a), \tag{5.33}$$

is true. This is easily seen if the first integral is approximated according to its definition by a sequence of finite sums.

If $g(a)$ is different from zero almost everywhere with respect to σ, then one can set

$$f(a) = \frac{e_s(a)}{g(a)}$$

in (5.33). Then

$$Z(S) = \int_S \frac{dZ'(a)}{g(a)}$$

and consequently $L_2(Z') = L_2(Z)$. If in particular, $\left| g(a) \right| = 1$, then by (5.32) $\sigma'(S) \equiv \sigma(S)$.

30. We now wish to consider some simple examples of spectral representations. The following § contains further applications of Theorem 10.

(1) Let $x(t)$ be a random function whose correlation function is given by (3.6). Then there exists an orthonormal system $\{z_k\}$ such that the representation

$$x(t) = \sum_k z_k f_k(t)$$

is valid. Referring back to paragraph 15, it follows that $x(t)$ is separable if and only if there exists a bilinear representation (3.6) for $r(s,t)$.

(2) Let T be the complex plane and $r(t)$ an entire analytical function, whose derivatives of all orders are real and non-negative at $t = 0$. Then for all t,

$$r(t) = \sum_{k=0}^{\infty} c_k^2 t^k$$

holds, where the c_k are real. Then $r(s\bar{t})$ is a positive-definite Hermite function. If it is the correlation function of the random function $x(t)$, then since

$$r(s\bar{t}) = \sum_{k=0}^{\infty} c_k s^k c_k \bar{t}^k$$

we have the representation

$$x(t) = \sum_{k=0}^{\infty} c_k z_k t^k \, ,$$

where $\{z_k\}$ is a complete orthonormal system in $L_2(x)$. Hence there exists no sequence of complex numbers $\{\varphi_k\}$ with $\sum_k |\varphi_k|^2 = 1$ such that

$$\sum_{k=0}^{\infty} \varphi_k c_k t^k \equiv 0.$$

In both of the above examples R is the set of all positive integer and $\sigma(S)$ is the number of elements in S.

(3) Let T be the real line. By $\xi(t)$ we denote a random function with the following properties:

$$\begin{cases} \|\xi(b) - \xi(a)\|^2 = |b - a| & \text{for all a and b,} \\ \xi(d) - \xi(c) \perp \xi(b) - \xi(a), & \text{if } a \le b \le c \le d, \qquad (5.34) \\ \xi(0) = 0 \end{cases}$$

It follows from these conditions that $\xi(s) \perp \xi(t) - \xi(s)$, if $0 < s < t$ or $t < s < 0$, and $\xi(s) \perp \xi(t)$ if $s < 0 < t$. Furthermore, we obtain $r(s,t) = r(s,s) = \|\xi(s)\|^2 = \|\xi(s) - \xi(0)\|^2$ if $0 < s < t$ or $t < s < 0$, and $r(s,t) = 0$ if $s < 0 < t$. Consequently

$$r(s,t) = \begin{cases} 0, & \text{if } s \cdot t \le 0 \\ |s|, & \text{if } s \cdot t > 0 \text{ and } |s| \le |t| \, . \end{cases}$$

We can write

$$r(s,t) = \frac{1}{2} \left[|s| + |t| - |s - t| \right]. \qquad (5.35)$$

It is seen that

$$|s| = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{\sin^2 \lambda s}{\lambda^2} \, d\lambda = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{2 - e^{2i\lambda s} - e^{-2i\lambda s}}{4\lambda^2} \, d\lambda$$

$$= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{2 - e^{i\lambda s} - e^{-i\lambda s}}{\lambda^2} \, d\lambda$$

and similarly

$$|t| = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{2 - e^{i\lambda t} - e^{-i\lambda t}}{\lambda^2} \, d\lambda \ ,$$

$$|s-t| = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{2 - e^{i\lambda(s-t)} - e^{-i\lambda(s-t)}}{\lambda^2} \, d\lambda \ .$$

Then it follows that

$$r(s,t) = \frac{1}{4\pi} \int_{-\infty}^{+\infty} \frac{2 - e^{i\lambda s} - e^{-i\lambda s} - e^{i\lambda t} - e^{-i\lambda t} + e^{i\lambda s} \cdot e^{-\lambda it} + e^{-i\lambda s} \cdot e^{i\lambda t}}{\lambda^2} \, d\lambda$$

$$= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{1 - e^{i\lambda s} - e^{-i\lambda t} + e^{i\lambda s} \cdot e^{-i\lambda t}}{\lambda^2} \, d\lambda$$

$$= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{(1 - e^{i\lambda s})(1 - e^{-i\lambda t})}{\lambda^2} \, d\lambda .$$

Therefore we can write

$$r(s,t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{e^{i\lambda s}-1}{i\lambda} \cdot \overline{\frac{e^{i\lambda t}-1}{i\lambda}} \, d\lambda \qquad (5.36)$$

By Theorem 10 there exists a spectral function Z(S) defined on the real

line, corresponding to the usual Lebesgue measure, such that

$$\xi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \frac{e^{i\lambda t} - 1}{i\lambda} \, dZ(\lambda)$$

The random function

$$\xi^*(\lambda) = \int_a^\lambda dZ(\lambda)$$

obviously has the properties (5.34). It is easily seen that $Z(S)$ is uniquely defined by $\xi^*(\lambda)$ for all Lebesgue - measurable sets.

If $f(\lambda)$ is a function which is quadratically integrable on the real line, then the integral

$$\int_{-\infty}^{+\infty} f(\lambda) d\,\xi^*(\lambda) \tag{5.37}$$

can be simply defined by means of

$$\int_{-\infty}^{+\infty} f(\lambda) d\,\xi^*(\lambda) = \int_{-\infty}^{+\infty} f(\lambda) dZ(\lambda) \tag{5.38}$$

Then

$$\xi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \frac{e^{i\lambda t} - 1}{i\lambda} d\,\xi^*(\lambda). \tag{5.39}$$

One can show that the converse to (5.39),

$$\xi^*(\lambda) = -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \frac{e^{-i\lambda t} - 1}{it} d\,\xi(t) \tag{5.40}$$

also holds. Thus it is necessary only to approximate the integral (5.40) by the corresponding finite sums, to express the ξ values by means of (5.39), and then to pass to the limit.

If $\xi(t)$ is bounded on the interval (0,1), one obtains a finite sum instead of the integral (5.39). Wiener [1] defined his "Fundamental Random Function" by means of these sums (cf. also Doob [1] , pp 134-139). He assumes that the random coefficients of the sum are normally distributed and pairwise independent, and proves that $\xi(t)$ is continuous with probability one (the same result follows from the single assumption that the coefficients are pairwise independent.)

We hope to return to similar questions in the future.

Considering $\xi(t)$ only for positive values of t, one obtains r(s, t) $= \frac{1}{2} \left[\left| s + t \right| - \left| s - t \right| \right]$, and an easy computation shows

$$r(s,t) = \frac{2}{\pi} \int_{0}^{+\infty} \frac{\sin \lambda s \, \sin \lambda t}{\lambda^2} \, d\lambda \qquad (5.41)$$

By Theorem 10, we have

$$\xi(t) = \sqrt{\frac{2}{\pi}} \int_{0}^{\infty} \frac{\sin \lambda t}{\lambda} \, d\xi_1^*(\lambda) \qquad (t \geqq 0) , \qquad (5.42)$$

where $\xi_1^*(\lambda)$ is defined for $\lambda \geqq 0$ and has the properties (5.34).

The representation (5.42) is real, i.e. if $\xi(t)$ is real, so is $\xi_1^*(\lambda)$.

VI. STATIONARY RANDOM FUNCTIONS

a. General Properties

31. Let T be the real line and τ the Lebesgue measure defined on T.

A random function $x(t)$ defined on T is called stationary if for each h,

$$E\left\{x(s + h)\overline{x(t + h)}\right\} = E\left\{x(s)\overline{x(t)}\right\} \tag{6.1}$$

i.e., if the correlation function $r(s,t)$ is a function of s - t only (Khinchin [1], Cramer [2]).[1] Instead of $r(s,t)$ we shall simply write $r(s - t)$. By (6.1) for every s we have

$$r(t) = \overline{r(-t)} = E\left\{x(s + t)\overline{x(s)}\right\} \tag{6.2}$$

and in particular

$$\left\|x(s)\right\|^2 = r(0) = \sigma^2 > 0 \tag{6.3}$$

By Theorem 4, $x(t)$ is measurable if and only if $r(t)$ is a measurable function. Then it follows from Theorem 1 that $x(t)$ is continuous if and only if

$$\lim_{t \to 0} r(t) = r(0) = \sigma^2 \tag{6.4}$$

If condition (6.4) is satisfied, $r(t)$ is continuous for all t (Khinchin [1]). Thus one may note that a (wide-sense) stationary random process is continuous at every point or discontinuous at every point.

32. Let

$$z = \sum_{k=1}^{n} c_k x(t_k) \tag{6.5}$$

---

[1]Tr. Note: This property is usually taken to define the class of processes stationary in the "wide-sense."

be an element of $L(x)$. Then

$$T_h z = \sum_{k=1}^{n} C_k x(t_k + h) \tag{6.6}$$

also defines an element of $L(x)$. One can easily show that $T_h z$ is unique (cf. para. 27). If y and z are elements in $L(x)$, then obviously

$$T_h(a y + b z) = a T_h y + b T_h z, \tag{6.7}$$

and by (6.1)

$$E(T_h y \overline{T_h z}) = E(y \bar{z}). \tag{6.8}$$

As a special case of (6.8), we have

$$\| T_h z \| = \| z \|. \tag{6.9}$$

Finally, from (6.5) and (6.8)

$$T_{h+k} z = T_h(T_k z) = T_k(T_h z) \tag{6.10}$$

The transformations T_h of the set $L(x)$ are linear and measure-preserving and form a group. It follows from (6.9) that

$$T_h z = \underset{n \to \infty}{\text{l.i.m.}} \ T_h z_n, \ z_n \text{ in } L(x), \ z = \underset{n \to \infty}{\text{l.i.m}} \ z_n$$

uniquely defines T_h for every element of $L_2(x)$. By Lemma 1, the formulas (6.7), (6.8), (6.9) and (6.10) remain valid. The extended transformation T_h are therefore unitary (cf. Nagy [1], p. 36).

Since

$$x(t) = T_t x(0)$$

instead of steadying the stationary random variable $x(t)$, one may study the properties of the corresponding group of unitary transformations T_t (cf. Doob $\begin{bmatrix}1\end{bmatrix}$, p. 123-124; Hopf $\begin{bmatrix}1\end{bmatrix}$, p. 56-65; Kolmogoroff $\begin{bmatrix}2\end{bmatrix}$). It appears simpler and more natural in many respects to work directly with the given random functions, although it is also useful to apply the transformation group.

33. Up to here we have considered only stationary continuous random functions. Many of their properties remain true for the more general class of measurable stationary random functions. The following theorem shows the correspondence between the two classes.

Theorem 12: Every measurable stationary random function $x(t)$ can be uniquely decomposed into two orthogonal stationary components $x_s(t)$ and $x_o(t)$

$$x(t) = x_s(t) + x_o(t), \qquad (6.11)$$

such that $x_s(t)$ is continuous and $x_o(t)$ is of zero-order.

Proof: It is easily seen that the decomposition is unique. If we had

$$x(t) = x_s^{(1)}(t) + x_o^{(1)}(t) = x_s^{(2)}(t) + x_o^{(2)}(t),$$

then the zero-order function $x_o^{(2)}(t) - x_o^{(1)}(t) = x_s^{(1)}(t) - x_s^{(2)}(t)$ would be continuous, which is impossible according to paragraph 21.

In order to prove the possibility of the decomposition, consider the random function

$$X(u,v) = \int_u^v x(t)dt$$

(when τ is the Lebesgue measure, we write dt instead of $d\tau(t)$). Obviously $L_2(X)$ is a subspace in $L_2(x)$. Let $x_s(t)$ be the projection of $x(t)$ on $L_2(X)$:

$$x_s(t) = P_{L_2(X)}x(t),$$

and

$$x_0(t) = x(t) - x_s(t) = P_{L_2(x)(-)L_2(X)}x(t).$$

Then $x_s(t)$ and $x_0(t)$ are orthogonal. We shall show that they satisfy the conditions of the theorem.

1. $x_s(t)$ and $x_0(t)$ are stationary. Let T_h be the transformation group defined in the previous paragraph. Then

$$T_h x_s(t) = T_h P_{L_2(X)}x(t) = P_{L_2(T_h X)}T_h x(t) = P_{L_2(T_h X)}x(t + h).$$

However,

$$T_h X(u,v) = \int_u^v x(t + h)dt = \int_{u+h}^{v+h} x(t)dt = X(u + h, v + h),$$

so $L_2(T_h X) = L_2(X)$. Then we obtain

$$T_h x_s(t) = P_{L_2(X)}x(t + h) = x_s(t + h)$$

and by (6.8)

$$E\left\{x_s(s+h)\overline{x_s(t+h)}\right\} = E\left\{T_h x_s(s)\overline{T_h x_s(t)}\right\} = E\left\{x_s(s)\overline{x_s(t)}\right\}$$

Since $x_s(s) \perp x_0(t)$, we have

$$E\left\{x_0(s+h)\overline{x_0(t+h)}\right\} = E\left\{x(s+h)\overline{x(t+h)}\right\} - E\left\{x_s(s+h)\overline{x_s(t+h)}\right\}$$

$$= E\left\{x(s)\overline{x(t)}\right\} - E\left\{x_s(s)\overline{x_s(t)}\right\} = E\left\{x_0(s)\overline{x_0(t)}\right\}$$

(2) $x_s(t)$ is continuous--by (4.2) and (6.2), for all u and v such that $|t - s| \leqq |v - u|$,

$$\left| E\left\{[x(s) - x(t)]\overline{X(u,v)}\right\}\right| = \left| E\left\{x(s)\overline{X(u,v)}\right\} - E\left\{x(t)\overline{X(u,v)}\right\}\right|$$

$$= \left|\int_u^v r(s-w)dw - \int_u^v r(t-w)dw\right| = \left|\int_{s-v}^{s-u} r(w)dw - \int_{t-v}^{t-u} r(w)dw\right|$$

$$= \left|\int_{s-u}^{t-u} r(w)dw\right| + \left|\int_{s-v}^{t-v} r(w)dw\right| \leqq 2\,\sigma^2|t - s|.$$

$E\left\{x(t)\overline{X(u,v)}\right\}$ is a continuous function of t. Thus it follows that $E\left\{x(t)\overline{z}\right\}$ is continuous for all z of $L(X)$. Now let z be any member of $L_2(X)$. If

$$z = \underset{n \to \infty}{\text{l.i.m.}} z_n, \quad z_n \in L(X),$$ then the sequence of continuous functions

$E\left\{x(t)\overline{z}_1\right\}$, $E\left\{x(t)\overline{z}_2\right\}$, ..., $E\left\{x(t)\overline{z}_n\right\}$,... converges uniformly to the function $E\left\{x(t)\overline{z}_n\right\}$, which is therefore continuous. We have used

$$\left| E\left\{x(t)\overline{z}\right\} - E\left\{x(t)\overline{z}_n\right\}\right| \leqq \|x(t)\| \cdot \|z - z_n\| = \sigma\|z - z_n\|$$

Since $x_s(t) \perp x_0(t)$, then for the correlation function $r_s(t)$ of $x_s(t)$,

$$r_s(t) = E\left\{x_s(t)\overline{x_s(0)}\right\} = E\left\{x(t)\overline{x_s(0)}\right\}$$

Because $x_s(0)$ is an element of $L_2(X)$, $r_s(t)$ and therefore $x_s(t)$ are continuous.

(3) $x_0(t)$ is of zero-order. For arbitrary a and b, the integral

$$\int_a^b x_0(t)dt$$

belongs to $L_2(x)(-)L_2(X)$ if $x_0(t)$ belongs to this space. Since

$$\int_a^b x_0(t)dt = \int_a^b x(t)dt - \int_a^b x_s(t)dt = X(a,b) - \int_a^b x_s(t)dt$$

belongs to $L_2(X)$ also, it must vanish identically. Q.E.D.

If $x(t)$ is separable, so is $x_0(t)$. By Theorem 8, $x_0(t)$ vanishes almost everywhere. Since $\|x_0(t)\|$ is a constant, this is possible only if $x_0(t)$ vanishes identically. Consequently,

Theorem 13: A measurable, stationary random function is separable if and only if it is continuous.

34. Khinchin[1] has shown that the statistical ergodic theorem holds for continuous stationary random variables (cf. Hopf [1] , p. 28), i.e., that

$$\underset{|v-u|\longrightarrow \infty}{\text{l.i.m.}} \frac{1}{v-u} \int_u^v x(t)dt \tag{6.12}$$

exists.

It follows from Theorem 12 that the ergodic theorem still holds if one assumes that $x(t)$ is measurable rather than stationary, if the integral is defined by §4 (Khinchin did not state what he understood by the integral in (6.12).

The Khinchin theory depends upon a spectral theory of the correlation function, which we shall further consider below. Here we wish to prove the ergodic theory by another method without the aid of the spectral

representation. The basic ideas of the proof are due to F. Riesz (cf.

Hopf [1], p. 23). In addition to their clarity, they possess the advantage

that the continuity of the random functions need not be assumed.

Let $x(t)$ be a measurable stationary random function. Then,

$$y(t,h) = x(t + h) - x(t) \tag{6.13}$$

is also a measurable stationary random function of τ for all h. So long as

$|v - u| \geq |h|$, we have

$$\int_u^v y(t,h)dt = \int_u^v x(t+h)dt - \int_u^v x(t)dt = \int_v^{v+h} x(t)dt - \int_u^{u+h} x(t)dt$$

and consequently

$$\left\| \frac{1}{v-u} \int_u^v y(t,h)dt \right\| \leq \frac{1}{|v-u|} \left\{ \left\| \int_v^{v+h} x(t)dt \right\| + \left\| \int_u^{u+h} x(t)dt \right\| \right\} \leq \frac{2\,h\,\sigma}{|v-u|}.$$

For all h,

$$\underset{|v-u| \to \infty}{\text{l.i.m.}} \quad \frac{1}{v-u} \int_u^v y(t,h)dt = 0. \tag{6.14}$$

Since $T_t y(s,h) = y(s + t,h)$ by (6.13) and (6.6), we also obtain

$$\underset{|v-u| \to \infty}{\text{l.i.m.}} \quad \frac{1}{v-u} \int_u^v T_t y(s,h)dt = 0$$

for all values of s and h. More generally, for every element $y = \sum_{k=1}^{n} c_k y(s_k,h_k)$

of $L(y)$

$$\underset{|v-u|\to\infty}{\text{l.i.m.}} \quad \frac{1}{v-u} \int_u^v T_t y\, dt = 0. \tag{6.15}$$

The same equation holds for every $y \epsilon h_2(y)$. Therefore, for every y there exists a $y_\epsilon \epsilon L(y)$ such that $\|y - y_\epsilon\| < \epsilon$. By (6.9) we also have $\|T_t y - T_t y_\epsilon\| < \epsilon$ and therefore

$$\left\| \frac{1}{v-u} \int_u^v T_t y\, dt - \frac{1}{v-u} \int_u^v T_t y_\epsilon\, dt \right\| < \epsilon,$$

from which the assertion follows:

Let

$$x_1(t) = P_{L_2(y)} x(t). \tag{6.16}$$

It follows from (6.13) that $L_2(T_t y) = L_2(y)$, and therefore

$$T_1 x_1(s) = P_{L_2(T_t y)} T_t x(s) = P_{L_2(y)} x(s + t) = x_1(s + t)$$

By (6.8), $x_1(t)$ is stationary. Furthermore,

$$x_1(t) = T_t x_1(0).$$

Since $x_1(0)$ is an element of $L_2(y)$, from (6.15) we have

$$\underset{|v-u|\to\infty}{\text{l.i.m.}} \quad \frac{1}{v-u} \int_u^v x_1(t)\, dt = 0. \tag{6.17}$$

We write

$$x_2(t) = x(t) - x_1(t) = P_{L_2(x)(-)L_2(y)} x(t).$$

Since $x_2(t) \perp L_2(y)$, we have

$$0 = E\left\{y(t,h)\overline{x_2(t)}\right\} = E\left\{x(t+h)\overline{x_2(t)}\right\} - E\left\{x(t)\overline{x_2(t)}\right\}$$

$$= E\left\{x(t+h)\overline{x_2(t)}\right\} - \left\|x_2(t)\right\|^2.$$

Because $x(t)$ is obviously stationary, so is $\left\|x_2(t)\right\|$ and thus $E\left\{x(t+h)\overline{x_2(t)}\right\}$ is a constant. Consequently for all s and t

$$E\left\{x(s)\overline{x_2(t)}\right\} = E\left\{x(s)\overline{x_2(0)}\right\}.$$

It follows from Lemma 2 that $x_2(t) = x_2(0)$, and denoting $x_2(0)$ by z_0,

$$x(t) = x_1(t) + z_0, \quad z_0 \perp x_1(t). \tag{6.18}$$

Then from (6.17),

$$z_0 = \underset{|v-u| \to \infty}{\text{l.i.m.}} \frac{1}{v-u} \int_u^v x(t)dt. \tag{6.19}$$

This completes the proof of Khinchin's theorem.

Slutsky[2] has given the representation (6.18) for continuous $x(t)$. His proof is essentially the same as Kolmogorov's.

35. If $x(t)$ is stationary, so is $e^{-i\lambda t}x(t)$ for real λ. Then we have

$$E\left\{e^{-i\lambda(s+h)}x(s+h) \overline{e^{-i\lambda(t+h)}x(t=h)}\right\} = e^{-i\lambda(s-t)}E\left\{x(s+h)\overline{x(t+h)}\right\}$$

$$= E\left\{e^{-i\lambda s}x(s)\overline{e^{-i\lambda t}x(t)}\right\}.$$

By (6.19) we have the existence of

$$z_\lambda = \underset{|v-u| \to \infty}{\text{l.i.m.}} \frac{1}{v-u} \int_u^v e^{-i\lambda t}x(t)dt. \tag{6.20}$$

By (6.18), $e^{-i\lambda t}x(t) - z_\lambda$ and therefore $x(t) - z_\lambda e^{i\lambda t}$ are stationary and hence

$$z_\lambda \perp x(t) - z_\lambda e^{i\lambda t}. \tag{6.21}$$

Because

$$\lim_{|v-u| \to \infty} \frac{1}{v-u} \int_u^v e^{-i\lambda t} dt = 0 \quad \text{for } \lambda \neq 0$$

we obtain, for $\lambda \neq 0$

$$\lim_{|v-u| \to \infty} \frac{1}{v-u} \int_u^v e^{-i\lambda t} x_1(t) dt = \lim_{|v-u| \to \infty} \frac{1}{v-u} \int_u^v e^{-i\lambda t} x(t) dt = z_\lambda.$$

It follows from (6.18) that $z_\lambda \perp z_0$. If we consider $e^{-i\lambda t} x(t)$ instead of $x(t)$, we obtain the more general result

$$z_\lambda \perp z_{\lambda'}, \text{ when } \lambda \neq \lambda'. \tag{6.22}$$

$x(t) - \sum_{k=1}^{n} z_{\lambda_k} e^{i\lambda_k t}$ is stationary by (6.21) and (6.22), and therefore

$$z_{\lambda_m} \perp x(t) - \sum_{k=1}^{n} z_{\lambda_k} e^{i\lambda_k t} \quad (m = 1, 2, \ldots, n)$$

Then it follows that

$$\left\| x(t) - \sum_{k=1}^{n} z_{\lambda_k} e^{i\lambda_k t} \right\|^2 = \sigma^2 - \sum_{k=1}^{n} \left\| z_{\lambda_k} \right\|^2 \geq 0.$$

Hence z_λ can be different from zero for at most a denumerable number of values of λ.

We denote these values by $\lambda_1, \lambda_2, \ldots, \lambda_n, \ldots$ Then the series $\sum_k z_{\lambda_k} e^{i\lambda_k t}$ converges, and

$$x(t) - \sum_k z_{\lambda_k} e^{i\lambda_k t}$$

is stationary (cf. Slutsky [2] , where the corresponding result for real

$x(t)$ is proved using Khinchin's method).

Obviously,

$$T_h \left(\frac{1}{v-u} \int_u^v e^{-i\lambda t} x(t) dt \right) = \frac{1}{v-u} \int_u^v e^{-i\lambda t} x(t+h) dt$$

$$= e^{i\lambda h} \frac{1}{v-u} \int_{u+h}^{v+h} e^{-i\lambda t} x(t) dt$$

and therefore,

$$T_h z_\lambda = e^{i\lambda h} z_\lambda . \qquad (6.23)$$

These two equations uniquely determine $z_\lambda \epsilon L_2(x)$ to within a numerical

factor. Then it follows from (6.23)

$$E\left\{ T_h z_\lambda \overline{x(t)} \right\} = e^{i\lambda h} E\left\{ z_\lambda \overline{x(t)} \right\} .$$

However, by (6.8)

$$E\left\{ T_h z_\lambda \overline{x(t)} \right\} = E\left\{ z_\lambda \overline{T_{-h} x(t)} \right\} = E\left\{ z_\lambda \overline{x(t-h)} \right\} .$$

For all h , it is true that

$$E\left\{ z_\lambda \overline{x(h)} \right\} = e^{-i\lambda h} E\left\{ z_\lambda \overline{x(0)} \right\}$$

and consequently

$$E\left\{ \frac{z_\lambda}{E\left\{ z_\lambda \overline{x(0)} \right\}} \overline{x(h)} \right\} = e^{-i\lambda h} .$$

By Lemma 2, $\dfrac{z_\lambda}{E\left\{z_\lambda \overline{x(0)}\right\}}$ is uniquely defined by this equation.

It follows from the above that z_0 (to within a scale factor) is the only element in $h_2(x)$ invariant under T_h (cf. Hopf [1] , p. 21-25).

Every number λ for which z_λ is different from zero is called an eigenfrequency of the stationary random function $x(t)$. We have seen that the number of eigenfrequencies is finite or countable.

b. The Cramér Representation

36. We consider a continuous random function $x(t)$ with correlation function $r(t)$, which must also be continuous. For arbitrary real a and b and for an arbitrary function $\varphi(t)$ continuous in the open interval (a,b), by (4.7) we have

$$\int_a^b \int_a^b r(s-t)\, \varphi(s)\, \overline{\varphi(t)}\, ds\, dt = \left\| \int_a^b \varphi(t) x(t)\, dt \right\|^2 \geqq 0.$$

By a well-known theorem of Bochner's [1] it follows (cf. Khinchin [1]) that $r(t)$ may be represented in the form

$$r(t) = \int_{-\infty}^{+\infty} e^{i\lambda t}\, dF(\lambda) \tag{6.24}$$

where $F(\lambda)$ is a real non-decreasing function of bounded variation. Since $r(0) = \sigma^2$,

$$F(+\infty) - F(-\infty) = \sigma^2. \tag{6.25}$$

Conversely, for every function of the form (6.24),

$$\sum_{j=1}^n r(t_i - t_j) a_i a_j = \int_{-\infty}^{+\infty} \sum_{i,j=1}^n e^{i\lambda t_i} e^{-i\lambda t_j} a_i \bar{a}_j\, dF(\lambda) = \int_{-\infty}^{+\infty} \left| \sum_{k=1}^n a_k e^{i\lambda t_k} \right|\, dF(\lambda) \geqq 0.$$

so that by paragraph 14, every such function can be considered as the correlation function of a continuous stationary random process (Khinchin   1   ).

Set

$$\sigma(S) = \int_S dF(\lambda).$$

so that we can write

$$r(s-t) = \int_{-\infty}^{+\infty} e^{i\lambda s} \cdot \overline{e^{i\lambda t}} d\sigma(\lambda)$$

by use of (6.24). According to Theorem 10, there exists a spectral function $Z(S)$, corresponding to the measure $\sigma(S)$, such that

$$x(t) = \int_{-\infty}^{+\infty} e^{i\lambda t} dZ(\lambda) \tag{6.26}$$

(Cramér $\begin{bmatrix} 3 \end{bmatrix}$). Writing

$$Z(\lambda) = \frac{1}{2} \left[\int_{-0}^{\lambda-0} dZ(\lambda) + \int_{+0}^{\lambda+0} dZ(\lambda) \right], \tag{6.27}$$

we obtain using (5.3),

$$\| Z(b) - Z(a) \|^2 = F(b) - F(a), \text{ if } b \geqq a. \tag{6.28}$$

At the points of discontinuity of $F(\lambda)$, we define

$$F(\lambda) = \frac{1}{2} \left[F(\lambda-0) + F(\lambda+0) \right]$$

Let $\varphi(\lambda)$ be any function quadratically integrable with respect to $F(\lambda)$ and possessing the property

$$\int_{-\infty}^{+\infty} e^{i\lambda t} \varphi(\lambda) dF(\lambda) = 0 \tag{6.29}$$

identically in t. Then, for any trigonometric polynomial $P(\lambda)$,

$$\int_{-\infty}^{+\infty} P(\lambda)\varphi(\lambda)dF(\lambda) = 0 \cdot \tag{6.30}$$

Let (a, b) be an arbitrary interval. Since $\varphi(\lambda)$ is quadratically integrable, one may choose a number m such that

$$\int_{-\infty}^{-m} |\varphi(\lambda)|^2 dF(\lambda) + \int_{+m}^{+\infty} |\varphi(\lambda)|^2 dF(\lambda) < \epsilon^2$$

and $-m < a < b < m$. It is well known that there exists a uniformly bounded sequence of trigonometric polynomial $P_1(\lambda), P_2(\lambda), \ldots, P_n(\lambda), \ldots$ which converges to unity for $-m < \lambda < a$ and converges to zero for $b < \lambda < m$ as well as for $-m < \lambda < a$. Each $P_n(\lambda)$ is quadratically integrable with respect to $F(\lambda)$ by (6.25), and therefore

$$\lim_{n \to \infty} \int_{-m}^{+m} P_n(\lambda)\varphi(\lambda)dF(\lambda) = \int_a^b \varphi(\lambda)dF(\lambda).$$

holds. The function $P_n(\lambda)$ may be chosen so that $P_n(\lambda) < 2$, for instance, for any n and all values of λ. Then the Schwarz inequality holds

$$\left| \int_{-\infty}^{-m} P_n(\lambda)\varphi(\lambda)dF(\lambda) \right|^2 < \int_{-\infty}^{-m} 4 \, dF(\lambda) \int_{-\infty}^{-m} |\varphi(\lambda)|^2 dF(\lambda) < 4 \, \sigma^2 \epsilon^2,$$

we have

$$\left| \int_{-\infty}^{-m} P_n(\lambda)\varphi(\lambda)dF(\lambda) \right| < 2\sigma\epsilon$$

and similarly

$$\left| \int_{+m}^{+\infty} P_n(\lambda)\varphi(\lambda)dF \right| < 2\sigma\epsilon .$$

According to (6.30),

$$\left| \int_{-m}^{+m} P_n(\lambda)\varphi(\lambda)dF(\lambda) \right| < 4\ \sigma\varepsilon$$

and consequently

$$\left| \int_a^b \varphi(\lambda)dF(\lambda) \right| = \lim_{n\to\infty} \left| \int_{-m}^{+m} P_n(\lambda)\varphi(\lambda)dF(\lambda) \right| \leq 4\ \sigma\varepsilon.$$

Since ε can be chosen arbitrarily small, it follows that

$$\int_a^b \varphi(\lambda)dF(\lambda) = 0$$

for all a and b and thus $\varphi(\lambda) = 0$ almost everywhere with respect to $F(\lambda)$.

For all $\varphi(\lambda)$ with the property (6.29) we have

$$\int_{-\infty}^{+\infty} |\varphi(\lambda)|^2 dF(\lambda) = 0\ .$$

Consequently from Theorem 10

$$L_2(Z) = L_2(x). \qquad (6.31)$$

37. The converse of (6.26) holds:

$$Z(\lambda) = \underset{t\to\infty}{\text{l.i.m.}}\ \frac{1}{2\pi} \int_{-t}^{+t} \frac{1-e^{-i\lambda t}}{it}\ x(t)dt\ . \qquad (6.32)$$

Then by Theorem 11,

$$\frac{1}{2\pi} \int_{-t}^{+t} \frac{1-e^{-i\lambda t}}{it}\ x(t)dt = \int_{-\infty}^{+\infty} \left[\frac{1}{2\pi} \int_{-t}^{+t} \frac{1-e^{-i\lambda t}}{it}\ e^{i\mu t}\ dt \right] dZ(\mu)$$

$$= \int_{-\infty}^{+\infty} \psi(\lambda,\ \mu,\ t)dZ(\mu).$$

where

$$\psi(\lambda, \mu, t) = \frac{1}{2\pi} \int_{-t}^{+t} \frac{1-e^{-i\lambda t}}{it} e^{i\mu t} dt$$

for all values of λ and t bounded, and thus by (6.25) ψ is quadratically integrable with respect to $F(\mu)$. We have

$$\lim_{t \to \infty} \psi(\lambda, \mu, t) = \lim_{t \to \infty} \frac{1}{2\pi} \int_{-t}^{+t} \frac{1-e^{-i\lambda t}}{it} e^{i\mu t} dt = \lim_{t \to \infty} \frac{1}{2\pi} \int_{-t}^{+t} \frac{e^{i\mu t} - e^{i(\mu-\lambda)t}}{it} dt$$

$$= \lim_{t \to \infty} \frac{1}{2\pi} \int_{-t}^{+t} \frac{e^{i\mu t} - e^{-i\mu t} - (e^{i(\mu-\lambda)t} - e^{-i(\mu-\lambda)t})}{2 it} dt$$

$$= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{\sin \mu t - \sin (\mu-\lambda)t}{t} dt$$

$$= \begin{cases} 1, & \text{if } 0 < \mu < \lambda \\[2mm] \frac{1}{2}, & \text{if } \mu = \lambda > 0 \text{ or } 0 = \mu < \lambda \\[2mm] 0, & \text{if } \mu > 0, \mu > \lambda \text{ or } \mu = \lambda = 0 \text{ or } \mu < 0, \mu < \lambda \\[2mm] -\frac{1}{2} & \text{if } \mu = \lambda < 0 \text{ or } 0 = \mu > \lambda \\[2mm] -1, & \text{if } 0 > \mu > \lambda \end{cases}$$

Denoting these limits by $\psi(\lambda, \mu)$, we have

$$\lim_{t \to \infty} \int_{-\infty}^{+\infty} |\psi(\lambda, \mu) - \psi(\lambda, \mu, t)|^2 dF(\mu)$$

$$= \int_{-\infty}^{+\infty} \lim_{t \to \infty} |\psi(\lambda, \mu) - \psi(\lambda, \mu, t)|^2 dF(\mu) = 0.$$

It now follows from Theorem 9 that

$$\underset{t \to \infty}{\mathrm{l.i.m.}} \int_{-\infty}^{+\infty} \gamma(\lambda, \mu, t)dZ(\mu) = \int_{-\infty}^{+\infty} \gamma(\lambda, \mu)dZ(\mu)$$

$$= \frac{1}{2}\left[Z(\lambda+0) + Z(\lambda-0)\right] - \frac{1}{2}\left[Z(+0) + Z(-0)\right] = Z(\lambda).$$

Hence formula (6.32) is proven.

38. Consider the expression (6.20). By Theorem 11 and (6.26) we have

$$\frac{1}{v-u}\int_u^v e^{-i\lambda t}x(t)dt = \int_{-\infty}^{+\infty}\left[\frac{1}{v-u}\int_u^v e^{i(\mu-\lambda)t}dt\right]dZ(\mu)$$

$$= \int_{-\infty}^{+\infty} Z(\mu-\lambda, u, v)dZ(\mu).$$

where

$$X(\lambda, u, v) = \frac{1}{v-u}\int_u^v e^{i\lambda t} dt = \begin{cases} \dfrac{e^{i\lambda v}-e^{i\lambda u}}{i\lambda(v-u)}, & \text{if } \lambda \neq 0 \\ 1, & \text{if } \lambda = 0 \end{cases}$$

for all values of λ, u and bounded v. $X(\lambda, u, v)$ is thus quadratically integrable with respect to $F(\lambda)$.

We can write

$$X(\lambda) = \lim_{v-u \to \infty} X(\lambda, u, v) = \begin{cases} 0, & \text{if } \lambda \neq 0 \\ 1, & \text{if } \lambda = 0. \end{cases}$$

$X(\lambda)$ is quadratically integrable with respect to $F(\lambda)$. As in the above paragraph, we obtain

$$\underset{\substack{v-u \to \infty}}{\text{l.i.m.}} \int_{\infty}^{+\infty} X(\mu-\lambda, u, v)dZ(\mu) = \int_{-\infty}^{+\infty} X(\mu-\lambda)dZ(\mu) = Z(\lambda+0) - Z(\lambda-0).$$

Consequently

$$z_\lambda = Z(\lambda+0) - Z(\lambda-0) \tag{6.33}$$

and by (6.28)

$$|z_\lambda|^2 = F(\lambda+0) - F(\lambda-0). \tag{6.34}$$

Thus the eigenfrequencies of x(t) are the points of discontinuity of $F(\lambda)$.

Multiplying both sides by the scalar quantity $x(0) = Z(+\infty) - Z(-\infty)$, it follows from equations (6.32), (6.33) and (6.20) that

$$F(\lambda) - F(0) = \lim_{t \to \infty} \frac{1}{2\pi} \int_{-t}^{+t} \frac{1-e^{-i\lambda t}}{it} r(t)dt, \tag{6.35}$$

$$F(\lambda+0) - F(\lambda-0) = \lim_{|v-u| \to \infty} \frac{1}{v-u} \int_u^v e^{-i\lambda t}r(t)dt. \tag{6.36}$$

At every point of discontinuity in (6.35), we set

$$F(\lambda) = \frac{1}{2}\left[F(\lambda+0) + F(\lambda-0) \right].$$

39. If r(t) is real, instead of (6.24) we can write

$$r(t) = \int_0^\infty \cos \lambda t dG(\lambda) \tag{6.37}$$

where

$$G(\lambda) = F(\lambda) - F(-\lambda) \tag{6.38}$$

(Khinchin [1]). It follows from (6.37) that

$$r(s-t) = \int_0^\infty (\cos \lambda s \cos \lambda t + \sin \lambda s \sin \lambda t) dG(\lambda).$$

By Theorem 10, there exist two spectral functions $Z_1(\lambda)$ and $Z_2(\lambda)$ such that

$$Z_1(\lambda) \perp Z_2(\lambda), \quad \left| Z_1(\lambda) \right|^2 = \left| Z_2(\lambda) \right|^2 = G(\lambda) \tag{6.39}$$

and the representation

$$x(t) = \int_0^\infty \cos \lambda t dZ_1(\lambda) + \int_0^\infty \sin \lambda t dZ_2(\lambda) \tag{6.40}$$

holds for $x(t)$ (Doob [3], he takes the lower limit of the integral to be $-\infty$. This is not convenient, because then $Z_1(\lambda)$ and $Z_2(\lambda)$ are not uniquely defined). It is easily shown that

$$L_2(x) = L_2(Z_1)(+) L_2(Z_2)$$

If $x(t)$, rather than $r(t)$, is real, it follows from (6.32) that

$$Z(-\lambda) = -\overline{Z(\lambda)}.$$

By an easy computation we obtain

$$\begin{cases} Z_1(\lambda) = Z(\lambda) - Z(-\lambda) = 2\, RZ(\lambda), \\ Z_2(\lambda) = i\left[Z(\lambda) + Z(-\lambda) \right] = -2IZ(\lambda), \end{cases}$$

and thus

$$Z_1(\lambda) = \underset{t \to \infty}{\text{l.i.m.}} \frac{1}{\pi} \int_{-t}^{+t} \frac{\sin\lambda t}{t} x(t)dt,$$

$$Z_2(\lambda) = \underset{t \to \infty}{\text{l.i.m.}} \frac{1}{\pi} \int_{-t}^{+t} \frac{1-\cos\lambda t}{t} x(t)dt. \qquad (6.41)$$

From equations (6.20) and (6.26),

$$z_\lambda' = Z_1(\lambda+0) - Z_1(\lambda-0) = \underset{|v-u| \to \infty}{\text{l.i.m.}} \frac{1}{v-u} \int_u^v x(t) \cos \lambda t \, dt,$$

$$z_\lambda'' = Z_2(\lambda+0) - Z_2(\lambda-0) = \underset{|v-u| \to \infty}{\text{l.i.m.}} \frac{1}{v-u} \int_u^v x(t) \sin \lambda t \, dt. \qquad (6.42)$$

Slutsky [2] has proved the existence of the right-hand integrals (cf. $\oint 35$).

40. From the above results and Theorem 12, we directly obtain

Theorem 13: Every complex measurable stationary random function x(t) can be expressed in the form

$$x(t) = z_o + \sum_{k=1}^{\infty} z_k e^{i\lambda_k t} + x_1(t) + x_o(t) \qquad (6.43)$$

If x(t) is real, the representation may be written

$$x(t) = z_0 + \sum_{k=1}^{x} (z_k' \cos \lambda_k t + z_k'' \sin \lambda_k t) + x_1(t) + x_0(t) \qquad (6.43)'$$

The components z_0, $z_k e^{i\lambda_k t}$ (or $z_k' \cos \lambda_k t$ and $z_k'' \sin \lambda_k t$), $x_1(t)$ and $x_0(t)$ are all stationary and mutually orthogonal; $x_1(t)$ is continuous and possesses no eigenfrequency, and $x_0(t)$ is of zero-order. The eigenfrequencies λ_k are the points of discontinuity of the distribution function $F(\lambda)$ defined in (6.35), while the random variables z_k $\left(z_k', z_k'' \right)$ are defined by (6.20) [(6.42)].

If x(t) is continuous and $F(\lambda)$ is purely a step function, then by (6.24) r(t) is an almost periodic function of Bohr's. By the above theorem and (6.34) we obtain

$$\| x(t) - \sum_{k=0}^{\infty} z_k e^{i\lambda_k t} \|^2 = \| x(t) \|^2 - \sum_{k=0}^{\infty} \| z_k \|^2$$

$$= F(+\infty) - F(-\infty) - \sum_{k=0}^{\infty} \left[F(\lambda_k + 0) - F(\lambda_k - 0) \right] = 0,$$

so the components $x_1(t)$ and $x_0(t)$ of (6.43) vanish and

$$x(t) = \sum_{k=0}^{\infty} z_k e^{i\lambda_k t}$$

Because r(t) is almost periodic, by a well-known theorem of Bohr's [1] for every $\epsilon > 0$, there is a positive number $\ell(\epsilon)$ such that for each integer, n, there

exists a point t in the interval $(n\ell(\epsilon), (n+1)\ell(\epsilon))$ such that

$$\left| r(0) - r(t) \right| < \frac{1}{2}\epsilon^2.$$

Since

$$\left\| x(s+t) - x(s) \right\|^2 = 2r(0) - 2R(t) \leq 2 \left| r(0) - r(t) \right|$$

it is true for all s that

$$\left\| x(s+t) - x(s) \right\| < \epsilon,$$

and x(t) is almost periodic in the mean (in the sense of Bohr). One can show that x(t) is, with probability one, an almost periodic function of Besicovitch (Slutsky[2]).

If conversely x(t) is almost periodic in the above sense, we note that $F(\lambda)$ must be a pure step function. Thus we have

Theorem 14. A continuous stationary random function x(t) is almost periodic in the mean (in the sense of Bohr) if and only if the distribution function defined by (6.35) is purely a step function.

c. Stationary Random Functions with Absolutely Continuous Spectra

40. Continuous stationary random variables with absolutely continuous distribution functions $F(\lambda)$ constitute an important class. We say that such a function possesses an absolutely continuous spectrum.

We wish to investigate this class of functions more closely. We shall assume

$$F(\lambda) = \int_0^\lambda F'(\lambda)\, d\lambda \qquad (6.44)$$

where $F'(\lambda)$ is almost everywhere equal to the derivative of $F(\lambda)$. Then by (6.24) we have

$$r(t) = \int_{-\infty}^{+\infty} e^{i\lambda t} F'(\lambda)\, d\lambda \qquad (6.45)$$

By the well known Riemann-Lebesgue theorem it follows that

$$\lim_{t \to \infty} r(t) = 0 \qquad (6.46)$$

Conversely, the representation (6.45) is known to hold if, for example, $r(t)$ monotonically tends toward zero or if $r(t)$ is integrable on the interval $(-\infty, \infty)$ and is of bounded variation on every finite interval (cf. Bochner [1]).

Since $F(\lambda)$ is non-decreasing, one may assume that $F'(\lambda)$ is never-negative. Then the function

$$f(\lambda) = +\sqrt{F'(\lambda)} \qquad (6.47)$$

is real and quadritically integrable on the interval $(-\infty, \infty)$. By Plancherel's theorem (cf. Bochner [1]) $f(\lambda)$ has a Fourier transform

$$g(a) = \text{l.i.m.}_{\lambda \to \infty} \frac{1}{\sqrt{2\pi}} \int_{-\lambda}^{+\lambda} e^{i\lambda a} f(\lambda)\, d\lambda. \qquad (6.48)$$

The function $g(a)$ is quadratically integrable on $(-\infty, \infty)$. Obviously, the Fourier transform $g_t(a)$ of the function $f_t(\lambda) = e^{i\lambda t} f(\lambda)$ is equal to $g(a+t)$. By the well known Parseval Theorem

$$\int_{-\infty}^{+\infty} g_t(a) g(a) da = \int_{-\infty}^{+\infty} f_t(\lambda) f(\overline{\lambda}) \, d\lambda$$

and by (6.47) it follows that

$$\int_{-\infty}^{+\infty} g(a+t) g(a) da = \int_{-\infty}^{+\infty} e^{i\lambda t} |f(\lambda)|^2 \, d\lambda = \int_{-\infty}^{+\infty} e^{i\lambda t} F'(\lambda) \, d\lambda. \qquad (6.49)$$

By (6.45) we have the representation for $r(t)$

$$r(t) = \int_{-\infty}^{+\infty} g(a+t) \overline{g(a)} da. \qquad (6.50)$$

Conversely, of the correlation function of the continuous stationary random function $x(t)$ can be represented in the form (6.50), $x(t)$ has an absolutely continuous spectrum. Then $g(a)$ is quadratically integrable since

$$\int_{-\infty}^{+\infty} |g(a)|^2 da = r(0)$$

If $f(\lambda)$ is the Fourier transform of the correlation function, and $F'(\lambda)$ is set equal to $|f(\lambda)|^2$, equations (6.49) and (6.45) hold.

It follows from (6.50) that

$$r(s-t) = \int_{-\infty}^{+\infty} g(a+s-t)\overline{g(a)}\,da = \int_{-\infty}^{+\infty} g(a+s)\overline{g(a+t)}\,da.$$

By Theorem 10 the representation of $x(t)$,

$$x(t) = \int_{-\infty}^{+\infty} g(a+t)\,d\,\zeta(a), \tag{6.51}$$

is valid, where the random function $\zeta(a)$ satisfies the conditions (5.34).

Let $\varphi(a)$ be a quadratically integrable function on $(-\infty, +\infty)$, and $\psi(\lambda)$ be its Fourier transform. By the Parseval theorem we obtain

$$\int_{-\infty}^{+\infty} g(a+t)\overline{\varphi(a)}\,da = \int_{-\infty}^{+\infty} e^{i\lambda t}f(\lambda)\overline{\psi(\lambda)}\,d\lambda. \tag{6.52}$$

Because the functions $f(\lambda)$ and $\psi(\lambda)$ are quadratically integrable, $f(\lambda)\overline{\psi(\lambda)}$ is absolutely integrable and defined uniquely almost every where by its Fourier transform (6.52) (cf. Bochner$\lfloor 1 \rfloor$, p. 47). Thus it follows that the expression (6.52) vanishes for all t if and only if

$$f(\lambda)\overline{\psi(\lambda)} = 0$$

almost everywhere. If $|f(\lambda)|^2 = F'(\lambda) > 0$ almost everywhere, this is possible only if $\psi(\lambda)$ and therefore $\varphi(a)$ vanish almost everywhere. However,

if $F'(\lambda) = 0$ on a set M of positive mass, then one can choose $\psi(\lambda) \neq 0$ on M, and $\psi(\lambda) = 0$ otherwise, such that

$$\int_{-\infty}^{+\infty} |\varphi(a)|^2 da = \int_{-\infty}^{+\infty} |\psi(\lambda)|^2 d\lambda = 1$$

From the above and Theorem 10, it follows that $L_2(\xi) = L_2(x)$ if and only if $F'(\lambda) = 0$ almost everywhere.

41. We say that the spectrum of a stationary continuous random function is complete if there is no set M of positive Lebesgue measure such that

$$\int_M dF(\lambda) = 0$$

Let $\varphi(t)$ be an absolutely integral random function on $(-\infty, +\infty)$ which does not vanish almost everywhere. We consider the expression

$$\int_{-\infty}^{+\infty} \varphi(t)x(t)\, dt.$$

By (4.7) and (6.24),

$$\left\| \int_{-\infty}^{+\infty} \varphi(t)x(t)dt \right\|^2 = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} r(u-v)\, \varphi(u)\overline{\varphi(v)}dudv$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \left\{ \int_{-\infty}^{+\infty} e^{i\lambda(u-v)}dF(\lambda) \right\} \varphi(u)\overline{\varphi(v)}dudv.$$

Since the convergence is absolute and uniform, the order of the

integrations can be reversed. Then we obtain

$$\left\| \int_{-\infty}^{+\infty} \varphi(t)x(t)dt \right\|^2 = \int_{-\infty}^{+\infty} \left\{ \int_{-\infty}^{+\infty} e^{i\lambda u}\varphi(u)du \right\} \left\{ \int_{-\infty}^{+\infty} e^{-i\lambda v}\overline{\varphi(v)}dv \right\} dF(\lambda)$$

$$= 2\pi \int_{-\infty}^{+\infty} \left| \psi(\lambda) \right|^2 dF(\lambda), \tag{6.53}$$

where $\phi(t)$ denotes the Fourier transform of $\psi(\lambda)$.

If the spectrum of $x(t)$ is complete, the last expression is always

positive. If the spectrum is not complete, there is a set M of positive

measure such that $\int_M dF(\lambda) = 0$. One can choose $\psi(\lambda) \neq 0$ on M, $\psi(\lambda) = 0$

elsewhere, such that (6.53) is equal to zero without $\varphi(t)$ vanishing almost

everywhere. Thus we see that the spectrum of $x(t)$ is complete if and only

if there exists no absolutely integrable function $\varphi(t)$, not vanishing

almost everywhere, such that

$$\int_{-\infty}^{+\infty} \varphi(t)x(t)dt = 0 \tag{6.54}$$

From (6.54) it follows more generally that

$$\int_{-\infty}^{+\infty} \varphi(t)x(s+t)dt = 0$$

for all values of s.

As a result of this and the previous paragraph, we have

Theorem 15: A continuous stationary random function can be expressed in the form (6.51) if and only if its spectrum is absolutely continuous. $L_2(\xi) = L_2(x)$ if and only if the spectrum is complete. The latter condition is equivalent to this: that there exist no function $\varphi(t)$, absolutely integrable on the interval $(-\infty, +\infty)$ and not vanishing almost everywhere, which possesses the property (6.54).

42. The representation (6.51) can obviously be considered as a generalization of the method of moving averages, which is often used in mathematical statistics for the treatment of stationary time series (cf. Wold $[1]$, for instance). Particularly important is the special case in which $x(t)$ depends only upon the "earlier" values of the auxiliary process $\xi(a)$. We must reformulate the expression (6.51) somewhat in order to give a precise meaning to "earlier". If we denote

$$\xi_1(a) = \xi(-a)$$

we can write

$$x(t) = \int_{-\infty}^{+\infty} g(a)d_a \xi_1(t-a) \qquad (6.55)$$

Now we consider the values $x(t)$ and $\xi_1(t)$ to be "simultaneous," and say that $\xi_1(t-a)$ is "earlier" than $x(t)$ if $a > 0$. We wish to investigate under what conditions $\xi_1(t-a)$ and $g(a)$ may be chosen such that only positive values of a appear in (6.55), i.e., that $g(a)$ disappears for $a < 0$.

Let $f(\lambda)$ be the Fourier transform of $g(a)$. We have seen that $f(\lambda)$ must satisfy the condition

$$\left| f(\lambda) \right|^2 = F'(\lambda) \tag{6.56}$$

It is necessary to ascertain when $f(\lambda)$ may be chosen such that its Fourier transform vanishes for negative values of the argument. The following theorem holds (Paley and Wiener[1], Theorem XII):

Let $\varphi(\lambda)$ be a real non-negative function which is quadratically integrable on the interval $(-\infty, +\infty)$ and which does not vanish almost everywhere. A necessary and sufficient condition for the existence of a function $g(t)$, which vanishes for $t \leq t_0$ and which possesses a Fourier transform $f(\lambda)$ satisfying the condition $\left| f(\lambda) \right| = \varphi(\lambda)$, is that the integral

$$\int_{-\infty}^{+\infty} \frac{\left| \log \varphi(\lambda) \right|}{1+\lambda^2} \, d\lambda$$

be finite.

Setting $\varphi(\lambda) = F'(\lambda)$, we obtain the necessary and sufficient condition

$$\int_{-\infty}^{+\infty} \frac{\left| \log \sqrt{F'(\lambda)} \right|}{1+\lambda^2} \, d\lambda = \frac{1}{2} \int_{-\infty}^{+\infty} \frac{\left| \log F'(\lambda) \right|}{1+\lambda^2} \, d\lambda < \infty \tag{6.57}$$

in order that $g(a)$ in (6.55) may be chosen such that it vanishes for $a < 0$. Thus we have

Theorem 16. In order that the continuous stationary random function x(t) may be represented in the form

$$x(t) = \int_0^\infty g(a)d_a \, \zeta_1(t-a) \tag{6.58}$$

where $\zeta_1(t)$ satisfies conditions (5.34), it is necessary and sufficient that the spectrum of x(t) be absolutely continuous and that the integral

$$\int_{-\infty}^{+\infty} \frac{|\log F'(\lambda)|}{1+\lambda^2} \, d\lambda$$

converge.

When the conditions of the theorem are satisfied, the spectrum of x(t) is also complete.

43. Examples. -(1) Set

$$F'(\lambda) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\lambda^2}{2}} . \tag{6.59}$$

Then

$$r(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{i\lambda t} e^{-\frac{\lambda^2}{2}} \, d\lambda = e^{-\frac{t^2}{2}} . \tag{6.60}$$

Setting

$$f(\lambda) = \sqrt{F'(\lambda)} = (2\pi)^{-\frac{1}{4}} e^{-\frac{\lambda^2}{4}}$$

we obtain

$$g(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{i\lambda t} f(\lambda) d\lambda = (2\pi)^{-\frac{3}{4}} \int_{-\infty}^{+\infty} e^{i\lambda t} e^{-\frac{\lambda^2}{4}} d\lambda = \sqrt{\frac{2}{\pi}} e^{-t^2}.$$

By theorem 15, a stationary random function with the correlation function (6.60) may be represented in the form

$$x(t) = \sqrt[4]{\frac{2}{\pi}} \int_{-\infty}^{+\infty} e^{-(a+t)^2} d\zeta(a). \tag{6.61}$$

Since $F'(\lambda) > 0$ everywhere, $L_2(x) = L_2(\zeta)$. The integral

$$\int_{-\infty}^{+\infty} \frac{|\log F'(\lambda)|}{1+\lambda^2} d\lambda = \frac{1}{2} \int_{-\infty}^{+\infty} \frac{\log 2\pi + \lambda^2}{1+\lambda^2} d\lambda$$

diverges, so that by Theorem 16 $x(t)$ may not be represented in the form (6.58).

(2) Let

$$F'(\lambda) = \frac{1}{\pi} \frac{1}{1+\lambda^2}. \tag{6.62}$$

Then

$$r(t) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{e^{i\lambda t}}{1+\lambda^2} d\lambda = e^{-|t|}. \tag{6.63}$$

Setting

$$f(\lambda) = \frac{1}{\sqrt{\pi}} \ \frac{1}{1+i\lambda}$$

obviously $f(\lambda)^2 = F'(\lambda)$. It is seen from an easy computation, that the Fourier transform of $f(\lambda)$ is

$$g(t) = \lim_{\lambda \to \infty} \ \frac{1}{\pi\sqrt{2}} \int_{-\lambda}^{+\lambda} \frac{e^{i\lambda t}}{1+i\lambda} \, d\lambda = \begin{cases} \sqrt{2}e^{-t}, & \text{if } t > 0 \\ \\ 0, & \text{if } t < 0. \end{cases}$$

A stationary random function with correlation function (6.63) can be represented in the form

$$x(t) = \sqrt{2} \int_{0}^{\infty} e^{-a} d_a \zeta_1(t-a) \, . \tag{6.64}$$

It follows from (6.62) that the conditions of Theorem 16 are fulfilled.

(3) Let

$$F'(\lambda) = \begin{cases} \frac{1}{2\pi}, & \text{if } |\lambda| \leqq \pi \\ \\ 0, & \text{if } |\lambda| > \pi \, . \end{cases} \tag{6.65}$$

Then

$$r(t) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} e^{i\lambda t} d\lambda = \frac{e^{\pi i t} - e^{-\pi i t}}{2\pi i t} = \frac{\sin \pi t}{\pi t} \, . \tag{6.66}$$

Setting $f(\lambda) = \sqrt{F'(\lambda)}$, we obtain

$$g(t) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} e^{i\lambda t} d\lambda = \frac{\sin \pi t}{\pi t} \; .$$

A stationary random function $x(t)$ with correlation function (6.66) can be expressed in the form

$$x(t) = \int_{-\infty}^{+\infty} \frac{\sin \pi (a+t)}{\pi (a+t)} \, d\zeta(a) \qquad (6.67)$$

By 6.65 the spectrum of $x(t)$ is not complete. By Theorem 15 $L_2(x)$ is a proper subset of $L_2(\zeta)$, so that no unique inversion of (6.67) exists. Hence by Theorem 16, $x(t)$ may not be represented in the form (6.58).

BIBLIOGRAPHY

Bochner, S., [1] Vorlesungen uber Fouriersche Integrale, Leipzig, 1932.

Bohr, H., [1] Fastperiodische Funktionen, Berlin, 1932.

Cramér, H. [1] "Random Variables and Probability Distributions," Cambridge Tracts in Mathematics, No. 36, Cambridge, 1937.

Cramér, H. [2] "On the Theory of Stationary Random Processes," Annals of Mathematics, Vol. 41, 1940.

Cramér, H. [3] "On Harmonic Analysis in Certain Functional Spaces," Arkiv for matematick, astronomi och fysik, 28 B, 1942.

Doob, J. L., [1] "Stochastic Processes Depending on a Continuous Parameter," Transactions of the American Mathematical Society, Vol. 42, 1937.

Doob, J. B., [2] "The Brownian Movement and Stochastic Equations," Annals of Mathematics, Vol. 43, 1942.

Doob, J. B., [3] "The Elementary Gaussian Processes," Annals of Mathematical Statistics, Vol. 15, 1944.

Doob, J. L., and Ambrose, W., [1] "On Two Formulations of the Theory of Stochastic Processes Depending Upon a Continuous Parameter," Annals of Mathematics, Vol. 41, 1940.

Fréchet, M., [1] Recherches théoriques modernes sur la théorie des probabilités, Premier livre, Paris, 1937.

Hopf, E., [1] Ergodentheorie, Berlin, 1937.

Jessen, B. [1] Abstrakt maal, og Integralteori, Kφbenhavn, 1947.

Karhunen, K., [1] "Zur Spektraltheorie Stochastischer Prozesse, Ann. Ac. Sci. Fennicae, A, I, p. 34, 1946.

Karhunen, K., [2] "Lineare Transformationen Stationärer Stochastischer Prozesse," X Skand. Mat. Kongr., Kφbenhavn, 1946.

Khinchin, A., [1] "Korrelationstheorie der stationären stochastischen Prozesse, Mathematische Annalen, Vol. 109, 1934.

Kolmogorov, A., [1] Grundbegriffe der Wahrscheinlichkeitsrechnung, Berlin 1933.

Kolmogorov, A., [2] "Stationary Sequences in Hilbert Space (Russian)," Bulletin de I'Université de Moscou, Vol. 2, 1941.

Lévy, P., [1] Théorie de l'addition des variables aléatoires, Paris, 1937.

Löwig, H., [1] "Komplexe euklidische Räume von beliebiger endlicher oder unendlicher Dimensionszahl, Acta Sci. Math. Szeged, Vol. 7, 1934.

Neumann, J. v., [1] Mathematische Grundlagen der Quantenmechanik, Berlin, 1932.

Paley, R. and Wiener, N., [1] Fourier Transforms in the Complex Domain, New York, 1934.

Riesz, F., [1] "Zur Theorie des Hilbertschen Raumes," Acta. Sci. Math. Szeged, Vol. 7, 1934.

Saks, S.,[1] Theory of the Integral, Warszawa - Lwow, 1937.

Slutsky, E.,[1] Qualche proposizioni relative alla teoria delle funzioni aleatorie, Giornale dell' Istituto Italiano degli Attuari, Anno VIII, No. 2, 1937.

Slutsky, E.,[2] Sur les fonctions aléatoires presque périodiques etc., Actualités scientifiques et industrielles, 738 (1938).

Stone, M. H., [1] Linear Transformations in Hilbert Space, New York, 1932.

SZ. Nagy, B. v., [1] Spektraldarstellung linearer Transformationen des Hilbertschen Raumes, Berlin, 1942.

Weiner, N.: cf. Paley, R., and Wiener, N.

Wold, H., [1] A Study in the Analysis of Stationary Time Series, Uppsala, 1938.

Several new papers, especially by Blanc-LaPierre, Brard, Fortet, Kampé de Fériet, Loève, and Ville were unfortunately unavailable to the author.