ON SOME CLASSES OF NONSTATIONARY STOCHASTIC PROCESSES

HARALD CRAMÉR
STOCKHOLM

1. Introduction

This paper will be concerned with stochastic processes with finite second-order moments. We start from a given probability space $(\Omega, \mathfrak{F}, \mathfrak{P})$, where Ω is a space of points ω, while $\mathfrak{F}$ is a Borel field of sets in Ω, and $\mathfrak{P}$ is a probability measure defined on sets of $\mathfrak{F}$.

Any $\mathfrak{F}$-measurable complex-valued function $X = x(\omega)$ defined for all $\omega \in \Omega$ will be denoted as a *random variable*. We shall always assume that

$$Ex = \int_\Omega x(\omega)\, d\mathfrak{P} = 0, \qquad E|x|^2 = \int_\Omega |x(\omega)|^2\, d\mathfrak{P} < \infty. \tag{1}$$

Two random variables which are equal except on a null set with respect to $\mathfrak{P}$ will be regarded as identical, and equations containing random variables are always to be understood in this sense.

A family of random variables $x(t) = x(t, \omega)$, defined for all t belonging to some given set T, will be called a *stochastic process* defined on T. With respect to T, we shall consider only two cases:

(i) T is the set of all integers $n = 0, \pm 1, \pm 2, \cdots$,

(ii) T is the set of all real numbers t.

With the usual terminology borrowed from the applications, we shall in these cases talk respectively of a stochastic process with *discrete time*, or with *continuous time*. In the first case, where we are concerned with a sequence of random variables, we shall usually write x_n in place of $x(n)$.

With due modifications, the majority of our considerations may be extended to cases where T is some other set of real numbers.

We shall also consider *finite-dimensional vector-valued stochastic processes*, writing

$$\mathbf{x}(t) = \{x^{(1)}(t), x^{(2)}(t), \cdots, x^{(q)}(t)\}, \tag{2}$$

where $\mathbf{x}(t)$ is a q-dimensional column vector, while the components $x^{(1)}(t), \cdots, x^{(q)}(t)$ are stochastic processes in the above sense.

The *covariance functions* of the $\mathbf{x}(t)$ process are

$$R_{jk}(t, u) = E\{x^{(j)}(t)\bar{x}^{(k)}(u)\} = \overline{R}_{kj}(u, t), \tag{3}$$

where $j, k = 1, 2, \cdots, q$. These are all finite, since we consider only variables with finite second-order moments. In the particular case $q = 1$, we are concerned with one single stochastic process $x(t)$, and there is only one covariance function

$$R(t, u) = E\{x(t)\bar{x}(u)\}. \tag{4}$$

2. Stationary stochastic processes

In the important particular case when all R_{jk} are functions of the difference $t - u$, so that we have

$$R_{jk}(t, u) = r_{jk}(t - u), \tag{5}$$

the $\mathbf{x}(t)$ process is known as a *stationary stochastic process*. We shall not here be concerned with the so-called *strictly stationary* processes, which satisfy more restrictive conditions.

The class of stationary processes possesses useful and interesting properties, which have been thoroughly studied. The present paper is the outcome of an attempt to generalize some of these properties to certain classes of nonstationary processes. In particular, various properties related to the problem of *linear least squares prediction* will be considered. For the sake of brevity, we shall in the sequel always use the word "prediction" in the sense of "linear least squares prediction."

In the first part of the paper, the general case of a vector-valued stochastic process with finite second-order moments will be studied. For such a process, there exists a uniquely defined decomposition into a *deterministic* and a *purely nondeterministic* component, which are mutually orthogonal.

In the case of a vector process with discrete time, the properties of this decomposition are a straightforward generalization of the well-known Wold decomposition [12] for stationary one-variable processes with discrete time. This case will be treated in detail in section 4 of the present paper.

For a process with continuous time, on the other hand, the properties of the nondeterministic component are somewhat more complicated than in the case of a stationary process. This case will be dealt with in section 5. We shall here give only the main lines of the argument, and state our main results. Complete proofs will be given in a forthcoming publication in the *Arkiv för Matematik*.

The decomposition into a deterministic and a purely nondeterministic component forms the basis of a *time-domain analysis* of a given stochastic process, generalizing the well-known properties of stationary processes.

For the class of stationary processes, "the chief advantage of turning from the time-domain analysis of stochastic processes to the *frequency-domain* or *spectral analysis* is the possibility of using the powerful methods of harmonic analysis" (Wiener and Masani [11], p. 140). This can be done for stationary

processes, since there exist for these processes *spectral representations* in the form of Fourier-Stieltjes integrals, both for the process variables themselves and for the associated covariance functions.

In the second part of the present paper, we shall consider certain classes of nonstationary processes admitting spectral representations of a similar kind. Our results in this part of the paper are of a much less definite character than those given in the first part. In fact, only some highly preliminary results concerned with the spectral analysis of processes will be given here. Also we shall here consider only one-variable processes with discrete time, although most of our results may be generalized to the vector case, and also to processes with continuous time.

Two classes of stochastic processes admitting spectral representations will be considered, each including the stationary processes as a particular case.

For a stationary process with discrete time, it is well known that there exists a representation in the form of a stochastic Fourier-Stieltjes integral

$$x_n = \int_0^{2\pi} e^{inu}\, dz(u), \tag{6}$$

where $z(u)$ is a stochastic process with *orthogonal increments*. If we consider a process x_n representable in the same form, but without requiring that $z(u)$ should necessarily have orthogonal increments, we shall be led to a class of stochastic processes first introduced by Loève [7], [8], and called by him *harmonizable processes*. Obviously this class contains the class of stationary processes. The harmonizable processes will be considered in section 6 of the present paper.

Finally, in section 7 we shall consider a different kind of generalization of the concept of a stationary process. With respect to a stationary process with discrete time, it is well known that there exists a *unitary shift operator* U, which takes every x_n into the immediately following variable x_{n+1}. The properties of this operator are intimately connected with the properties of the stationary process. The more general class of processes obtained when it is only assumed that the shift operator is *normal* has been studied by Getoor [5]. In section 7 we shall give some very preliminary results concerning the spectral analysis of this class of processes, restricting ourselves to the case of processes with discrete time.

PART I. THE GENERAL VECTOR-VALUED PROCESS WITH FINITE SECOND-ORDER MOMENTS

3. Notation, deterministic and nondeterministic processes

All random variables $x, y, \cdots$ defined on the given probability space $(\Omega, \mathfrak{F}, \mathfrak{P})$, and satisfying (1), form a Hilbert space $\mathfrak{H}$, if the inner product and the norm are defined by the usual expressions

$$(x, y) = E(x\bar{y}), \qquad ||x||^2 = E|x|^2. \tag{7}$$

Whenever we use the term *convergence* with respect to a sequence of random variables, it will be understood that we refer to convergence in the topology induced by this norm, that is, convergence in quadratic mean.

We now consider a vector-valued stochastic process $\mathbf{x}(t) = \{x^{(1)}(t), \cdots, x^{(q)}(t)\}$, where the $x^{(j)}(t)$ are complex-valued stochastic processes, defined for all $t \in T$. Every random variable $x^{(j)}(t)$ is assumed to satisfy (1), and is thus an element of $\mathfrak{H}$. In order to avoid trivial difficulties we shall always suppose that, for every $j = 1, \cdots, q$,

$$E|x^{(j)}(t)|^2 > 0 \tag{8}$$

for at least one $t \in T$. We shall say that two processes $\mathbf{x}(t)$ and $\mathbf{y}(t)$ of this type are *orthogonal*, in symbols $\mathbf{x}(t) \perp \mathbf{y}(t)$, if

$$E\{x^{(j)}(t)\bar{y}^{(k)}(u)\} = 0, \tag{9}$$

for $j, k = 1, \cdots, q$ and all $t, u \in T$.

Let $\mathfrak{H}(\mathbf{x}, t)$ denote the subspace of $\mathfrak{H}$ spanned by the random variables $x^{(j)}(u)$ for $j = 1, \cdots, q$ and all u such that $u \in T$ and $u \leqq t$. We shall write this

$$\mathfrak{H}(\mathbf{x}, t) = \mathfrak{S}\{x^{(j)}(u), j = 1, \cdots, q, u \in T, u \leqq t\}, \tag{10}$$

where $\mathfrak{S}$ stands for "span." Instead of $\mathfrak{H}(\mathbf{x}, +\infty)$, we shall write simply $\mathfrak{H}(\mathbf{x})$.

As t decreases, the set $\mathfrak{H}(\mathbf{x}, t)$ can never increase. It follows that, when $t \to -\infty$, the set $\mathfrak{H}(\mathbf{x}, t)$ must tend to a limiting set, which we denote by $\mathfrak{H}(\mathbf{x}, -\infty)$. We thus have for any $t_1 < t_2$

$$\mathfrak{H}(\mathbf{x}, -\infty) \subset \mathfrak{H}(\mathbf{x}, t_1) \subset \mathfrak{H}(\mathbf{x}, t_2) \subset \mathfrak{H}(\mathbf{x}) \subset \mathfrak{H}. \tag{11}$$

It can be said that $\mathfrak{H}(\mathbf{x}, t)$ contains all the information available when we know the development of all the component processes $x^{(j)}(u)$ up to and including the point t. In the terminology used by Wiener and Masani ([11], p. 135), we can say that $\mathfrak{H}(\mathbf{x}, t)$ represents the *past and present* of $\mathbf{x}(t)$, while $\mathfrak{H}(\mathbf{x}, -\infty)$ corresponds to the *remote past* of the process.

If $\mathfrak{M}$ is any subspace of $\mathfrak{H}$, that is, any closed linear manifold in $\mathfrak{H}$, we denote by $P_{\mathfrak{M}}y$ the projection on $\mathfrak{M}$ of an arbitrary element y of $\mathfrak{H}$. When $\mathfrak{M} = \mathfrak{H}(\mathbf{x}, t)$, we write simply P_t instead of $P_{\mathfrak{M}}$.

Similarly, if $y^{(1)}, \cdots, y^{(q)}$ are any elements of $\mathfrak{H}$, and $\mathbf{y}$ denotes the column vector

$$\mathbf{y} = (y^{(1)}, \cdots, y^{(q)}), \tag{12}$$

we shall write

$$P_{\mathfrak{M}}\mathbf{y} = (P_{\mathfrak{M}}y^{(1)}, \cdots, P_{\mathfrak{M}}y^{(q)}), \tag{13}$$

replacing $P_{\mathfrak{M}}$ by P_t in the particular case when $\mathfrak{M} = \mathfrak{H}(\mathbf{x}, t)$.

The projection $P_{t-h}x^{(j)}(t)$ is, among all elements y of the subspace $\mathfrak{H}(\mathbf{x}, t-h)$, that which minimizes the norm $||x^{(j)}(t) - y||$. Accordingly $P_{t-h}x^{(j)}(t)$ is, from the point of view of linear least squares prediction, the best possible prediction of $x^{(j)}(t)$ in terms of all variables $x^{(1)}(u), \cdots, x^{(q)}(u)$ with $u \leqq t - h$. The norm

$$\sigma_{ik}^{(j)} = ||x^{(j)}(t) - P_{t-h}x^{(j)}(t)|| \tag{14}$$

is the corresponding *error of prediction*. For $0 < h < k$, we obviously have

$$0 \leqq \sigma_{ik}^{(j)} \leqq \sigma_{ik}^{(j)}. \tag{15}$$

Going back to relation (11), we shall now consider two extreme possibilities with respect to the subspace $\mathfrak{H}(\mathbf{x}, -\infty)$, namely

$$\begin{aligned} &\text{(A)} && \mathfrak{H}(\mathbf{x}, -\infty) = \mathfrak{H}(\mathbf{x}), \\ &\text{(B)} && \mathfrak{H}(\mathbf{x}, -\infty) = 0. \end{aligned} \tag{16}$$

In case (A), it follows from (11) that $\mathfrak{H}(\mathbf{x}, t) = \mathfrak{H}(\mathbf{x}, -\infty)$ for all t. Thus, in particular, $x^{(j)}(t) \in \mathfrak{H}(\mathbf{x}, -\infty)$ for all j and t, and it follows that the prediction error $\sigma_{ik}^{(j)}$ reduces to zero for $j = 1, \cdots, q$, for all $t \in T$ and all $h > 0$. Hence for every $t \in T$ the components $x^{(j)}(t)$ of $\mathbf{x}(t)$ all can be exactly predicted by means of the information provided by the arbitrarily remote past of the process. In this case we shall say that the $\mathbf{x}(t)$ process is *deterministic*. Every $\mathbf{x}(t)$ process not satisfying condition (A) will be called *nondeterministic*.

On the other hand, in case (B) we can say that the information provided by the remote past of the $\mathbf{x}(t)$ process is, in the limit, of no value for the prediction of the component variables $x^{(j)}(t)$ at any given time t. Thus every piece of information contained in the process at the instant t must have entered the process as an *innovation* at some definite instant $u \leqq t$ in the past or present. Accordingly, a process satisfying condition (B) will be called a *purely nondeterministic* process. (Wiener and Masani [11] use the term *regular* process.) For such a process, the irrelevance of the remote past for prediction purposes may be expressed by the relation

$$\lim_{h \to \infty} \sigma_{ik}^{(j)} = ||x^{(j)}(t)||, \tag{17}$$

which holds for $j = 1, \cdots, q$ and every $t \in T$.

4. The discrete case

When T is the set of all integers $n = 0, \pm 1, \cdots$, we are concerned with a discrete vector process

$$\mathbf{x}_n = (x_n^{(1)}, \cdots, x_n^{(q)}), \tag{18}$$

where it is assumed only that each component $x_n^{(j)}$ is a complex-valued stochastic process with discrete time parameter n, zero mean-values, and finite second-order moments. Using the notations introduced above, we have

$$P_{n-1}\mathbf{x}_n = (P_{n-1}x_n^{(1)}, \cdots, P_{n-1}x_n^{(q)}). \tag{19}$$

Writing

$$\xi_n = \mathbf{x}_n - P_{n-1}\mathbf{x}_n, \qquad \xi_n^{(j)} = x_n^{(j)} - P_{n-1}x_n^{(j)} \tag{20}$$

it follows that

$$\xi_n = (\xi_n^{(1)}, \cdots, \xi_n^{(q)}). \tag{21}$$

The sequence of vector-valued random variables $\boldsymbol{\xi}_n$ defines a vector-valued stochastic process with discrete time parameter n. The $\boldsymbol{\xi}_n$ process will be called the *innovation process* corresponding to the given $\mathbf{x}_n$ process. This name may be justified by the following remarks.

If, for a certain value of n, we have

$$\boldsymbol{\xi}_n = (0, \cdots, 0), \tag{22}$$

this signifies that every component $\xi_n^{(j)}$ is zero, that is, every $x_n^{(j)}$ is contained in the subspace $\mathfrak{H}(\mathbf{x}, n-1)$. From the prediction point of view this means that every component $x_n^{(j)}$ of $\mathbf{x}_n$ can be predicted exactly by means of the information available when we know the development of the $\mathbf{x}$ process up to and including the instant $n-1$. Obviously this can be expressed by saying that no innovation enters the $\mathbf{x}$ process at the instant n or, equivalently, that the innovation received by the process at this instant reduces to zero.

Suppose, on the other hand, that for a certain n, the vector variable $\boldsymbol{\xi}_n$ has at least one component $\xi_n^{(j)}$ such that $E|\xi_n^{(j)}|^2 > 0$. Then $\xi_n^{(j)} = x_n^{(j)} - P_{n-1}x_n^{(j)}$ does not reduce to zero, so that $x_n^{(j)}$ cannot be predicted exactly in terms of the variables $x_m^{(1)}, \cdots, x_m^{(q)}$ with $m \leqq n-1$. The variable $\xi_n^{(j)}$ then represents the innovation received by the component $x_n^{(j)}$ at the instant n, and consequently $\boldsymbol{\xi}_n = (\xi_n^{(1)}, \cdots, \xi_n^{(q)})$, which is not identically zero, is the *innovation* entering into the vector process $\mathbf{x}$ at the instant n.

The set of all those values of n, for which the innovation $\boldsymbol{\xi}_n$ does not reduce to zero, may be said to form the *innovation spectrum* of the $\mathbf{x}_n$ process. This set contains precisely all those time points where a new impulse, or an innovation, enters into the process.

The innovation spectrum may be empty, finite, or infinite; and it will be readily seen that, to any given set of integers n, we can construct an $\mathbf{x}_n$ process having this set for its innovation spectrum. For a deterministic process, the innovation spectrum is evidently empty, while for any nondeterministic process it must contain at least one value of n. For a nondeterministic stationary process, the innovation spectrum includes all integers n.

Let now n be a given integer, and consider the set of q random variables $\xi_n^{(1)}, \cdots, \xi_n^{(q)}$, with covariance matrix

$$\mathbf{R}_n = \{E(\xi_n^{(j)}\bar{\xi}_n^{(k)})\}, \qquad j, k = 1, \cdots, q. \tag{23}$$

The rank r_n of $\mathbf{R}_n$ is equal to the maximum number of linearly independent variables among $\xi_n^{(1)}, \cdots, \xi_n^{(q)}$. Thus $0 \leqq r_n \leqq q$, and $r_n > 0$ if, and only if, n belongs to the innovation spectrum of the $\mathbf{x}$ process.

Let $\mathfrak{f}(\mathbf{x}, n)$ denote the r_n-dimensional space spanned by the variables $\xi_n^{(1)}, \cdots, \xi_n^{(q)}$. Since $\xi_n^{(j)} \in \mathfrak{H}(\mathbf{x}, n)$, it is seen that $\mathfrak{f}(\mathbf{x}, n)$ is a subspace of $\mathfrak{H}(\mathbf{x}, n)$.

It follows from (20) that $\xi_n^{(j)}$ is always orthogonal to $\mathfrak{H}(\mathbf{x}, n-1)$, and consequently $\mathfrak{f}(\mathbf{x}, n) \perp \mathfrak{H}(\mathbf{x}, n-1)$. It also follows from (20) that any two variables $\xi_m^{(j)}$ and $\xi_n^{(k)}$ with $m \neq n$ are orthogonal, so that $\mathfrak{f}(\mathbf{x}, m) \perp \mathfrak{f}(\mathbf{x}, n)$ when $m \neq n$.

If we orthogonalize the set of variables $\xi_n^{(1)}, \cdots, \xi_n^{(q)}$, we shall obtain a set of

r_n variables, say $\eta_n^{(1)}, \cdots, \eta_n^{(r_n)}$, forming a complete orthonormal system in $\mathfrak{f}(\mathbf{x}, n)$, and in addition $q - r_n$ zero variables. If n does not belong to the innovation spectrum, $r_n = 0$, and the space $\mathfrak{f}(\mathbf{x}, n)$ reduces to zero.

The vector sum of the orthogonal family of subspaces $\mathfrak{f}(\mathbf{x}, m)$ with $m \leqq n$,

$$\mathfrak{R}(\mathbf{x}, n) = \mathfrak{f}(\mathbf{x}, n) + \mathfrak{f}(\mathbf{x}, n-1) + \cdots \tag{24}$$

is the space spanned by all $\xi_m^{(j)}$ with $j = 1, \cdots, q$ and $m \leqq n$. Obviously this is a subspace of $\mathfrak{H}(\mathbf{x}, n)$. It will be seen that the set of all variables $\eta_m^{(j)}$, where $j = 1, \cdots, r_m$ and $m \leqq n$, forms a complete orthonormal system in $\mathfrak{R}(\mathbf{x}, n)$. This remark will be used later.

We now proceed to the proof of the following lemma which, in the stationary case, corresponds to part (b) of lemma 6.10 of Wiener and Masani [11].

LEMMA 1. *The space* $\mathfrak{R}(\mathbf{x}, n)$ *is, within* $\mathfrak{H}(\mathbf{x}, n)$, *the orthogonal complement of* $\mathfrak{H}(\mathbf{x}, -\infty)$. *In symbols*

$$\begin{aligned} &\mathfrak{R}(\mathbf{x}, n) \perp \mathfrak{H}(\mathbf{x}, -\infty), \\ &\mathfrak{H}(\mathbf{x}, n) = \mathfrak{R}(\mathbf{x}, n) + \mathfrak{H}(\mathbf{x}, -\infty). \end{aligned} \tag{25}$$

We have already seen that $\mathfrak{f}(\mathbf{x}, m)$ is orthogonal to $\mathfrak{H}(\mathbf{x}, m-1)$, and a fortiori orthogonal to $\mathfrak{H}(\mathbf{x}, -\infty)$. Consequently the vector sum $\mathfrak{R}(\mathbf{x}, n)$ is orthogonal to $\mathfrak{H}(\mathbf{x}, -\infty)$.

Further, since $\mathfrak{R}(\mathbf{x}, n)$ and $\mathfrak{H}(\mathbf{x}, -\infty)$ are both subspaces of $\mathfrak{H}(\mathbf{x}, n)$, we have

$$\mathfrak{R}(\mathbf{x}, n) + \mathfrak{H}(\mathbf{x}, -\infty) \subset \mathfrak{H}(\mathbf{x}, n). \tag{26}$$

On the other hand we have $x_n^{(j)} = \xi_n^{(j)} + P_{n-1}x_n^{(j)}$, so that every element of $\mathfrak{H}(\mathbf{x}, n)$ is the limit of a convergent sequence of variables, each of which is the sum of a linear combination of $\xi_n^{(1)}, \cdots, \xi_n^{(q)}$ and an element of $\mathfrak{H}(\mathbf{x}, n-1)$. Since $\mathfrak{f}(\mathbf{x}, n)$ and $\mathfrak{H}(\mathbf{x}, n-1)$ are orthogonal, it follows that every element of $\mathfrak{H}(\mathbf{x}, n)$ is the sum of one element of $\mathfrak{f}(\mathbf{x}, n)$ and one of $\mathfrak{H}(\mathbf{x}, n-1)$, so that

$$\mathfrak{H}(\mathbf{x}, n) \subset \mathfrak{f}(\mathbf{x}, n) + \mathfrak{H}(\mathbf{x}, n-1) \subset \mathfrak{R}(\mathbf{x}, n) + \mathfrak{H}(\mathbf{x}, n-1). \tag{27}$$

By repeated application of this relation we obtain

$$\mathfrak{H}(\mathbf{x}, n) \subset \mathfrak{R}(\mathbf{x}, n) + \mathfrak{H}(\mathbf{x}, n-p) \tag{28}$$

for every $p > 0$, and finally, as $p \to \infty$,

$$\mathfrak{H}(\mathbf{x}, n) \subset \mathfrak{R}(\mathbf{x}, n) + \mathfrak{H}(\mathbf{x}, -\infty). \tag{29}$$

This, together with (26), completes the proof of the lemma.

We can now prove the analogue of the Wold decomposition for the $\mathbf{x}_n$ process, thus generalizing theorem 6.11 of Wiener and Masani [11].

THEOREM 1. *For any given* $\mathbf{x}_n$ *process, there is a uniquely determined decomposition*

$$\mathbf{x}_n = \mathbf{u}_n + \mathbf{v}_n \tag{30}$$

having properties (a) *and* (b).

(a) $\mathbf{u}_n = (u_n^{(1)}, \cdots, u_n^{(q)})$ and $\mathbf{v}_n = (v_n^{(1)}, \cdots, v_n^{(q)})$, *where all* $u_n^{(j)}$ *and* $v_n^{(j)}$ *belong to* $\mathfrak{H}(\mathbf{x}, n)$.

(b) *The* $\mathbf{u}_n$ *and* $\mathbf{v}_n$ *processes are orthogonal, and* $\mathbf{u}_n$ *is purely nondeterministic, while* $\mathbf{v}_n$ *is deterministic. The nondeterministic component* $\mathbf{u}_n$ *has, in addition, property* (c).

(c) $\mathbf{u}_n$ *can be expressed as a linear combination of those innovations* ξ_p *of the* $\mathbf{x}_n$ *process that have entered into the process before or at the instant* n,

$$\mathbf{u}_n = \sum_{p=-\infty}^{n} \mathbf{A}_{np}\xi_p, \tag{31}$$

where the $\mathbf{A}_{np} = \{a_{np}^{(jk)}\}$ *are* $q \times q$ *matrices, such that the development formally obtained for any component* $u_n^{(j)}$,

$$u_n^{(j)} = \sum_{p=-\infty}^{n} \sum_{k=1}^{q} a_{np}^{(jk)}\xi_p^{(k)}, \tag{32}$$

is convergent in the topology of $\mathfrak{H}$. *Thus, writing*

$$c_{np}^{(j)} = \left\| \sum_{k=1}^{q} a_{np}^{(jk)}\xi_p^{(k)} \right\| = \left\{ E\left| \sum_{k=1}^{q} a_{np}^{(jk)}\xi_p^{(k)} \right|^2 \right\}^{1/2}, \tag{33}$$

we have

$$\sum_{p=-\infty}^{n} (c_{np}^{(j)})^2 < \infty \tag{34}$$

for all n *and for* $j = 1, \cdots, q$. *The coefficients* $a_{np}^{(jk)}$ *are uniquely determined if, and only if, the rank* r_p *has the maximum value* q, *while the* $c_{np}^{(j)}$ *are uniquely determined for all* n, p, *and* j.

PROOF. For all n and for $j = 1, \cdots, q$ we take for $u_n^{(j)}$ and $v_n^{(j)}$ the projections of $x_n^{(j)}$ on the subspaces $\mathfrak{N}(\mathbf{x}, n)$ and $\mathfrak{H}(\mathbf{x}, -\infty)$ respectively. It then follows from lemma 1 that $u_n^{(j)}$ and $v_n^{(j)}$ belong to $\mathfrak{H}(\mathbf{x}, n)$, and that we have

$$\begin{aligned} x_n^{(j)} &= u_n^{(j)} + v_n^{(j)}, \\ \mathbf{x}_n &= \mathbf{u}_n + \mathbf{v}_n, \end{aligned} \tag{35}$$

where

$$\mathbf{u}_n = (u_n^{(1)}, \cdots, u_n^{(q)}), \qquad \mathbf{v}_n = (v_n^{(1)}, \cdots, v_n^{(q)}). \tag{36}$$

Since $u_m^{(j)}$ belongs to $\mathfrak{N}(\mathbf{x}, m)$, while $v_n^{(k)}$ belongs to $\mathfrak{H}(\mathbf{x}, -\infty)$, it further follows from lemma 1 that $u_m^{(j)}$ and $v_n^{(k)}$ are always orthogonal. Thus the $\mathbf{u}_n$ and $\mathbf{v}_n$ processes are orthogonal, according to the definition given in section 3. If, in accordance with section 3, we define

$$\begin{aligned} \mathfrak{H}(\mathbf{u}, n) &= \mathfrak{S}(u_m^{(j)}, j = 1, \cdots, q, m \leqq n), \\ \mathfrak{H}(\mathbf{v}, n) &= \mathfrak{S}(v_m^{(j)}, j = 1, \cdots, q, m \leqq n), \end{aligned} \tag{37}$$

we thus find that $\mathfrak{H}(\mathbf{u}, m)$ and $\mathfrak{H}(\mathbf{v}, n)$ are orthogonal for all m and n.

Since all $u_m^{(j)}$ and $v_m^{(j)}$ with $m \leqq n$ belong to $\mathfrak{H}(\mathbf{x}, n)$, we have

$$\mathfrak{H}(\mathbf{u}, n) + \mathfrak{H}(\mathbf{v}, n) \subset \mathfrak{H}(\mathbf{x}, n). \tag{38}$$

On the other hand, it follows from (35) and from the orthogonality of the $\mathbf{u}_n$ and $\mathbf{v}_n$ processes that

$$\mathfrak{H}(\mathbf{x}, n) \subset \mathfrak{H}(\mathbf{u}, n) + \mathfrak{H}(\mathbf{v}, n). \tag{39}$$

Hence by lemma 1, $\mathfrak{H}(\mathbf{u}, n) + \mathfrak{H}(\mathbf{v}, n) = \mathfrak{H}(\mathbf{x}, n) = \mathfrak{N}(\mathbf{x}, n) + \mathfrak{H}(\mathbf{x}, -\infty)$. By the definition of the $u_n^{(j)}$ and $v_n^{(j)}$ we have, however, $\mathfrak{H}(\mathbf{u}, n) \subset \mathfrak{N}(\mathbf{x}, n)$ and $\mathfrak{H}(\mathbf{v}, n) \subset \mathfrak{H}(\mathbf{x}, -\infty)$, and thus obtain

$$\begin{aligned} \mathfrak{H}(\mathbf{u}, n) &= \mathfrak{N}(\mathbf{x}, n), \\ \mathfrak{H}(\mathbf{v}, n) &= \mathfrak{H}(\mathbf{x}, -\infty). \end{aligned} \tag{40}$$

From lemma 1 we then obtain $\mathfrak{H}(\mathbf{u}, -\infty) = \mathfrak{N}(\mathbf{x}, -\infty) = 0$, so that the $\mathbf{u}_n$ process is purely nondeterministic. On the other hand $\mathfrak{H}(\mathbf{v}, -\infty) = \mathfrak{H}(\mathbf{x}, -\infty) = \mathfrak{H}(\mathbf{v}, n)$ for every n, and so the $\mathbf{v}_n$ process is deterministic.

The properties (a) and (b) of the decomposition considered here are thus established, and we shall now prove that this is the only decomposition of the given $\mathbf{x}_n$ process that has these properties. Suppose, in fact, that $\mathbf{u}_n$ and $\mathbf{v}_n$ are any processes satisfying (35), and having the properties (a) and (b) stated in the theorem. Then (38) and (39) will still hold, so that we obtain as before

$$\mathfrak{H}(\mathbf{x}, n) = \mathfrak{H}(\mathbf{u}, n) + \mathfrak{H}(\mathbf{v}, n) \tag{41}$$

for all n. It is readily seen that, owing to the orthogonality of $\mathfrak{H}(\mathbf{u}, n)$ and $\mathfrak{H}(\mathbf{v}, n)$, this holds even for $n = -\infty$, and we obtain

$$\mathfrak{H}(\mathbf{x}, -\infty) = \mathfrak{H}(\mathbf{u}, -\infty) + \mathfrak{H}(\mathbf{v}, -\infty). \tag{42}$$

However, on account of property (b), we have $\mathfrak{H}(\mathbf{u}, -\infty) = 0$, and thus

$$\mathfrak{H}(\mathbf{x}, -\infty) = \mathfrak{H}(\mathbf{v}, -\infty) = \mathfrak{H}(\mathbf{v}, n) \tag{43}$$

for all n. Hence by (41) and lemma 1 we find that relations (40) will still hold. In the decomposition $x_n^{(j)} = u_n^{(j)} + v_n^{(j)}$, the first component must then be the projection of $x_n^{(j)}$ on $\mathfrak{N}(\mathbf{x}, n)$, and the second the projection on $\mathfrak{H}(\mathbf{x}, -\infty)$, so that the decomposition is unique.

It finally remains to prove property (c). In order to do this, we use the remark made above about the completeness of the orthonormal system $\eta_m^{(j)}$ in the space $\mathfrak{N}(\mathbf{x}, n)$. The corresponding Fourier development of the element $u_n^{(j)} \in \mathfrak{N}(\mathbf{x}, n)$ will have the form

$$u_n^{(j)} = \sum_{p=-\infty}^{n} \sum_{k=1}^{r_p} b_{np}^{(jk)} \eta_p^{(k)} \tag{44}$$

with

$$\sum_{p=-\infty}^{n} \sum_{k=1}^{r_p} |b_{np}^{(jk)}|^2 < \infty. \tag{45}$$

For any fixed p, the orthogonal variables $\eta_p^{(1)}, \cdots, \eta_p^{(r_p)}$ are certain linear combinations of the innovation components $\xi_p^{(1)}, \cdots, \xi_p^{(q)}$. The coefficients appearing in these linear combinations will be uniquely determined if, and only if, the rank r_p has its maximum value q. Replacing the $\eta_p^{(k)}$ in the above development of $u_n^{(j)}$ by their expressions in terms of the $\xi_p^{(k)}$, we obtain the development given under (c), and we find that

$$(c_{np}^{(j)})^2 = \sum_{k=1}^{r_p} |b_{np}^{(jk)}|^2. \tag{46}$$

The proof of theorem 1 is thus completed.

We can now immediately state the following result, which gives the application of theorem 1 to the prediction problem for the $\mathbf{x}_n$ process.

THEOREM 2. *Let h be any positive integer. With the notation of theorem* 1, *the best prediction of the component $x_n^{(j)}$ in terms of all variables $x_p^{(1)}, \cdots, x_p^{(q)}$ with $p \leqq n - h$ will be*

$$P_{n-h}x_n^{(j)} = \sum_{p=-\infty}^{n-h} \sum_{k=1}^{q} a_{np}^{(jk)} \xi_p^{(k)} + v_n^{(j)}, \tag{47}$$

with the corresponding error of prediction

$$\begin{aligned} \sigma_{nh}^{(j)} &= ||x_n^{(j)} - P_{n-h}x_n^{(j)}|| \\ &= \left\{ \sum_{p=n-h+1}^{n} (c_{np}^{(j)})^2 \right\}^{1/2}. \end{aligned} \tag{48}$$

This follows directly from theorem 1 and from the definition (14) of the error of prediction, if we observe that $v_n^{(j)}$, as well as all $\xi_p^{(k)}$ with $p \leqq n - h$, belong to $\mathfrak{H}(\mathbf{x}, n - h)$, while all $\xi_p^{(k)}$ with $p > n - h$ are orthogonal to this space.

It should be noted that the coefficients $a_{np}^{(jk)}$ in the expression for the best prediction of $\mathrm{x}_n^{(j)}$ given in theorem 2 depend on certain covariances of the x process up to the time n. Accordingly, any statistical estimation of this prediction by means of theorem 2 must be based on information concerning the covariance structure of the process up to the time n, either from a priori knowledge (as in the case when the process is assumed to be stationary), or from previous statistical experience.

5. The continuous case

REMARK. I am indebted to Professor K. Itô for the observation that there are interesting points of contact of this section and a work by T. Hida on "Canonical representations of Gaussian processes," which will shortly appear in the *Memoirs of the College of Science, University of Kyoto.*

We now consider a q-dimensional stochastic vector process

$$\mathbf{x}(t) = \{x^{(1)}(t), \cdots, x^{(q)}(t)\}, \tag{49}$$

the parameter set T being the set of all real numbers t. Each component $x^{(j)}(t)$ is a complex-valued stochastic process with continuous time t, and the covariance functions $R_{jk}(t, u)$ defined by (3) are all assumed to be finite.

When t increases from $-\infty$ to $+\infty$, the point $x^{(j)}(t)$ describes a curve in the Hilbert space $\mathfrak{H}(\mathbf{x})$, and the $\mathbf{x}(t)$ process is made up by the set of q curves corresponding to the components $x^{(1)}(t), \cdots, x^{(q)}(t)$. The subspace $\mathfrak{H}(\mathbf{x}, t_0)$ is spanned by the arcs of these curves that belong to the domain $t \leqq t_0$. The properties of the

family of all subspaces $\mathfrak{H}(\mathbf{x}, t)$, where t ranges from $-\infty$ to $+\infty$, will play an important part in the sequel.

The following theorem is directly analogous to the first part of theorem 1, and can be proved along similar lines, so that we may content ourselves here with stating the theorem.

THEOREM 3. *There is a unique decomposition of the* $\mathbf{x}(t)$ *process,*

$$\mathbf{x}(t) = \mathbf{u}(t) + \mathbf{v}(t), \tag{50}$$

having properties (a) *and* (b).

(a) $\mathbf{u}(t) = \{u^{(1)}(t), \cdots, u^{(q)}(t)\}$ *and* $\mathbf{v}(t) = \{v^{(1)}(t), \cdots, v^{(q)}(t)\}$, *where all* $u^{(j)}(t)$ *and* $v^{(j)}(t)$ *belong to* $\mathfrak{H}(\mathbf{x}, t)$.

(b) *The* $\mathbf{u}(t)$ *and* $\mathbf{v}(t)$ *processes are orthogonal, and* $\mathbf{u}(t)$ *is purely nondeterministic, while* $\mathbf{v}(t)$ *is deterministic.*

The second part of theorem 1 is concerned with the representation of the nondeterministic component of a given process with *discrete* time as a linear function of the innovations associated with the past and present of the process. For the nondeterministic component of a process with *continuous* time there exists, in fact, an analogous representation. However, the circumstances are somewhat more complicated than in the discrete case, and we shall here only give some preliminary discussion and state our main results, reserving complete proofs for a forthcoming publication.

For our present purpose, it will be sufficient to deal with the purely nondeterministic component $\mathbf{u}(t)$ of the given $\mathbf{x}(t)$ process, and we may then as well assume that $\mathbf{x}(t)$ itself is purely nondeterministic, that is, the deterministic component $\mathbf{v}(t)$ is identically zero. Further, we shall find it convenient to introduce a certain regularity condition relating to the behavior of the $\mathbf{x}(t)$ process in points of discontinuity. Thus it will be assumed throughout the rest of the present section that we are dealing with a vector process $\mathbf{x}(t)$ satisfying the following two conditions.

(C_1) $\mathbf{x}(t)$ is purely nondeterministic, that is, $\mathfrak{H}(\mathbf{x}, -\infty) = 0$.

(C_2) The limits $x^{(j)}(t - 0)$ and $x^{(j)}(t + 0)$ exist (as always in the $\mathfrak{H}$ topology) for $j = 1, \cdots, q$ and for every real t.

We shall then write

$$\mathbf{x}(t - 0) = \{x^{(1)}(t - 0), \cdots, x^{(q)}(t - 0)\}, \tag{51}$$

and similarly for $\mathbf{x}(t + 0)$.

It follows without difficulty from condition (C_2) that *the space* $\mathfrak{H}(\mathbf{x})$ *is separable,* and that the set of *points of discontinuity of* $\mathbf{x}(t)$, that is, the set of all t such that at least one of the relations

$$\mathbf{x}(t - 0) = \mathbf{x}(t) = \mathbf{x}(t + 0) \tag{52}$$

is not satisfied, is at most enumerable.

Let us now consider the family of subspaces $\mathfrak{H}(\mathbf{x}, t)$ of the space $\mathfrak{H}(\mathbf{x})$. As t increases from $-\infty$ to $+\infty$, the $\mathfrak{H}(\mathbf{x}, t)$ form a never decreasing set of subspaces, with $\mathfrak{H}(\mathbf{x}, -\infty) = 0$ and $\mathfrak{H}(\mathbf{x}, +\infty) = \mathfrak{H}(\mathbf{x})$. The limits $\mathfrak{H}(\mathbf{x}, t \pm 0)$ will exist

for all t. If $(t, t + h)$ and $(u, u + k)$ are disjoint intervals, the orthogonal complements

$$\mathfrak{H}(\mathbf{x}, t + h) - \mathfrak{H}(\mathbf{x}, t) \quad \text{and} \quad \mathfrak{H}(\mathbf{x}, u + k) - \mathfrak{H}(\mathbf{x}, u) \tag{53}$$

are mutually orthogonal.

The set of all t such that for any $h > 0$ we have

$$\mathfrak{H}(\mathbf{x}, t + h) - \mathfrak{H}(\mathbf{x}, t - h) \neq 0 \tag{54}$$

will be called the *innovation spectrum* of the $\mathbf{x}(t)$ process. A point t such that at least one of the relations

$$\mathfrak{H}(\mathbf{x}, t - 0) = \mathfrak{H}(\mathbf{x}, t) = \mathfrak{H}(\mathbf{x}, t + 0) \tag{55}$$

is not satisfied, is a *point of discontinuity* of the innovation spectrum. The space $\mathfrak{H}(\mathbf{x})$ being separable, it follows immediately that the set of all discontinuity points is at most enumerable.

A discontinuity point of the innovation spectrum will not necessarily be a discontinuity point of the process, nor conversely. We shall make some remarks concerning the relations between these two kinds of discontinuities.

Let us first consider the case of a left discontinuity of the innovation spectrum, that is, a point t such that

$$\mathfrak{M}(t) = \mathfrak{H}(\mathbf{x}, t) - \mathfrak{H}(\mathbf{x}, t - 0) \neq 0. \tag{56}$$

Then it is easily shown that

$$\mathbf{x}(t) - \mathbf{x}(t - 0) \neq 0, \tag{57}$$

so that t is also a left discontinuity point of the process. Further, if $y^{(j)}$ denotes the projection of $x^{(j)}(t) - x^{(j)}(t - 0)$ on $\mathfrak{M}(t)$, we have $y^{(j)} \neq 0$ for at least one j, and the subspace $\mathfrak{M}(t)$ is spanned by the variables $y^{(1)}, \cdots, y^{(q)}$, and has thus at most q dimensions.

Thus in particular (56) implies (57). The converse statement is however not true: a left discontinuity of the process may, in fact, be a continuity point of the innovation spectrum.

Proceeding now to the case of a right discontinuity, it can be shown that neither of the two relations

$$\mathfrak{N}(t) = \mathfrak{H}(\mathbf{x}, t + 0) - \mathfrak{H}(\mathbf{x}, t) \neq 0 \tag{58}$$

and

$$\mathbf{x}(t + 0) - \mathbf{x}(t) \neq 0 \tag{59}$$

implies the other. In fact, it can be shown by examples that (58) may be satisfied even in a continuity point of the process, while on the other hand (59) may be satisfied even in a continuity point of the innovation spectrum.

The only implication that exists between the relations (56), (57), (58), and (59) is thus that (56) implies (57).

As in section 3, we now denote by $P_t z$ the projection of any point $z \in \mathfrak{H}(\mathbf{x})$ on the subspace $\mathfrak{H}(\mathbf{x}, t)$. When t increases from $-\infty$ to $+\infty$, the P_t form a never

decreasing set of projections, with $P_{-\infty} = 0$ and $P_{+\infty} = I$. For $h > 0$, the difference $P_{t+h} - P_t$ is the projection on $\mathfrak{H}(\mathbf{x}, t + h) - \mathfrak{H}(\mathbf{x}, t)$. The limits $P_{t\pm 0}$ exist for every t, and are the projections on $\mathfrak{H}(\mathbf{x}, t \pm 0)$ respectively.

For an arbitrary random variable z in $\mathfrak{H}(\mathbf{x})$, we now define a stochastic process by writing for all real t

$$z(t) = P_t z. \tag{60}$$

It then follows from the above that $z(t)$ defines a complex-valued stochastic process with *orthogonal increments*, such that

$$\begin{aligned} z(-\infty) &= 0, \qquad z(+\infty) = z, \\ Ez(t) &= 0, \qquad E|z(t)|^2 = F(t, z), \end{aligned} \tag{61}$$

where $F(t, z)$ is, for any fixed z, a real, never decreasing and bounded function of t, such that

$$F(-\infty, z) = 0, \qquad F(+\infty, z) = E|z|^2. \tag{62}$$

The points of increase of $z(t)$, that is, the points t such that for any $h > 0$

$$E|z(t + h) - z(t - h)|^2 = F(t + h, z) - F(t - h, z) > 0, \tag{63}$$

form a subset of the innovation spectrum of $\mathbf{x}(t)$. Similarly, the left (right) discontinuities of $z(t)$ form a subset of the left (right) discontinuities of the innovation spectrum. Any increment $z(t + h) - z(t)$ belongs to the subspace $\mathfrak{H}(\mathbf{x}, t + h) - \mathfrak{H}(\mathbf{x}, t)$, and may thus be regarded as a part of the innovation received by the $\mathbf{x}(t)$ process during the interval $(t, t + h)$.

We now denote by $\mathfrak{L}(z)$ the subspace of $\mathfrak{H}(\mathbf{x})$ spanned by all the variables $z(u)$ for $-\infty < u < +\infty$, and by $\mathfrak{L}^*(z)$ the set of all random variables y representable in the form

$$y = \int_{-\infty}^{\infty} g(u)\, dz(u) \tag{64}$$

with

$$E|y|^2 = \int_{-\infty}^{\infty} |g(u)|^2\, dF(u, z) < \infty. \tag{65}$$

If no u is at the same time a left and a right discontinuity of $z(u)$, then $\mathfrak{L}(z)$ and $\mathfrak{L}^*(z)$ are identical (see Doob [4], pp. 425–429). The variable y given by (64 will belong to $\mathfrak{H}(\mathbf{x}, t)$ if, and only if, we have $g(u) = 0$ for almost all $u > t$, "almost all" referring to the $F(u, z)$-measure on the u-axis.

By means of the theory of spectral multiplicity in Hilbert space (see, for example, Stone [10], chapter VII, and Halmos [6]), we can now show that with any $\mathbf{x}(t)$ process satisfying (C_1) and (C_2) it is possible to associate a number N, which may be a finite positive integer or equal to $+\infty$, such that we can find N random variables $z_1, \cdots, z_N$ belonging to $\mathfrak{H}(\mathbf{x})$, with the properties

(a) $\mathfrak{L}(z_n) = \mathfrak{L}^*(z_n), \qquad n = 1, \cdots, N.$

(b) $\mathfrak{L}(z_m) \perp \mathfrak{L}(z_n), \qquad m \neq n.$

(c) $$\mathfrak{H}(\mathbf{x}) = \mathfrak{L}(z_1) + \cdots + \mathfrak{L}(z_N).$$

(d) N is the smallest number having the properties (a), (b), and (c).

In particular, since any component variable $x^{(j)}(t)$ of $\mathbf{x}(t)$ evidently belongs to $\mathfrak{H}(\mathbf{x}, t)$, we have the expression

$$x^{(j)}(t) = \sum_{k=1}^{N} \int_{-\infty}^{t} g_k^{(j)}(t, u)\, dz_k(u) \tag{66}$$

for $j = 1, \cdots, q$. If $N = \infty$, the series appearing here will converge in quadratic mean, so that

$$\sum_{k=1}^{N} \int_{-\infty}^{t} |g_k^{(j)}(t, u)|^2\, dF(u, z_k) < \infty. \tag{67}$$

If now we define a column vector

$$\mathbf{z}(u) = \{z_1(u), \cdots, z_N(u)\} \tag{68}$$

and a $q \times N$ matrix

$$\mathbf{G}(t, u) = \{g_k^{(j)}(t, u)\}, \qquad j = 1, \cdots, q;\ k = 1, \cdots, N, \tag{69}$$

we finally obtain the required expression for the vector variable $\mathbf{x}(t)$ in terms of past and present innovations of the process, as stated in the following theorem.

THEOREM 4. *The vector variable* $\mathbf{x}(t)$ *of any stochastic process satisfying conditions* (C_1) *and* (C_2) *can be expressed in the form*

$$\mathbf{x}(t) = \int_{-\infty}^{t} \mathbf{G}(t, u)\, d\mathbf{z}(u), \tag{70}$$

where $\mathbf{z}(u)$ *is an* N*-dimensional vector process with orthogonal increments, while* $\mathbf{G}(t, u)$ *is a* $q \times N$ *matrix, in accordance with* (68) *and* (69). *The development* (66) *formally obtained for the component* $x^{(j)}(t)$ *is then convergent as shown by* (67).

It will be seen that this is directly analogous to the last part of theorem 1, except that certain sums have been replaced by integrals, and that the q-dimensional random vector $\boldsymbol{\xi}_p$ has been replaced by the N-dimensional vector $\mathbf{z}(u)$. It is the fact that the multiplicity N may have any integral value from 1 to ∞ that introduces additional complication into the continuous case. It is possible to construct examples corresponding to any given value of N, even when it is required that $\mathbf{x}(t)$ be everywhere continuous (or even differentiable in quadratic mean). We finally state the following theorem, which is the continuous analogy of theorem 2.

THEOREM 5. *Let* $h > 0$ *be given. For any* $\mathbf{x}(t)$ *process satisfying conditions* (C_1) *and* (C_2), *the best prediction of the component* $x^{(j)}(t)$ *in terms of all variables* $x^{(1)}(u), \cdots, x^{(q)}(u)$ *with* $u \leqq t - h$ *will be*

$$P_{t-h}x^{(j)}(t) = \sum_{k=1}^{N} \int_{-\infty}^{t-h} g_k^{(j)}(t, u)\, dz_k(u), \tag{71}$$

with the corresponding error of prediction

$$\sigma_{ih}^{(j)} = ||x^{(i)}(t) - P_{t-h}x^{(i)}(t)|| = \left\{\sum_{k=1}^{N}\int_{t-h}^{t} |g_k^{(j)}(t, u)|^2\, dF(u, z_k)\right\}^{1/2} \tag{72}$$

PART II. ON TWO CLASSES OF PROCESSES ADMITTING SPECTRAL REPRESENTATIONS

6. Harmonizable processes

We shall now consider a one-dimensional process with discrete time, such that x_n is given by a stochastic Fourier-Stieltjes integral

$$x_n = \int_0^{2\pi} e^{inu}\, dz(u), \tag{73}$$

where $z(u)$ denotes, for $0 \leqq u \leqq 2\pi$, a complex-valued random variable satisfying the conditions

$$Ez(u) = 0, \qquad E\{z(u)\bar{z}(v)\} = F(u, v). \tag{74}$$

It will be assumed that the complex-valued covariance function $F(u, v)$ is of bounded variation over the square C defined by $0 \leqq u, v \leqq 2\pi$, in the sense that, for every subdivision of C in a finite number of rectangles, we have

$$\sum |\Delta_2 F| < K, \tag{75}$$

the sum being extended over all the rectangles, and the constant K being independent of the subdivision.

The integral (73) will then exist as a limit in quadratic mean of certain Riemann sums. Processes of this type have been introduced by Loève [7], [8], and have been called by him *harmonizable processes*. The covariance function corresponding to the process defined by (73) is

$$R(m, n) = E(x_m\bar{x}_n) = \int_0^{2\pi}\int_0^{2\pi} e^{i(mu-nv)}\, dF(u, v). \tag{76}$$

Conversely, if the covariance function of a certain x_n process is given by (76), where $F(u, v)$ is a covariance function satisfying (75) it is known (Loève [8], Cramér [1]) that there exists a process $z(u)$ satisfying (74), and such that x_n is given by the integral (73).

Without changing the value of the integral (73), we can always suppose that $z(u)$ is everywhere continuous to the right in quadratic mean, so that $z(u + 0) = z(u)$. The function $F(u, v)$ then defines a *complex mass distribution* over C, such that the mass carried by any rectangle $h < u \leqq h + \Delta h$, $k < v \leqq k + \Delta k$ is equal to the second-order difference $\Delta_2 F(u, v)$ corresponding to this rectangle.

It follows from the Hermite-symmetric properties of covariances that the masses carried by two sets of points symmetrically situated with respect to the diagonal $u = v$ of the square C are always complex conjugates. If a point set

belonging to the diagonal $u = v$ carries a mass different from zero, this mass will be real and positive.

The function $F(u, v)$ will be called the *spectral function* of the x_n process, while the distribution defined by F is the *spectral distribution* of the process.

In the particular case when the whole spectral mass is situated on the diagonal $u = v$, it follows from the general symbolic relation

$$E\{dz(u)\, d\bar{z}(v)\} = d_{u,v}F(u, v) \tag{77}$$

that the $z(u)$ process has orthogonal increments, and so in this case the x_n process is *stationary*.

In the general case, $F(u, v)$ may be represented as a sum of three components, each of which is a covariance of bounded variation over C,

$$F = F_1 + F_2 + F_3. \tag{78}$$

Here F_1 is absolutely continuous, with a *spectral density* $f_1(u, v)$ such that

$$F_1(u, v) = \int_0^u \int_0^v f_1(s, t)\, ds\, dt. \tag{79}$$

On the other hand, the F_2 and F_3 distributions both have their total masses concentrated in sets of two-dimensional Lebesgue measure zero. For F_2 this set is at most enumerable, each point carrying a mass different from zero, while the F_3 set is nonenumerable, and each single point carries the mass zero. In the stationary case, the F_1 component is absent, while the F_2 and F_3 components have their total masses situated on the diagonal $u = v$.

A sufficient condition that the harmonizable x_n process given by (73) will be deterministic can be obtained in the following way. The x_n process will be deterministic if, and only if, for every n and every $h > 0$ we can find a finite number of constants $c_0, c_1, \cdots, c_r$ such that the quantity

$$W = E|x_n - c_0x_{n-h} - c_1x_{n-h-1} - \cdots - c_rx_{n-h-r}|^2 \geqq 0 \tag{80}$$

will be arbitrarily small. Writing

$$g(u) = e^{inu} - c_0e^{i(n-h)u} - \cdots - c_re^{i(n-h-r)u}, \tag{81}$$

it follows from (76) that we have

$$W = \int_0^{2\pi} \int_0^{2\pi} g(u)\bar{g}(v)\, dF(u, v) \tag{82}$$

and hence by the Schwarz inequality

$$W^2 \leqq \int_0^{2\pi} \int_0^{2\pi} |g(u)|^2\, |dF(u, v)| \int_0^{2\pi} \int_0^{2\pi} |g(v)|^2\, |dF(u, v)|. \tag{83}$$

By the symmetry of the spectral distribution, the two factors in the last member are equal, so that we obtain

$$W \leqq \int_0^{2\pi} |g(u)|^2\, dG(u), \tag{84}$$

where

$$G(u) = \int_0^u \int_0^{2\pi} |dF(s, t)|. \tag{85}$$

Now $G(u)$, being a never decreasing and bounded function of u, has almost everywhere in $(0, 2\pi)$ a nonnegative derivative $G'(u)$, and the integral

$$\int_0^{2\pi} \log G'(u)\, du \tag{86}$$

will be finite or equal to $-\infty$. In particular, if $G'(u) = 0$ on a set of positive measure, the integral will certainly have the value $-\infty$.

If the integral (86) has the value $-\infty$, it follows from well-known theorems in the prediction theory for stationary processes that the coefficients c_j can be chosen so as to make the second member of (80) as small as we please. Thus we have the following result (Cramér [2]).

THEOREM 6. *If we have*

$$\int_0^{2\pi} \log G'(u)\, du = -\infty \tag{87}$$

the x_n process is deterministic.

In particular, if the F_1 and F_3 components in (78) are absent, so that the whole mass of the F distribution is concentrated in isolated points, it will be seen that $G'(u) = 0$ almost everywhere, so that (87) will certainly hold, and the x_n process will be deterministic.

Consider now, on the other hand, a process x_n with a spectral function F having an absolutely continuous component F_1 not identically zero. Moreover, let us suppose that the spectral density $f_1(u, v)$ corresponding to F_1 belongs to L_2 over the square C. For such a process, we shall give a sufficient condition that it is nondeterministic. Let

$$f_1(u, v) = \sum_{p=1}^{\infty} \mu_p \varphi_p(u) \bar{\varphi}_p(v) \tag{88}$$

be the expansion of $f_1(u, v)$ in terms of its eigenvalues μ_p and eigenfunctions $\varphi_p(u)$. The μ_p are real and positive, the $\varphi_p(u)$ are a set of orthonormal functions in $(0, 2\pi)$, and the series converges in quadratic mean over C. We then have

THEOREM 7. *Suppose that, in the expansion* (88), *there is a p such that the Fourier series of the eigenfunction $\varphi_p(u)$*

$$\varphi_p(u) \sim \sum_{q=-\infty}^{\infty} b_{pq} e^{-iqu}, \qquad \sum_{q=-\infty}^{\infty} |b_{pq}|^2 < \infty \tag{89}$$

is "one-sided" in the sense that for a certain m it satisfies the conditions

$$b_{pq} = 0 \text{ for } q < m, \qquad b_{pm} \neq 0. \tag{90}$$

Then the x_n process is nondeterministic, and the point $n = m$ belongs to its innovation spectrum.

Taking $n = m$ and $h = 1$ in the expressions (80) and (81) for W and $g(u)$ we have, in fact (see Riesz and Nagy [9], p. 240),

$$W = \int_0^{2\pi}\int_0^{2\pi} g(u)\bar{g}(v)\,dF(u,v) \geqq \int_0^{2\pi}\int_0^{2\pi} g(u)\bar{g}(v)f_1(u,v)\,du\,dv = \sum_{p=1}^{\infty} \mu_p \left|\int_0^{2\pi} g(u)\varphi_p(u)\,du\right|^2 \tag{91}$$

and thus by hypothesis

$$W \geqq \mu_p \left|\int_0^{2\pi} g(u)\varphi_p(u)\,du\right|^2 = 4\pi^2\mu_p|b_{pm}|^2 \tag{92}$$

independently of the choice of the coefficients c_j. This obviously signifies that x_m cannot be predicted exactly in terms of the variables $x_{m-1}, x_{m-2}, \cdots$, so that the prediction error

$$||x_m - P_{m-1}x_m|| \tag{93}$$

is positive, and m belongs to the innovation spectrum of the x_n process, which is thus nondeterministic.

It follows from well-known theorems that, when the conditions of theorem 7 are satisfied, we have

$$\int_0^{2\pi} \log|\varphi_p(u)|\,du > -\infty. \tag{94}$$

The converse of this statement is, however, not true; (94) may be satisfied even in a case when $\varphi_p(u)$ does not have a one-sided Fourier expansion. A simple example is obtained by taking

$$2\pi f_1(u,v) = \varphi(u)\bar{\varphi}(v) \tag{95}$$

with

$$\varphi(u) = \begin{cases} e^{iu}, & 0 \leqq u \leqq \pi, \\ e^{-iu}, & \pi < u \leqq 2\pi. \end{cases} \tag{96}$$

It can also be shown by examples that there are nondeterministic processes with a spectral density $f_1(u, v)$ belonging to L_2 that do not satisfy the conditions of theorem 7 for any value of p.

By imposing a further restrictive condition on the behavior of the spectral density it is possible, however, to obtain a criterion which is both necessary and sufficient in order that a given harmonizable process be nondeterministic, and even have an a priori given set of integers as its innovation spectrum (Cramér [3]). Thus, in particular, it follows that, any set of integers being given, there always exists a harmonizable process having this set as its innovation spectrum.

7. Processes with normal shift operator

For a stationary process with discrete time, we have the integral representation

$$x_n = \int_0^{2\pi} e^{inu}\,dz(u), \tag{97}$$

where $z(u)$ has orthogonal increments. In the preceding section we considered the generalization obtained by dropping the assumption that $z(u)$ has orthogonal increments, and we saw that this leads to the class of harmonizable processes.

We now consider a different kind of generalization of (97), which leads to a different class of processes. To this effect, we now regard the integration variable u in (97) as a *complex* variable, and suppose that the integration is extended over a certain domain D in the plane of u, and that $z(u)$ is defined for all u belonging to D.

After an appropriate change of variables, the integral corresponding to (97) then takes the form

$$x_n = \int_D w^n \, dz(\rho, \lambda), \tag{98}$$

where $w = \rho \exp(i\lambda)$, while $z(\rho, \lambda)$ is a random variable satisfying the conditions

$$Ez(\rho, \lambda) = 0, \qquad E|z(\rho, \lambda)|^2 < K \tag{99}$$

for all ρ, λ such that w belongs to D. As in the stationary case, we still suppose that $z(\rho, \lambda)$ has orthogonal increments, so that we have in the usual symbolism

$$\begin{aligned} E\{dz(\rho_1, \lambda_1)\, \overline{dz}(\rho_2, \lambda_2)\} &= 0, \qquad w_1 \neq w_2, \\ E|dz(\rho, \lambda)|^2 &= dF(\rho, \lambda), \end{aligned} \tag{100}$$

where $F(\rho, \lambda)$ is a nonnegative and never decreasing function of ρ and λ, which is bounded throughout D. The integral (98) can then be defined in the same way as before, and we obtain

$$R(m, n) = E(x_m \bar{x}_n) = \int_D w^m \overline{w^n} \, dF(\rho, \lambda). \tag{101}$$

The function $F(\rho, \lambda)$ will be called the *spectral function* of the x_n process, and defines the *spectral distribution* of the process, which is a distribution of real and positive mass over the domain D.

We now introduce the further assumption that the domain D is entirely situated within the ring

$$\rho_1 \geqq \rho \geqq \rho_0 > 0. \tag{102}$$

We observe that this includes the particular case of a stationary process, when the domain D reduces to the unit circle.

In the present case we obtain from (101) for any complex constants c_j and any positive integer Q

$$0 \leqq E\left|\sum_{j=-Q}^{Q} c_j x_j\right|^2 = \sum_{j,k=-Q}^{Q} c_j \bar{c}_k R(j, k) = \int_D \left|\sum_{-Q}^{Q} c_j w^j\right|^2 dF, \tag{103}$$

and hence

$$\begin{aligned} E\left|\sum_{-Q}^{Q} c_j x_{j+1}\right|^2 &\leqq \rho_1^2 \, E\left|\sum_{-Q}^{Q} c_j x_j\right|^2, \\ E\left|\sum_{-Q}^{Q} c_j x_{j-1}\right|^2 &\leqq \frac{1}{\rho_0^2} E\left|\sum_{-Q}^{Q} c_j x_j\right|^2. \end{aligned} \tag{104}$$

According to Getoor [5], these inequalities imply that there is a *shift operator* N uniquely defined and bounded throughout the Hilbert space $\mathfrak{H}(x)$ of the x_n process, and such that

$$N^m x_n = x_{m+n} \tag{105}$$

for all m, $n = 0, \pm 1, \cdots$. It also follows from the work of Getoor that in our case N is a *normal* operator in $\mathfrak{H}(x)$. We have, in fact,

$$\begin{aligned}(Nx_m, x_n) &= E(x_{m+1}\bar{x}_n) = \int_D w^{m+1}\overline{w^n}\, dF \\ &= \int_D w^m(\overline{\bar{w}w^n})\, dF = E(x_m\bar{y}_n),\end{aligned} \tag{106}$$

where

$$y_n = N^*x_n = \int_D \bar{w}w^n\, dz. \tag{107}$$

It follows that

$$NN^*x_n = N^*Nx_n = \int_D |w|^2 w^n\, dz. \tag{108}$$

Thus N commutes with its adjoint N^*, and consequently N is normal. In the particular case of a stationary process, when the spectral mass is wholly situated on the unit circle, it is well known that N is even a unitary operator.

By an argument quite similar to that used for the deduction of the inequalities (104), we obtain for $n > 0$

$$\begin{aligned}\rho_0^{2n} E\left|x_0 - \sum_{j=1}^{Q} c_j x_{-j}\right|^2 &\leqq E\left|x_n - \sum_{j=1}^{Q} c_j x_{n-j}\right|^2 \\ &\leqq \rho_1^{2n} E\left|x_0 - \sum_{j=1}^{Q} c_j x_{-j}\right|^2.\end{aligned} \tag{109}$$

The coefficients c_j being arbitrary, this shows that we have for the prediction errors σ_{nh}, where h is any positive integer,

$$\rho_0^{2n}\sigma_{0h}^2 \leqq \sigma_{nh}^2 \leqq \rho_1^{2n}\sigma_{0h}^2. \tag{110}$$

For $n < 0$ we obtain in the same way

$$\rho_1^{2n}\sigma_{0h}^2 \leqq \sigma_{nh}^2 \leqq \rho_C^{2n}\sigma_{0h}^2. \tag{111}$$

From these inequalities, we obtain directly the following theorem.

THEOREM 8. *If the x_n process defined by* (98) *and* (102) *is nondeterministic, we have $\sigma_{nh} > 0$ for all n and all $h > 0$. In particular, the innovation spectrum of the process then contains all $n = 0, \pm 1, \cdots$.*

Suppose now that the spectral distribution defined by $F(\rho, \lambda)$ has a nonvanishing absolutely continuous component. We may then write

$$\begin{aligned}J = E\left|x_0 - \sum_1^Q c_j x_{-j}\right|^2 &= \int_D \left|1 - \sum_1^Q \frac{c_j}{w^j}\right|^2 dF \\ &\geqq \int_{\rho_0}^{\rho_1}\int_0^{2\pi} \left|1 - \sum_1^Q \frac{c_j}{w^j}\right|^2 f(\rho, \lambda)\, d\rho\, d\lambda,\end{aligned} \tag{112}$$

where $f(\rho, \lambda)$ is nonnegative and integrable. From the inequalities between arithmetic and geometric means we further obtain

$$\frac{1}{2\pi(\rho_1 - \rho_0)} J \geqq \exp\left\{\frac{1}{2\pi(\rho_1 - \rho_0)} \int_{\rho_0}^{\rho_1} \int_0^{2\pi} \left[\log f(\rho, \lambda) + 2 \log\left|1 - \sum_1^Q \frac{c_j}{w^j}\right|\right] d\rho\, d\lambda\right\} = G(f) \exp\left\{\frac{1}{\pi(\rho_1 - \rho_0)} \int_{\rho_0}^{\rho_1} \int_0^{2\pi} \log\left|1 - \sum_1^Q \frac{c_j}{w^j}\right| d\rho\, d\lambda\right\}, \tag{113}$$

where

$$G(f) = \exp\left\{\frac{1}{2\pi(\rho_1 - \rho_0)} \int_{\rho_0}^{\rho_1} \int_0^{2\pi} \log f(\rho, \lambda)\, d\rho\, d\lambda\right\}. \tag{114}$$

From Jensen's theorem we obtain, however,

$$\exp\left\{\frac{1}{\pi(\rho_1 - \rho_0)} \int_{\rho_0}^{\rho_1} \int_0^{2\pi} \log\left|1 - \sum_1^Q \frac{c_j}{w^j}\right| d\rho\, d\lambda\right\} = \exp\left\{\frac{2}{\rho_1 - \rho_0} \int_{\rho_0}^{\rho_1} \log \frac{|w_1 \cdots w_k|}{\rho^k}\, d\rho\right\} \geqq 1, \tag{115}$$

where $w_1, \cdots, w_k$ are the zeros of $1 - \sum_1^Q c_j w^{-j}$ outside the circle $|w| = \rho$. Consequently

$$\frac{1}{2\pi(\rho_1 - \rho_0)} J \geqq G(f), \tag{116}$$

and it follows that, if $G(f) > 0$, then the x_n process is nondeterministic, and we have

$$\sigma_{n1}^2 \geqq 2\pi(\rho_1 - \rho_0) G(f). \tag{117}$$

In the particular case when there is an expansion

$$\log f(\rho, \lambda) = \sum_{-\infty}^{\infty} \frac{b_j}{\rho^{|j|}} e^{-ij\lambda}, \qquad b_{-j} = \bar{b}_j, \tag{118}$$

absolutely convergent for $\rho \geqq \rho_0$, it can even be proved that the sign of equality holds in (117).

Finally, we may observe that it is also possible to give a sufficient condition for a deterministic process, corresponding at least partly to theorem 6. In fact, it can be shown that if the spectral distribution defined by $F(\rho, \lambda)$ is discrete, and if the set of points carrying a positive mass has at most a finite number of limiting points, then the x_n process is deterministic.

REFERENCES

[1] H. CRAMÉR, "A contribution to the theory of stochastic processes," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 1951, pp. 329–339.

[2] ———, "Remarques sur le problème de prédiction pour certaines classes de processus stochastiques," *Colloque sur le Calcul des Probabilités*, Paris, 1959, pp. 103–112.
[3] ———, "On the linear prediction problem for certain stochastic processes," *Ark. Mat.*, Vol. 4 (1959), pp. 45–53.
[4] J. L. DOOB, *Stochastic Processes*, New York, Wiley, 1953.
[5] R. K. GETOOR, "The shift operator for non-stationary stochastic processes," *Duke Math. J.*, Vol. 23 (1956), pp. 175–187.
[6] P. R. HALMOS, *Introduction to Hilbert Space and the Theory of Spectral Multiplicity*, New York, Chelsea, 1951.
[7] M. LOÈVE, "Fonctions aléatoires du second ordre," appendix to P. Lévy, *Processus Stochastiques et Mouvement Brownien*, Paris, Gauthier-Villars, 1948.
[8] ———, *Probability Theory*, New York, Van Nostrand, 1955.
[9] F. RIESZ and B. SZ.-NAGY, *Leçons d'Analyse Fonctionnelle*, Paris, 1955 (3rd ed.).
[10] M. H. STONE, *Linear Transformations in Hilbert Space and their Applications to Analysis*, New York, The American Mathematical Society, 1932.
[11] N. WIENER and P. MASANI, "The prediction theory of multivariate stochastic processes, I," *Acta Math.*, Vol. 98 (1957), pp. 111–150.
[12] H. WOLD, *A Study in the Analysis of Stationary Time Series*, Stockholm, Almqvist & Wiksell, 1954.