ON PREDICTION IN STATIONARY TIME SERIES

By Herman O. A. Wold

Uppsala University

Summary. In time series analysis there are two lines of approach, here called the *functional* and the *stochastic*. In the former case, the given time series is interpreted as a mathematical function, in the latter case as a random specimen out of a universe of mathematical functions. The close relation between the two approaches is in section 2 shown to amount to a genuine isomorphism. Considering the problem of prediction from this viewpoint, the author gives in sections 3–4 the functional equivalence of his earlier theorem on the decomposition of a stationary stochastic process with a discrete time parameter (see [9], theorem 7). In section 5 the decomposition theorem is applied to the problem of linear prediction. Finally in section 6 a few comments are made. Since various aspects of the isomorphism in question are known, this paper might be regarded as essentially expository.

1. Introductory. Let the sequence

$$\cdots, x_{t-1}, x_t, x_{t+1} \cdots \tag{1}$$

be an empirical time series such that no clear trend is present in the average level, in the variance or in any other structural properties of the series which we might choose to consider. Such series are usually called *stationary* as distinct from *evolutive*, terms which of course are somewhat loose when referring to empirical data. We shall consider two approaches in the theoretical analysis of stationary series. It is convenient to allow x_t to be complex; the conjugate complex of x_t is denoted $\bar{x}_t$.

In the functional approach, the sequence (1) is regarded as forming an infinite sequence, say $\{x_t\}$, where t runs from $-\infty$ to $+\infty$. To define stationarity, let us for any infinite sequence $\{z_t\}$ write

$$M[z_t] = \lim \frac{1}{t_2 - t_1 + 1} \sum_{t=t_1}^{t_2} z_t \qquad (t_1 \to -\infty, t_2 \to +\infty). \tag{2}$$

The limit $M[z_t]$, which will be called "the average of z_t", is clearly independent of t. It is also seen that a necessary and sufficient condition for $M[z_t]$ to exist is that the same average should be obtained when t_1 is kept fixed while $t_2 \to +\infty$, and when t_2 is kept fixed while $t_1 \to -\infty$. The stationarity of the sequence (1) may now be brought out by assumptions of the type that the averages $M[x_t]$ and $M[x_t \cdot \bar{x}_{t+k}]$ exist, say

$$M[x_t] = m, \qquad M[x_t \cdot \bar{x}_{t+k}] = r_k \qquad (k = 0, \pm 1, \pm 2, \cdots). \tag{3}$$

In the stochastic (or probabilistic) approach, we introduce an infinite sequence of random variables, say

$$\cdots, \xi_{t-1}, \xi_t, \xi_{t+1}, \cdots \qquad (-\infty < t < +\infty), \tag{4}$$

or briefly $\{\xi_t\}$. The sequence $\{\xi_t\}$ may be regarded as the generalization of the notion of multi-dimensional variable, say $[\xi_1, \cdots, \xi_n]$, to an infinite number of components ξ_i. According to a basic theorem by A. Kolmogoroff (see e.g. [9], §11), the probability distribution of the sequence $\{\xi_t\}$ may be defined by specifying for any finite set of variables, say $[\xi_{t_1}, \cdots, \xi_{t_n}]$, its multi-dimensional distribution function, say

$$(5) \qquad F(u_1, \cdots, u_n; t_1, \cdots, t_n) = \text{Prob}\,(\xi_{t_1} \leq u_1, \cdots, \xi_{t_n} \leq u_n).$$

The sequence $\{\xi_t\}$ thus defined is said to constitute a *stochastic process*. As is sufficient for our purpose, we confine ourselves to the case when the time parameter t is restricted to discrete values, $t = 0, \pm 1, \pm 2, \cdots$.

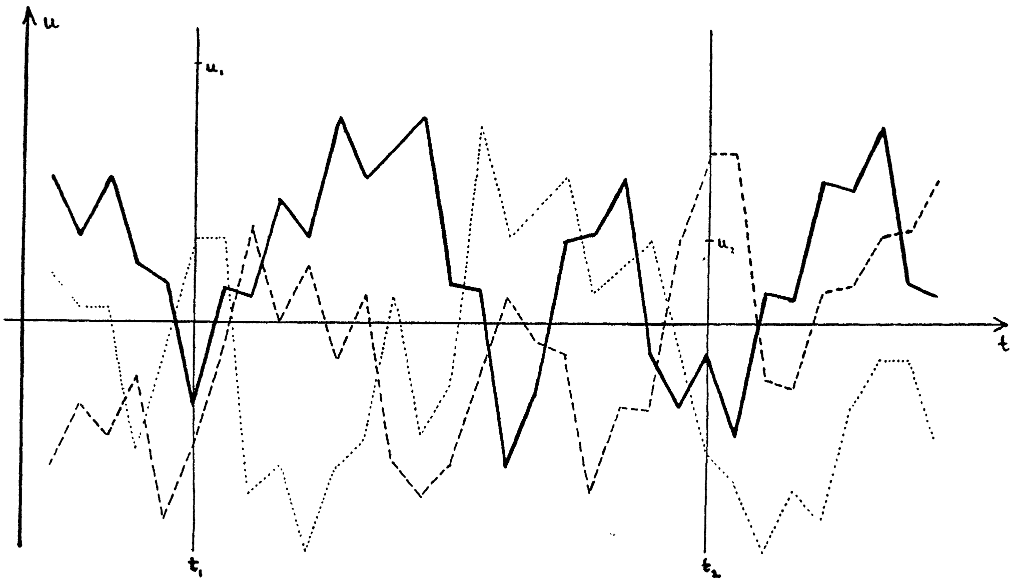

FIG. 1

Now in the stochastic approach, the empirical time series (1) is regarded as a sample specimen, a *realization*, of the stochastic process $\{\xi_t\}$, just as a point $[x_1, \cdots, x_n]$ in an n-dimensional space may be regarded as a sample specimen of a multidimensional variable $[\xi_1, \cdots, \xi_n]$. In line with this interpretation, the process $\{\xi_t\}$ may be regarded as a universe of individual realizations such as (1) (see the graph). Taking out a realization at random from this universe, we shall have the probability,

$$F(u_1; t_1) = \text{Prob}\,(\xi_{t_1} \leq u_1),$$

that the value taken on by the realization at the time point t_1 will be $\leq u_1$; similarly,

$$F(u_1, u_2; t_1, t_2) = \text{Prob}\,(\xi_{t_1} \leq u_1, \xi_{t_2} \leq u_2),$$

is the joint probability that the values taken on by the realization at t_1 and t_2 will be $\leq u_1$ and $\leq u_2$ respectively.

Any expectation referring to the variables (4) may be expressed in terms of the distribution functions (5), for instance

$$E[\xi_t] = \int_{-\infty}^{\infty} u\, d_u F(u; t), \qquad E[\xi_{t_1} \cdot \xi_{t_2}] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u \cdot v\, d^2_{u,v} F(u, v; t_1, t_2).$$

Again interpreting in terms of the universe of realizations, $E[\xi_t]$, say, is the average, over this universe, of the value taken by the realizations at the time point t.

The above definition of a stochastic process (4) being perfectly general, we have to impose special assumptions if we wish to take into account particular properties of the given time series (1). Thus stationarity of the process (4) may be defined by assuming that any probability of the type (5) will remain the same if $t_1, \cdots, t_n$ is replaced by $t_1 + t, \cdots, t_n + t$, where t is arbitrary. Alternatively, and more generally, the stationarity of the sequence (1) may be brought out in this approach by assuming that the expectations

$$E[\xi_t] = \mu, \qquad E[\xi_t \cdot \bar{\xi}_{t+k}] = \rho_k$$

exist and are independent of t.

2. The functional and stochastic approaches are closely related as to problems and results. A typical example is that r_k and ρ_k as defined above allow the representations[1]

$$(6) \qquad r_k = \int_{-\pi}^{\pi} e^{ik\lambda}\, dF(\lambda), \quad \rho_k = \int_{-\pi}^{\pi} e^{ik\lambda}\, d\Phi(\lambda), \qquad (k = 0, \pm 1, \pm 2, \cdots),$$

where $F(\lambda)$ and $\Phi(\lambda)$ are real, bounded and never decreasing functions. We shall now show that the parallelism between the two approaches amounts to a mathematical isomorphism. On the one hand, we recall that A. Kolmogoroff [3], [4] has introduced and studied the notion of a stationary sequence in Hilbert space,—let such a sequence be denoted $\{X_t\}$—, and shown that a stationary stochastic process $\{\xi_t\}$ forms a particular realization of this general, abstract $\{X_t\}$. On the other hand the following elementary lemma shows that another realization of $\{X_t\}$ may be formed on the basis of a stationary sequence $\{x_t\}$ such as (1).

LEMMA. *Let* $\{x_t\}$ *be a sequence of type* (1) *which satisfies the conditions* (3) *but is arbitrary in other respects. We write*

$$(7) \qquad \{x_t\} = \cdots, \mathbf{x}_{t-1}, \mathbf{x}_t, \mathbf{x}_{t+1}, \cdots,$$

where $\mathbf{x}_t = \{x_t\}$, *and* $\mathbf{x}_{t+k}$ *is obtained from* $\mathbf{x}_t$ *by replacing* x_t *by* x_{t+k} *for every* t.

[1] As to r_k, see N. Wiener [8], who treats the case of a continuous time parameter t. As to ρ_k, see H. Wold [9], p. 66, and A. Kolmogoroff [4], p. 5.

For the elements $\boldsymbol{x}_t$, let multiplication by a real or complex constant and addition be defined by

$$a\boldsymbol{x}_t = \{ax_t\}, \qquad \boldsymbol{x}_t + \boldsymbol{y}_t = \{x_t + y_t\},$$

and let $\boldsymbol{R}$ be the class formed by all elements of the type

$$c_{-n}\boldsymbol{x}_{t-n} + c_{-n+1}\boldsymbol{x}_{t-n+1} + \cdots + c_0\boldsymbol{x}_t + \cdots + c_n\boldsymbol{x}_{t-n},$$

where n and $c_{-n}, \cdots, c_n$ are arbitrary. Let the inner product $(\boldsymbol{x}_t, \boldsymbol{y}_t)$ of two elements $\boldsymbol{x}_t = \{x_t\}, \boldsymbol{y}_t = \{y_t\}$ in $\boldsymbol{R}$ be defined by

$$(x_t, y_t) = M[x_t \cdot \bar{y}_t],$$

and let $\boldsymbol{R}'$ be the closure of $\boldsymbol{R}$.

Then $\boldsymbol{R}'$ is a space the dimension of which is denumerable or finite. In the former case, $\boldsymbol{R}'$ satisfies the conditions of a Hilbert space H, in the latter case it can be extended to a Hilbert space H. In any case, the relations

$$U\boldsymbol{x}_t = \boldsymbol{x}_{t+1}, \qquad -\infty < t < +\infty, \tag{8}$$

define a unitary transformation U in H.

The first statement of the theorem is obvious. It is also easily verified that $\boldsymbol{R}'$ satisfies the conditions A–C of an abstract Hilbert space as defined by B. v. Sz. Nagy [7]. If $\boldsymbol{R}'$ is of finite dimension, a suitable extension will make $\boldsymbol{R}'$ satisfy the conditions A–E of a Hilbert space as defined by M. H. Stone [6]. The transformation U is clearly unitary; it is also plain that the definition (8) of U extends to the whole of H.

Now since both (4) and (7) are particular realizations of a stationary sequence $\{X_t\}$ in Hilbert space, any theorem on such a sequence $\{X_t\}$ will give, as immediate corollaries, similar theorems on a stationary sequence $\{x_t\}$ of type (1) and on a stationary stochastic process $\{\xi_t\}$. Generally speaking, the former corollary will involve averages of one or more functional sequences $\{x_t\}$, $\{y_t\}$, $\cdots$ over time t, while the latter will involve averages, for fixed t, over the realizations of one or more stochastic processes $\{\xi_t\}$, $\{y_t\}$, $\cdots$.

Let us consider the following problem of prediction in the light of the isomorphism established: Suppose the data (1) are known up to $t - 1$, say for $t - 1, t - 2, \cdots, t - n$, what can then be said about x_t, or, more generally, about x_{t+k}? One approach to the problem is to apply harmonic analysis to the given data, and to extrapolate the function obtained up to the time point $t + k$. Another approach, the one which we shall consider, is to approximate x_{t+k} directly in terms of the given data. Confining ourselves to linear prediction, and making use of n observations, the prediction formula will then be

$$\text{pred. } x_{t+k} = a_0^{(n,k)} + a_1^{(n,k)}x_{t-1} + a_2^{(n,k)}x_{t-2} + \cdots + a_n^{(n,k)}x_{t-n}. \tag{9}$$

The error of prediction, also called the residual, is denoted

$$y_{t+k}^{(n,k)} = x_{t+k} - \text{pred. } x_{t+k}. \tag{10}$$

Considering first the functional approach, we apply formula (9) for all t, thus obtaining the residuals

$$\cdots, \quad y_{t-1}^{(n,k)}, \quad y_t^{(n,k)}, \quad y_{t+1}^{(n,k)}, \quad \cdots.$$

In this approach we are led to regard the residual variance, i.e.

$$M[\,|\,y_t^{(n,k)}\,|^2], \tag{11}$$

as a total measure of the accuracy of the prediction. If we follow the stochastic approach, on the other hand, the formula (9) is applied, for fixed t, to all realizations $\{x_t\}$ of the process $\{\xi_t\}$. In this case, the variance expectation,

$$E[|\,y_t^{(n,k)}\,|^2], \tag{12}$$

is regarded as a total measure of the accuracy of the prediction. The prediction coefficients $a_i^{(n,k)}$ are determined by minimizing the expressions (11) and (12), respectively.[2] It needs no further comment that the two lines of approach in prediction theory will, thanks to the isomorphism indicated, lead to parallel results.

In a study of stationary stochastic processes, the author has earlier found a decomposition theorem which has a direct bearing on the prediction problem (see [9], theorem 7). The main purpose of the present note is to develop the corresponding decomposition for a functional sequence of the type (1). Two theorems on this line are given in sections 3–4. The proofs are briefly indicated; for further details, the reader is referred to my treatment on the stationary process [9]. In section 5, the decomposition is applied to the prediction problem. A few comments follow in section 6.

3. Auto-regression analysis of stationary time series. Let $\{x_t\}$ be an infinite sequence (1) such that the conditions (3) are fulfilled. By (9)–(10), the residuals $y_t^{(n,0)}$ will be well-defined for every n and t. According to elementary properties of least square residuals, we have

$$M[y_t^{(n,0)}] = 0; \qquad M\,[y_t^{(n,0)}\cdot \bar{x}_{t-k}] = 0 \text{ for } k = 1, 2, \cdots, n. \tag{13}$$

Since the minimum variance cannot increase if we replace n by $n + 1$, we further have

$$M[\,|\,x_t\,|^2] \geq M[\,|\,y_t^{(n,0)}\,|^2] \geq M\,[\,|\,y_t^{(n+1,0)}\,|^2] \geq 0.$$

Making $n \to \infty$, we infer that there is a constant d^2 such that

$$\lim_{n\to\infty} M[|\,y_t^{(n,0)}\,|^2] = d^2 \geq 0.$$

[2] For real sequences $\{x_t\}$ and $\{\xi_t\}$, this minimization is, of course, nothing else than the method of least squares.

Making use of the Gram-Schmidt orthogonalization procedure, it is further possible to show that there exists a sequence $\{y_t\}$ such that

$$\lim_{n\to\infty} M[|y_t^{(n,0)} - y_t|^2] = 0.$$

In the usual terminology, the sequence $\{y_t\}$ is the *limit in the mean* of the sequence $\{y_t^{(n,0)}\}$,

$$\text{(14)} \qquad \underset{n\to\infty}{\text{l.i.m.}}\,(\cdots, y_{t-1}^{(n,0)}, y_t^{(n,0)}, y_{t+1}^{(n,0)}, \cdots) = \cdots y_{t-1}, y_t, y_{t+1}, \cdots.$$

We may remark that (14) does not necessarily imply that $y_t^{(n)}$ will for a fixed t have y_t for an ordinary limit. We also note that the limiting sequence $\{y_t\}$ is not uniquely determined; for instance, the relation (14) remains valid if a finite number of the elements y_t are modified.

As is easily shown, we have

$$\text{(15)} \qquad \lim_{n\to\infty} M[|y_t^{(n,0)}|^2] = M[|y_t|^2] = M[y_t\cdot\bar{x}_t] = d^2 \geq 0,$$

and [cf. (13)]

$$\text{(16)} \qquad M[y_t\bar{x}_{t-k}] = 0, \qquad k = 1, 2, \cdots.$$

Moreover, the sequence $\{y_t\}$ is non-autocorrelated, i.e.

$$\text{(17)} \qquad M[y_t\bar{y}_{t+k}] = 0, \qquad k = \pm1, \pm2, \cdots.$$

In fact, observing that

$$M[y_t\, y_{t+k}] = \lim_{n\to\infty} M[y_t^{(n,0)}\cdot y_{t-k}^{(n,0)}] \qquad k = 1, 2, \cdots,$$

and supposing that (17) is not true, we would have

$$\text{(18)} \qquad |\, M[y_t^{(\nu,0)}\cdot\bar{y}_{t-k}^{(\nu,0)}]\,| > a > 0,$$

as ν runs through some sequence $n_1, n_2, \cdots$, such that $n_i \to \infty$. The relation (18), however, would imply

$$\text{(19)} \qquad M[\,|\, y_t^{(\nu,0)} - c\, y_{t-k}^{(\nu,0)}\,|^2] \leq d^2\,(1 - \tfrac{1}{2}a^2)$$

for some sufficiently large ν and for some suitable c. Since $y_t^{(\nu,0)} - c\, y_{t-k}^{(\nu,0)}$ is a linear expression of the type appearing in the right hand member of (9), the relation (19) is incompatible with (15). Thus (18) is not possible and (17) must hold good.

Part of the above analysis is summed up in

THEOREM 1. *Given a time series $\{x_t\}$ which satisfies* (3), *let $\epsilon > 0$ be arbitrary. Then an integer n and a set of coefficients $a_i^{(n,0)}$ exist for which* (9) *defines a residual series $\{y_t^{(n,0)}\}$ such that*

$$M[y_t^{(n,0)}] = 0, \qquad |\, M[y_t^{(n,0)}\cdot\bar{y}_{t+k}^{(n,0)}]\,| < \epsilon \qquad k = \pm1, \pm2, \cdots.$$

4. A decomposition theorem. We shall first consider the special case where (15) gives

$$M[\,|\,y_t\,|^2] = d^2 = 0\,, \tag{20}$$

which is the same as

$$\underset{n\to\infty}{\text{l.i.m.}}\,(\cdots\, y_{t-1}^{(n,0)}, y_t^{(n,0)}, \cdots) = (\cdots, 0, 0, \cdots).$$

In this case we shall say that the sequence $\{x_t\}$ is deterministic,[3] the interpretation of this term being as follows: Given the sequence $\{x_t\}$ for all time points up to and including $t - 1$, we may, by the use of a finite number of the given values, predict x_{t+k} with any accuracy; i.e., with a residual error of arbitrarily small variance. This can be shown by induction. In fact, suppose that we are able to predict each of $x_t, \cdots, x_{t+k-1}$ in such a way that the prediction error has a variance $< \epsilon$, where ϵ is arbitrarily prescribed. Letting $\delta > 0$ be arbitrary, we can then find a formula of type (9) which predicts x_{t+k} in terms of the exact values $x_{t+k-1}, x_{t+k-2}, \cdots$ and which gives a residual variance $\delta/(k+1)$. Replacing here $x_{t+k-1}, \cdots, x_t$ by values so predicted that the residual variances are less than $\delta/(k+1)\,|\,a_1^{(n,0)}\,|, \cdots, \delta/(k+1)\,|\,a_k^{(n,0)}\,|$, it is seen that the total error of (9) will have a variance $< \delta$.

We proceed to the *general case*, $d^2 \geq 0$. According to the above analysis, y_t is that part of x_t which cannot be linearly predicted from the previous observations $x_{t-1}, x_{t-2}, \cdots$. In other words, each time point t brings in an unpredictable, random-like element y_t in the series $\{x_t\}$. Now while from (16) y_t is uncorrelated with the previous observations $x_{t-1}, x_{t-2}, \cdots$, it will in general be correlated with the future observations $x_{t+1}, x_{t+2}, \cdots$. Thus the unpredictable element y_t may be regarded as influencing the future development $x_{t+1}, x_{t+2}, \cdots$ of the series $\{x_t\}$. In order to examine this influence we proceed as follows.

We approximate x_t linearly in terms of $y_t, y_{t-1}, \cdots, y_{t-n}$, writing

$$x_t = b_0 y_t + b_1 y_{t-1} + \cdots + b_n y_{t-n} + u_t^{(n)} = z_t^{(n)} + u_t^{(n)}.$$

Determining the coefficients b_k by minimizing

$$M[\,|\,x_t - z_t^{(n)}\,|^2],$$

the coefficients b_k will thanks to (16)–(17) be independent of n. We obtain

$$b_0 = 1; \qquad b_k = M[x_t \cdot \bar{y}_{t-k}]/d^2, \qquad k = 1, 2, \cdots.$$

The sequence $\{z_t^{(n)}\}$ thus being determined for every n, it is further easily shown that $\{z_t^{(n)}\}$ converges in the mean, say to $\{z_t\}$,

$$\underset{n\to\infty}{\text{l.i.m.}}\,(\cdots, z_{t-1}^{(n)}, z_t^{(n)}, \cdots) = (\cdots, z_{t-1}, z_t, \cdots). \tag{21}$$

[3] The term is due to J. Doob [1]; in my study [9] I used the term *singular*.

We may thus write

$$z_t = y_t + b_1 y_{t-1} + b_2 y_{t-2} + \cdots ,$$

where the sum converges in the mean. Finally, we write

$$x_t = z_t + u_t , \tag{22}$$

which gives a decomposition of the series $\{x_t\}$ into two components $\{z_t\}$ and $\{u_t\}$.

In the decomposition (22) the component z_t is that part of x_t which is linearily built up by the unpredictable elements $\{y_t\}$ up to and including the time point t. From (17) we know that the sequence $\{y_t\}$ is non-autocorrelated. It can further be shown that the square modulus sum of the coefficients b_k is convergent,

$$\sum_{k=0}^{\infty} | b_k |^2 < \infty.$$

As to the component u_t, it can be shown that $\{u_t\}$ is deterministic. More precisely, we have

$$\underset{n\to\infty}{\text{l.i.m.}} \{u_t - (a_0^{(n,0)} + a_1^{(n,0)} u_{t-1} + \cdots + a_n^{(n,0)} u_{t-n})\} = \{0\}$$

where the $a_i^{(n,0)}$ are the same as the minimizing coefficients of (9). It can further be shown that u_t is uncorrelated with y_{t+k} and z_{t+k} for all k,

$$M[u_t \bar{y}_{t+k}] = M[u_t \bar{z}_{t+k}] = 0, \qquad (k = 0, \pm 1, \pm 2, \cdots).$$

Summing up the above results, we obtain

THEOREM 2. *Any time series $\{x_t\}$ which satisfies the conditions* (3) *allows the decomposition*

$$\{x_t\} = \{z_t + u_t\}, \tag{23}$$

with

$$\{z_t\} = \underset{n\to\infty}{\text{l.i.m.}} \{y_t + b_1 y_{t-1} + b_2 y_{t-2} + \cdots + b_n y_{t-n}\},$$

where the series $\{y_t\}$, $\{z_t\}$ and $\{u_t\}$ have the following properties.

A. *The elements y_t, z_t and u_t are obtained from x_t, x_{t-1}, $\cdots$ by the limit formulae* (14), (21) *and* (22).

B. *The series $\{y_t\}$ has zero mean,*

$$M[y_t] = 0,$$

is non-autocorrelated,

$$M[y_t \bar{y}_{t+k}] = 0, \qquad k = \pm 1, \pm 2, \cdots ,$$

and is uncorrelated with $\{x_{t-1}\}$, $\{x_{t-2}\}$, $\cdots$,

$$M[y_t \cdot \bar{x}_{t-k}] = 0, \qquad k = 1, 2, \cdots .$$

C. *The series* $\{u_t\}$ *is uncorrelated with* $\{y_t\}$ *and* $\{z_t\}$,

$$M[u_t\bar{y}_{t+k}] = M[u_t\bar{z}_{t+k}] = 0, \qquad (k = 0, \pm 1, \pm 2, \cdots).$$

D. *The series* $\{u_t\}$ *is deterministic.*

5. Application to the problem of prediction. In section 1 we have considered the problem of predicting x_{t+k} linearly in terms of $x_{t-1}, x_{t-2}, \cdots$. Now it is seen that theorem 2 gives the following formula for predicting x_{t+k} with an error of minimal variance,

$$\text{pred. } x_{t+k} = u_{t+k} + b_{k+1}y_{t-1} + b_{k+2}y_{t-2} + \cdots .$$

In fact, by theorem 2, A and D, the right-hand member can be calculated with any prescribed accuracy from a finite set of observations $x_{t-1}, x_{t-2}, \cdots, x_{t-N}$, where N of course depends on the accuracy desired; on the other hand, the prediction error being

$$y_{t+k} + b_1 y_{t+k-1} + \cdots + b_k y_t ,$$

we infer from theorem 2 (B) that this error is of minimal variance,

$$M[\,|\, x_{t+k} - \text{pred } x_{t+k} \,|^2] = (1 + |\, b_1 \,|^2 + \cdots + |\, b_k \,|^2)d^2.$$

6. Comments. As mentioned in section 2, the above theorem 2 is the analogue of a theorem on the decomposition of a stationary stochastic process given by the author previously (see [9], theorem 7). The starting point is then to apply formula (9), not as above to the same sequence $\{x_t\}$ for varying t, but to all realizations $\{x_t\}$ of the process, holding t fixed. The close connection between the decomposition in the two approaches is further brought out by the following theorem.

THEOREM 3. *Given a stochastic process,*

$$\cdots, \xi(t-1), \xi(t), \xi(t+1), \cdots,$$

which is stationary in the sense of (5), *let* $\{x_t\}$ *be an individual realization of this process. Then* $\{x_t\}$ *will with probability* 1 *allow the decomposition of theorem* 2.

In fact, according to the ergodic theorem of Birkhoff-Khintchine,[4] the averages (2) will exist with probability 1, and so theorem 3 follows from theorem 2. It should be observed that the coefficients b_k will in general vary from one realization to another.

The theory of the decomposition (23) has been carried further in a brilliant study by A. Kolmogoroff [3]. His analysis deals with the general case of a stationary sequence in a Hilbert space. Establishing a decomposition of type (23)

[4] See A. Kolmogoroff [2]. His proof refers to averages (2) of the special type where t_1 is hold fixed while $t_2 \to \infty$. According to the stationarity, however, the average exists, and is the same, when t_2 is fixed and $t_1 \to -\infty$, and so the general average (2) will likewise exist.

for such sequences Kolmogoroff also shows that the decomposition is uniquely determined by properties corresponding to A–D. Making use of the powerful methods of spectral analysis of linear transformations in Hilbert space, Kolmogoroff further presents a highly developed theory of the decomposition.

As immediate corollaries of this general theory Kolmogoroff [4] obtains corresponding results for a stationary stochastic process $\{\xi_t\}$ such as (4). Now thanks to our lemma in section 2, similar theorems hold good for the functional sequence (1). These results include detailed theorems on the connection between the decomposition (23) and, on the other hand, the function $F(\lambda)$ which by (6) generates the coefficients r_k. For example, it turns out that $\{x_t\}$ is completely deterministic if the derivative $F'(\lambda)$ is constant over an interval of positive measure. An explicit formula for the coefficients b_k in terms of the function $F(\lambda)$ may also be obtained. For proofs and further results, we must refer to Kolmogoroff's papers [3]–[4].

The theory of the decomposition (23) has later been generalized in various directions. V. Zasuhin [11] and J. Doob [1] have shown that the decomposition applies to multi-dimensional stationary sequences. As shown by the present author [10], the decomposition may be employed for the analysis of linear equation systems with an infinite number of unknowns. This device makes use of the decomposition of non-stationary sequences, a generalization indicated also by M. Loève [5].

REFERENCES

[1] J. L. Doob, "The elementary Gaussian processes," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 229–282.

[2] A. Kolmogoroff, "Ein vereinfachter Beweis des Birkhoff-Khintchineschen Ergodensatzes," *Rec. Math. (Sbornik) N.S.*, Vol. 2 (1937), pp. 367–368.

[3] A. Kolmogoroff, "Stationary sequences in Hilbert space (Russian)," *Bolletin Moskovskovo Gosudarstvenovo Universiteta, Matematika*, Vol. 2 (1941).

[4] A. Kolmogoroff, "Interpolation und Extrapolation von stationären zufälligen Folgen," *Bull. Acad. Sci. URSS, Sér. Math.*, Vol. 5 (1941), pp. 3–14.

[5] M. Loève, "Fonctions aléatoires de second ordre," *Revue Sci.*, Vol. 84 (1946), pp. 195–206.

[6] M. H. Stone, *Linear Transformations in Hilbert Space*, Amer. Math. Soc. Colloq. Publ. 15, New York, 1932.

[7] B. v. Sz. Nagy, *Spektraldarstellung linearer Transformationen des Hilbertschen Raumes*, Ergebn. Math. u. Grenzgeb., Vol. 5, Berlin, 1942.

[8] N. Wiener, "Generalized harmonic analysis," *Acta Math.*, Vol. 55 (1933), pp. 117–258.

[9] H. Wold, *A Study in the Analysis of Stationary Time Series*, Dissertation (Stockholm), Uppsala, 1938.

[10] H. Wold, "On infinite, non-negative definite, Hermitian matrices, and corresponding linear equation systems," *Arkiv Mat. Astr. Fys.*, Vol. 29A (1943).

[11] V. Zasuhin, "On the theory of multidimensional stationary random processes," *C.R. (Doklady) Akad. Sci. URSS, N.S.*, Vol. 33 (1941), pp. 435–437.