COLLOQUIUM MATHEMATICUM

VOL. XII 1964 FASC. 1

PREDICTION OF STRICTLY STATIONARY SEQUENCES

BY

K. URBANIK (WROCŁAW)

Let P be a probability measure defined on a σ-field $\mathscr{F}$ of subsets of a space Ω consisting of elementary events ω. Let $\mathfrak{S}(\Omega, \mathscr{F}, P)$ be the space of all random variables x defined on Ω, i. e. the space of all $\mathscr{F}$-measurable real-valued functions $x(\omega)$ defined on Ω. Throughout this paper we identify random variables which are equal P-almost everywhere. The space $\mathfrak{S}(\Omega, \mathscr{F}, P)$ is a linear space under usual addition and multiplication by real numbers. Moreover, it becomes a complete metric space under the Fréchet norm

$$\|x\| = \int_{\Omega} \frac{|x(\omega)|}{1+|x(\omega)|} P(d\omega).$$

It should be noted that this norm is non-homogeneous. It is clear that the convergence in Fréchet norm is equivalent to the convergence in probability P. The random variables which we consider in this paper are supposed to be defined on the same space Ω of elementary events.

A sequence $\{x_n\}$ $(n = 0, \pm 1, \pm 2, \ldots)$ of random variables is called *strictly stationary*, or — shortly — *stationary*, if for every system $m, n_1, n_2, \ldots, n_k$ of integers the multivariate distribution of the random variables $x_{n_1+m}, x_{n_2+m}, \ldots, x_{n_k+m}$ is independent of m. To each stationary sequence $\{x_n\}$ there corresponds a *shift transformation* $Tx_n = x_{n+1}$ $(n = 0, \pm 1, \pm 2, \ldots)$, which can be extended to an invertible isometry T in the space $\mathfrak{S}(\Omega, \mathscr{F}_0, P)$, where $\mathscr{F}_0$ is the smallest σ-field with respect to which all random variables x_n are measurable (see [2], Chapter X, § 1). Moreover, the isometry T is an extension of a P-measure-preserving set transformation. Consequently, it preserves the independence of random variables and constant random variables are invariant under the transformation T.

Given a sequence $\{y_n\}$ $(n = 0, \pm 1, \pm 2, \ldots)$, by $[y_n]$ and $[y_n: n \leqslant k]$ we shall denote the closed linear subspaces of $\mathfrak{S}(\Omega, \mathscr{F}, P)$ spanned by all random variables y_n and by random variables y_n with $n \leqslant k$

respectively. It is clear that the subspace $[x_n]$ generated by a stationary sequence $\{x_n\}$ is invariant under the shift transformation corresponding to $\{x_n\}$.

We say that a stationary sequence $\{x_n\}$ *admits a prediction*, if there exists a continuous linear operator A_0 from $[x_n]$ onto $[x_n: n \leqslant 0]$ such that

(i) $A_0 x = x$ whenever $x \in [x_n: n \leqslant 0]$,

(ii) if for every $y \in [x_n: n \leqslant 0]$ the random variables x and y are independent, then $A_0 x = 0$,

(iii) for every $x \in [x_n]$ and $y \in [x_n: n \leqslant 0]$ the random variables $x - A_0 x$ and y are independent.

The random variable $A_0 x$ can be regarded as a linear prediction of x based on the full past of the sequence $\{x_n\}$ up to the time $n = 0$. An optimality criterion is given by (iii). In what follows the operator A_0 will be called a *predictor* based on the past of the sequence $\{x_n\}$ up to time $n = 0$. The conditions (i), (ii) and (iii) determine the predictor A_0 uniquely. Indeed, if an operator A_0' satisfies these conditions, then for all $x \in [x_n]$ and $y \in [x_n: n \leqslant 0]$ the random variables $x - A_0' x$ and y are independent. Thus, by (ii), $A_0 x - A_0 A_0' x = 0$. Since $A_0' x \in [x_n: n \leqslant 0]$, we have, by (i), $A_0 A_0' x = A_0' x$, which together with the last equation implies $A_0 x = A_0' x$.

It should be noted that Gaussian stationary sequences with zero mean always admit a prediction. This follows from the fact that in this case the concepts of independence and orthogonality are equivalent and, moreover, the square-mean convergence and the convergence in probability are equivalent. Therefore the predictor A_0 is simply the best linear least squares predictor, i.e. the orthogonal projector from $[x_n]$ onto $[x_n: n \leqslant 0]$ (see [2], Chapter XII, § 1).

Since our stationary sequences need not have a finite variance, the problem of prediction discussed in this paper is not contained in the Wiener-Kolmogorov theory of the best linear least squares prediction for wide sense stationary sequences.

Let $\{x_n\}$ be a stationary sequence admitting a prediction. The predictor A_0 and the shift T induced by $\{x_n\}$ determine the predictor A_k based on the full past of $\{x_n\}$ up to the time $n = k$. Namely, setting

$$A_k = T^k A_0 T^{-k} \quad (k = 0, \pm 1, \pm 2, \ldots), \tag{1}$$

and taking into account that T preserves the independence, we obtain a continuous linear operator from $[x_n]$ onto $[x_n: n \leqslant k]$ satisfying the conditions

(2) $\quad A_k x = x$ whenever $x \in [x_n: n \leqslant k]$,

(3) if for every $y \in [x_n: n \leqslant k]$ the random variables x and y are independent, then $A_k x = 0$,

(4) for every $x \in [x_n]$ and $y \in [x_n: n \leqslant k]$ the random variables $x - A_k x$ and y are independent.

A stationary sequence $\{x_n\}$ admitting a prediction is called *deterministic*, if $A_0 x = x$ for every $x \in [x_n]$. Further, a stationary sequence x_n admitting a prediction is called *completely non-deterministic*, if $\lim_{k \to -\infty} A_k x = 0$ for every $x \in [x_n]$.

The aim of this paper is to prove that any stationary sequence admitting a prediction can be decomposed into a deterministic and a completely non-deterministic components. Moreover, we shall give a representation of completely non-deterministic sequences by moving averages. These theorems are an analogue of the well-known Wold's decomposition and representation theorems in the linear least squares prediction theory (see [2], Chapter XII and [4]).

It should be noted that in general, for a given $x \in [x_n]$, the prediction $A_k x$ does not furnish the best approximation of x in the Fréchet norm $\| \ \|$ by elements from the subspace $[x_n: n \leqslant k]$, i.e. in general $\inf\{\|x-y\|: y \in [x_n: n \leqslant k]\}$ is not equal to $\|x - A_k x\|$. But it will be shown that there exists an equivalent norm $\| \ \|_0$ in $[x_n]$ such that

$$\|x - A_k x\|_0 = \inf\{\|x-y\|_0: y \in [x_n: n \leqslant k]\}$$

for every $x \in [x_n]$ and $k = 0, \pm 1, \pm 2, \ldots$

We begin by proving some Lemmas from which we deduce the decomposition and the representation theorems.

LEMMA 1. *For $k \leqslant r$ the predictors satisfy the equation $A_k = A_k A_r = A_r A_k$.*

Proof. Let A_k and A_r $(k \leqslant r)$ be the predictors for a stationary sequence $\{x_n\}$. Since $A_k x \in [x_n: n \leqslant r]$ for every $x \in [x_n]$, we have, by (2), the relation $A_r A_k x = A_k x$, which implies $A_r A_k = A_k$. Further, by (4), for every $x \in [x_n]$ and $y \in [x_n: n \leqslant k]$ the random variables $x - A_r x$ and y are independent. Hence, by (3), $A_k x - A_k A_r x = 0$, which implies the equation $A_k = A_k A_r$.

LEMMA 2. *0 is the only constant random variable belonging to the subspace $[x_n]$ spanned by a stationary sequence $\{x_n\}$ admitting a prediction.*

Proof. Let c be a constant random variable belonging to $[x_n]$. For every positive number ε there exists a linear combination $\sum_{j=1}^{m} a_j x_{n_j}$ with real coefficients such that

$$\|c - \sum_{j=1}^{m} a_j x_{n_j}\| < \varepsilon.$$

Setting $q = \max(n_1, n_2, \ldots, n_m)$ and taking into account that c is invariant under the shift transformation T induced by the sequence $\{x_n\}$, we have the inequality

$$\left\|T^{-q}c - \sum_{j=1}^{m} a_j T^{-q} x_{n_j}\right\| = \left\|c - \sum_{j=1}^{m} a_j x_{n_j - q}\right\| < \varepsilon .$$

Since $\sum_{j=1}^{m} a_j x_{n_j-q} \in [x_n : n \leqslant 0]$ and ε was arbitrarily chosen, the relation $c \in [x_n \colon n \leqslant 0]$ is established. Thus, by (i), $A_0 c = c$. On the other hand, for any $y \in [x_n \colon n \leqslant 0]$ the random variables c and y are independent and, consequently, by (ii), $A_0 c = 0$. Thus $c = 0$, which completes the proof.

Let $\{y_k\}$ $(k = 1, 2, \ldots)$ be a sequence of random variables. If there are constants $a_1, a_2, \ldots$ such that $\sum_{k=1}^{\infty} (y_k - a_k)$ converges with probability 1, the series $\sum_{k=1}^{\infty} y_k$ will be said to *converge with probability* 1 *when centered* and $a_1, a_2, \ldots$ will be called *centering constants* ([2], Chapter III, § 2).

LEMMMA 3. *Let $\{y_k\}$ $(k = 1, 2, \ldots)$ be a sequence of independent random variables such that* 0 *is the only constant random variable belonging to $[y_k]$. If the series $\sum_{k=1}^{\infty} y_k$ converges with probability* 1 *when centered, then it converges with probability* 1, *regardless of the order of summation.*

Proof. By Theorem 2.6 in [2] (p. 112) we can find a sequence $a_1, a_2, \ldots$ of centering constants such that the series $\sum_{k=1}^{\infty} (y_k - a_k)$ is convergent with probability 1, regardless of the order of summation. Consequently, to prove the Lemma it suffices to prove that the numerical series $\sum_{k=1}^{\infty} a_k$ is absolutely convergent or, in other words, that $\sum_{k=1}^{\infty} a_k$ converges for any ordering of the terms. Further, since the conditions of the Lemma do not depend upon an ordering of terms y_k, it is sufficient to show that the series $\sum_{k=1}^{\infty} a_k$ is convergent. Contrary to this let us suppose that there are indices p_n and q_n such that $p_n \leqslant q_n$, $p_n \to \infty$ and the sequence $b_n = \sum_{k=p_n}^{q_n} a_n$ converges to a finite or infinite limit different from 0 as $n \to \infty$. From the equation

$$b_n^{-1} \sum_{k=p_n}^{q_n} y_k = b_n^{-1} \sum_{k=p_n}^{q_n} (y_k - a_k) + 1$$

it follows that $b_n^{-1}\sum_{k=p_n}^{q_n} y_k$ tends to 1 with probability 1 as $n\to\infty$. Consequently, $1\in[y_k]$, which contradicts the hypothesis. The Lemma is thus proved.

LEMMA 4. *Let A_k $(k=0,\pm1,\pm2,\ldots)$ be predictors for a stationary sequence $\{x_n\}$. There exists a continuous linear operator $A_{-\infty}$ on $[x_n]$ commuting with the shift induced by $\{x_n\}$ and such that for every $x\in[x_n]$*

$$\lim_{k\to\infty} A_{-k}x = A_{-\infty}x.$$

Proof. Given an element $x\in[x_n]$ we put

$$y_1 = x - A_{-1}x,\qquad y_j = A_{1-j}x - A_{-j}x\qquad (j=2,3,\ldots).$$

Since, by Lemma 1,

$$y_j = A_{1-j}x - A_{-j}A_{1-j}x\qquad (j=2,3,\ldots),$$

we infer that, according to (4), for $j=1,2,\ldots$ and $z\in[x_n\colon n\leqslant -j]$ the random variables y_j and z are independent. Moreover, we have the relation $y_j\in[x_n\colon n\leqslant 1-j]$ $(j=2,3,\ldots)$. Thus for every system $a_j, a_{j+1}, \ldots, a_{k+1}$ of real numbers the random variables a_jy_j and $a_{j+1}y_{j+1}+a_{j+2}y_{j+2}+\ldots+a_ky_k+a_{k+1}A_{-k}x$ are independent. Consequently,

$$E\exp\Big(i\sum_{r=j}^{k} a_ry_r + ia_{k+1}A_{-k}x\Big) = E\exp(ia_jy_j)\,E\exp\Big(i\sum_{r=j+1}^{k} a_ry_r + ia_{k+1}A_{-k}x\Big),$$

where E denotes the expectation. Hence we get the equation

$$E\exp\Big(i\sum_{r=1}^{k} a_ry_r + ia_{k+1}A_{-k}x\Big) = E\exp(ia_{k+1}A_{-k}x)\prod_{r=1}^{k} E\exp(ia_ry_r).$$

Thus the multivariate characteristic function of the random variables $y_1, y_2, \ldots, y_k, A_{-k}x$ is equal to the product of the characteristic functions of $y_1, y_2, \ldots, y_k$ and $A_{-k}x$ respectively. Hence it follows that the random variables $y_1, y_2, \ldots, y_k, A_{-k}x$ are independent. Since

$$x = \sum_{j=1}^{k} y_j + A_{-k}x\qquad (k=1,2,\ldots), \tag{5}$$

the series $\sum_{j=1}^{\infty} y_j$ converges with probability 1 when centered (see [2], Theorem 2.8, p. 119). Since, by Lemma 2, 0 is the only constant random variable belonging to $[x_n]$ and, consequently, to $[y_k]$, the series $\sum_{j=1}^{\infty} y_j$, according to Lemma 3, converges with probability 1. Hence and from (5) it follows that the limit

$$A_{-\infty}x = \lim_{k\to\infty} A_{-k}x$$

exists with probability 1. It is clear that the operator $A_{-\infty}$ defined by the last formula is linear. Moreover, by Banach theorem ([1], Theorem 4, p. 23) it is also continuous. Let T be the shift induced by the sequence $\{x_n\}$. From (1) we get the equation $A_{-k}T = TA_{-k-1}$, which implies $A_{-\infty}T = TA_{-\infty}$. The Lemma is thus proved.

We say that two sequences $\{x'_n\}$ and $\{x''_n\}$ of random variables are *independent*, if the random variables y' and y'' are independent whenever $y' \in [x'_n]$ and $y'' \in [x''_n]$.

THEOREM 1. *Each stationary sequence admitting a prediction is the sum of two independent stationary sequences admitting a prediction, one deterministic and the other completely non-deterministic. Moreover, if $x_n = x'_n + x''_n$ is such a decomposition, then $[x_n]$ is a direct sum of subspaces $[x'_n]$ and $[x''_n]$.*

Proof. Let $\{x_n\|$ be a stationary sequence admitting a prediction and let A_k $(k = 0, \pm 1, \pm 2, \ldots)$ be its predictors. The limit operator $A_{-\infty}$ defined by Lemma 4 satisfies, in view of Lemma 1, the equation

(6) $$A_k A_{-\infty} = A_{-\infty}A_k = A_{-\infty} \quad (k = 0, \pm 1, \pm 2, \ldots).$$

Hence, in particular, it follows that

(7) $$A^2_{-\infty} = A_{-\infty}$$

and, consequently,

(8) $$(I - A_{-\infty})^2 = I - A_{-\infty},$$

where I is the unit operator. Setting

(9) $$x'_n = A_{-\infty}x_n, \quad x''_n = (I - A_{-\infty})x_n \quad (n = 0, \pm 1, \pm 2, \ldots),$$

we have the relation

(10) $$x_n = x'_n + x''_n.$$

Moreover, by (7) and (8),

(11) $$[x'_n] = A_{-\infty}[x_n], \quad [x''_n] = (I - A_{-\infty})[x_n],$$

(12)

$$[x'_n : n \leqslant 0] = A_{-\infty}[x_n : n \leqslant 0], \quad [x''_n : n \leqslant 0] = (I - A_{-\infty})[x_n : n \leqslant 0],$$

and

(13) $$A_{-\infty}y' = y', \quad (I - A_{-\infty})y'' = y'' \text{ whenever } y' \in [x'_n] \text{ and } y'' \in [x''_n].$$

Since, by Lemma 4, the operator $A_{-\infty}$ commutes with the shift T induced by the sequence $\{x_n\}$, we infer that

$$T^n x_0' = T^n A_{-\infty} x_0 = A_{-\infty} T^n x_0 = A_{-\infty} x_n = x_n'$$

and, according to (10),

$$T^n x_0'' = T^n (x_0 - x_0') = x_n - x_n' = x_n''.$$

Thus both sequences $\{x_n'\}$ and $\{x_n''\}$ are stationary.

Let $y' \in [x_n']$ and $y'' \in [x_n'']$. By (2) and (4) for every integer k the random variables $A_k y'$ and $(I-A_k)y''$ are independent, whence the independence of $A_{-\infty}y'$ and $(I-A_{-\infty})y''$ follows. Hence and from (13) we obtain the independence of y' and y''. In other words, the sequences $\{x_n'\}$ and $\{x_n''\|$ are independent.

Now we shall prove that A_0 restricted to $[x_n']$ and $[x_n'']$ is a predictor of $\{x_n'\}$ and $\{x_n''\|$ respectively based on the past up to the time $n = 0$. First of all we note that, by (6), (11) and (12), the operator A_0 maps $[x_n']$ onto $[x_n'\colon n \leqslant 0]$ and $[x_n'']$ onto $[x_n''\colon n \leqslant 0]$. Consider the space $[x_n']$. By (6) and (13) we conclude that $A_0 = I$ on $[x_n']$. Thus the conditions (i) and (iii) are obvious. Since $[x_n'\colon n \leqslant 0] = [x_n']$, the only random variables z' such that z' and y' are independent for all $y' \in [x_n';\ n \leqslant 0]$ are constant ones. Thus, by Lemma 2, $z' = 0$, which shows that condition (ii) is also satisfied. Consequently, the sequence $\{x_n'\}$ is deterministic.

Now let us turn to the space $[x_n'']$. By (12) we have the inclusion $[x_n''\colon n \leqslant 0] \subset [x_n\colon n \leqslant 0]$. Hence it follows that the operator A_0 fulfils conditions (i) and (iii) on $[x_n'']$. To prove condition (ii) on $[x_n'']$ it suffices to show that the independence for all $y \in [x_n\colon n \leqslant 0]$ of random variables $(I-A_{-\infty})y$ and x'', where $x'' \in [x_n'']$, implies the independence of y and x''. But this implication is a direct consequence of the independence of sequences $\{x_n'\}$ and $\{x_n''\}$. Indeed, for every pair a_1, a_2 of real numbers the random variables $a_1 x'' + a_2(I-A_{-\infty})y$ and $a_2 A_{-\infty} y$ are independent. Moreover, the random variables $a_2(I-A_{-\infty})y$ and $a_2 A_{-\infty} y$ are also independent. Thus

$$E\exp(ia_1 x'' + ia_2 y) = E\exp\big(ia_1 x'' + ia_2(I-A_{-\infty})y\big) E\exp(ia_2 A_{-\infty} y)$$

and

$$E\exp(ia_1 x'') E\exp\big(ia_2(I-A_{-\infty})y\big) E\exp(ia_2 A_{-\infty} y)$$
$$= E\exp(ia_1 x'') E\exp(ia_2 y),$$

which implies the independence of x'' and y. Thus condition (ii) is also fulfilled. Finally, from (6) and (13) we obtain the relation

$$\lim_{k\to\infty} A_{-k} y'' = \lim_{k\to\infty} A_{-k}(I-A_{-\infty})y'' = \lim_{k\to\infty} (A_{-k}-A_{-\infty})y'' = 0$$

for all $y'' \epsilon [x_n'']$. Consequently, the sequence $\{x_n''\}$ is completely non-deterministic.

It remains to prove that $[x_n]$ is the direct sum of $[x_n']$ and $[x_n'']$. Since the sequence $\{x_n'\}$ and $\{x_n''\}$ are independent and 0 is the only constant random variable belonging to $[x_n]$ (see Lemma 2), we have the relation $[x_n'] \cap [x_n''] = \{0\}$. Further, from (10) it follows that the direct sum $[x_n'] \oplus [x_n'']$ contains the space $[x_n]$. On the other hand, by (11), $[x_n'] \oplus [x_n''] \subset [x_n]$, which implies $[x_n'] \oplus [x_n''] = [x_n]$. The Theorem is thus proved.

Before proving the representation theorem we shall prove two Lemmas concerning some properties of subspaces spanned by sequences of random variables.

LEMMA 5. *Let $\{v_n\}$ $(n = 0, \pm 1, \pm 2, \ldots)$ be a sequence of independent random variables such that 0 is the only constant random variable belonging to $[v_n]$. For every $x \epsilon [v_n]$ there exists then a sequence $\{a_n\}$ of real numbers such that*

$$x = \sum_{n=-\infty}^{\infty} a_n v_n,$$

where the series converges with probability 1, regardless of the order of summation.

Proof. Without loss of generality we may assume that

$$v_n \neq 0 \quad (n = 0, \pm 1, \pm 2 \ldots). \tag{14}$$

Given $x \epsilon [v_n]$, there exists a sequence of linear combinations $\sum_{n=-k}^{k} a_n^{(k)} v_n$ tending to x in probability as $k \to \infty$. Let $\varphi(t)$, $\varphi_r(t)$ and $\psi_{rk}(t)$ be the characteristic functions of the random variables x, v_r and $\sum_{n=-k}^{k} a_n^{(k)} v_n - a_r^{(k)} v_r$ respectively. Suppose that there exist an index r and a subsequence $k_1, k_2, \ldots$ tending to ∞ such that

$$\lim_{s\to\infty} |a_r^{(k_s)}| = \infty.$$

Then the sequence of random variables

$$\frac{1}{a_r^{(k_s)}} \sum_{n=-k_s}^{k_s} a_n^{(k_s)} v_n$$

tends to 0 in probability as $s \to \infty$, which in the language of characteristic functions can be written as follows:

$$\lim_{s\to\infty} \varphi_r(t) \psi_{rk_s}\left(\frac{t}{a_r^{(k_s)}}\right) = 1.$$

Hence it follows that $|\varphi_r(t)| = 1$ for all t or, in other words, that v_r is a constant random variable. But this contradicts the hypothesis and (14). Thus for every index r the coefficients $a_r^{(k)}$ $(k = r, r+1, \ldots)$ are bounded in common. Consequently, passing to a subsequence if necessary, we may assume that for all indices r the limits $\lim_{k\to\infty} a_r^{(k)} = a_r$ exist. Hence it follows that for every positive integer m the sequence of random variables

$$\sum_{n=-m}^{m} a_n v_n + \sum_{m<|n|\leqslant k} a_n^{(k)} v_n$$

tends to x in probability as $k \to \infty$. Thus

$$\lim_{k\to\infty} \prod_{n=-m}^{m} \varphi_n(a_n t) \prod_{m<|n|\leqslant k} \varphi_n(a_n^{(k)} t) = \varphi(t)$$

and, consequently, for any positive integer m

$$\prod_{n=-m}^{m} \left|\varphi_n(a_n t)\right| \geqslant |\varphi(t)|.$$

Hence it follows that the infinite product $\prod_{n=-\infty}^{\infty} |\varphi_n(a_n t)|$ converges on a set of positive Lebesgue measure. This implies that the series $\sum_{n=-\infty}^{\infty} a_n v_n$ converges with probability 1 when centered (see [2], Theorem 2.7, p. 115). Applying Lemma 3 we conclude that the series $\sum_{n=-\infty}^{\infty} a_n v_n$ converges with probability 1, regardless of the order of summation. Setting

$$y = x - \sum_{n=-\infty}^{\infty} a_n v_n, \tag{15}$$

for every positive integer m we have the convergence of

$$\sum_{m<|n|\leqslant k} a_n^{(k)} v_n - \sum_{m<|n|} a_n v_n$$

to y in probability as $k \to \infty$. Thus $y \in [v_n : |n| > m]$ for every m, and, consequently, the random variable y is measurable on the sample space of v_n $(|n| > m)$, which, by zero-one law (see [2], Theorem 1.1, p. 102) implies that y is a constant random variable. Since 0 is the only constant random variable in $[v_n]$, we have $y = 0$, which, by (15), completes the proof of the Lemma.

LEMMA 6. *Suppose that for every $n = 1, 2, \ldots$ the random variables x_n and y_n are independent. If 0 is the only constant random variable belonging to $[x_n]$, then the convergence $x_n + y_n \to 0$ in probability implies the convergence $x_n \to 0$ in probability.*

Proof. From the relation $x_n + y_n \to 0$ in probability and from the independence of x_n and y_n it follows that the absolute values of characteristic functions of x_n tend to 1 as $n \to \infty$. Consequently, there exists a sequence $c_1, c_2, \ldots$ of constants such that

$$x_n - c_n \to 0 \tag{16}$$

in probability (see [3], Theorem 3, p. 57). If the sequence $\{c_n\}$ contains a subsequence $\{c_{n_k}\}$ tending to a finite or infinite limit different from 0, then by (16)

$$c_{n_k}^{-1} x_{n_k} - 1 \to 0$$

in probability as $k \to \infty$. But this would imply $1 \in [x_n]$, which yields a contradiction. Consequently, $c_n \to 0$, which, by (16), completes the proof.

Now we shall prove a representation theorem for completely non-deterministic sequences.

THEOREM 2. *Let $\{x_n\}$ be a stationary completely non-deterministic sequence. Then there exists a sequence $\{v_n\}$ of independent identically distributed random variables such that $[v_n\colon n \leqslant 0] = [x_n\colon n \leqslant 0]$ and x_n is a moving average*

$$x_n = \sum_{k=-\infty}^{0} a_k v_{k+n} \quad (n = 0, \pm 1, \pm 2, \ldots), \tag{17}$$

where the series converges with probability 1, regardless of the order of summation.

Conversely, if $\{v_n\}$ is a sequence of independent identically distributed random variables such that 0 is the only constant random variable in $[v_n]$, then the moving average (17) is a stationary completely non-deterministic provided $[x_n\colon n \leqslant 0] = [v_n\colon n \leqslant 0]$.

Proof. Let $\{x_n\}$ be a stationary completely non-deterministic sequence and A_k $(k = 0, \pm 1, \pm 2, \ldots)$ its predictors. Put

$$v_k = x_k - A_{k-1} x_k \quad (k = 0, \pm 1, \pm 2, \ldots). \tag{18}$$

Denoting by T the shift transformation induced by $\{x_n\}$ we obtain, by (1), the equation

$$T^k v_0 = T^k x_0 - T^k A_{-1} x_0 = x_k - T^k A_{-1} T^{-k} x_k = x_k - A_{k-1} x_k = v_k, \tag{19}$$

which shows that the random variables v_k are identically distributed. Moreover,

$$v_k \in [x_n\colon n \leqslant k] \quad (k = 0, \pm 1, \pm 2, \ldots) \tag{20}$$

and, by (4), for every $y \in [x_n\colon n \leqslant k-1]$ the random variables v_k and y are independent. Thus for every system $\beta_k, \beta_{k-1}, \ldots, \beta_{k-r}$ of real numbers the random variables $\beta_k v_k$ and $\sum_{j=1}^{r} \beta_{k-j} v_{k-j}$ are independent. Consequently,

$$E \exp\Big(i \sum_{j=0}^{r} \beta_{k-j} v_{k-j}\Big) = E \exp(i\beta_k v_k) E \exp\Big(i \sum_{j=1}^{r} \beta_{k-j} v_{k-j}\Big).$$

Hence, by induction, we obtain the formula for characteristic functions

$$E \exp\Big(i \sum_{j=0}^{k} \beta_{k-j} v_{k-j}\Big) = \prod_{j=0}^{r} E \exp(i\beta_{k-j} v_{k-j}),$$

which implies the independence of $v_k, v_{k-1}, \ldots, v_{k-r}$. Thus the sequence $\{v_k\}$ consists of independent random variables.

Setting

$$w_k = A_k x_0 - A_{k-1} x_0 \quad (k = -1, -2, \ldots),$$

we have the formula

$$(21) \qquad x_0 = x_0 - A_{-1} x_0 + \sum_{k=1-m}^{-1} w_k + A_{-m} x_0 \quad (m = 2, 3, \ldots).$$

Moreover,

$$(22) \qquad w_k \in [x_n \colon n \leqslant k] \quad (k = -1, -2, \ldots).$$

Since, by Lemma 1, $A_{k-1} A_k = A_{k-1}$, the element w_k can be rewritten in the form

$$w_k = A_k x_0 - A_{k-1} A_k x_0,$$

which, by (4), shows that for every $y \in [x_n \colon n \leqslant k-1]$ the random variables w_k and y are independent. Hence and from (3) it follows that

$$(23) \qquad A_{k-1} w_k = 0 \quad (k = -1, -2, \ldots).$$

Further, by (22), there exists a sequence of linear combinations $\sum_{j=n}^{k} a_j^{(n)} x_j$ $(n = k, k-1, \ldots)$ tending to w_k in probability as $n \to -\infty$. Replacing, by (18), x_k by $v_k + A_{k-1} x_k$ and denoting the expression

$$\sum_{j=n}^{k-1} a_j^{(n)} x_j + a_k^{(n)} A_{k-1} x_k$$

briefly by z_{nk}, we get the convergence

$$(24) \qquad a_k^{(n)} v_k + z_{nk} \to w_k$$

in probability as $n \to -\infty$. Since z_{nk} belongs to the subspace $[x_n : n \leqslant k-1]$, we have, by (2) and (18),

$$A_{k-1}(a_k^{(n)} v_k + z_{nk}) = z_{nk},$$

which, by (23) and (24), implies that $z_{nk} \to 0$ in probability as $n \to -\infty$. Thus, according to (24), $a_k^{(n)} v_k \to w_k$ in probability as $n \to -\infty$. Consequently, there exists a constant a_k such that $w_k = a_k v_k$ $(k = -1, -2, \ldots)$ Setting in addition $a_0 = 1$, we obtain from (21) the equation

$$x_0 = \sum_{k=1-m}^{0} a_k v_k + A_{-m} x_0 \quad (m = 2, 3, \ldots).$$

Since the sequence $\{x_n\}$ is completely non-deterministic, $A_{-m} x_0$ tends to 0 in probability as $m \to \infty$. Consequently, the last equation yields

$$x_0 = \sum_{k=-\infty}^{0} a_k v_k,$$

where, according to Lemma 5, the series converges with probability 1 regardless of the order of summation. Hence and from (19) formula (17) follows. Consequently, $[x_n\colon n \leqslant 0] \subset [v_n\colon n \leqslant 0]$, which together with (20) implies the identity $[x_n\colon n \leqslant 0] = [v_n\colon n \leqslant 0]$. The first part of the Theorem is thus proved.

Suppose now that $\{v_n\}$ is a sequence of independent identically distributed random variables such that 0 is the only constant random variable belonging to $[v_n]$. Let T be the shift transformation defined by means of the formula $Tv_n = v_{n+1}$ $(n = 0, \pm 1, \pm 2, \ldots)$. Further, let $\{x_n\}$ be a sequence of moving averages (17) satisfying the condition $[x_n\colon n \leqslant 0] = [v_n\colon n \leqslant 0]$. Of course, $Tx_n = x_{n+1}$, which shows that the sequence $\{x_n\}$ is stationary. Moreover, $[x_n] = [v_n]$. Thus, by Lemma 5, each element x of $[x_n]$ can be represented by a series

$$x = \sum_{k=-\infty}^{\infty} \beta_k v_k \tag{25}$$

which converges with probability 1 regardless of the order of summation. It should be noted that this representation is unique except a trivial case $v_n = 0$ $(n = 0, \pm 1, \pm 2, \ldots)$. Since in this trivial case the sequence $\{x_n\}$ is obviously completely non-deterministic, we shall assume in the sequel that $v_n \neq 0$.

For elements x having the expansion (25) we put

$$A_0 x = \sum_{k=-\infty}^{0} \beta_k v_k. \tag{26}$$

We note that by Theorem 2.6 in [2] (p. 112) and Lemma 5 this series converges with probability 1 regardless of the order of summation. We shall prove that A_0 is the predictor for $\{x_n\}$ based on the full past up to time $n = 0$. First of all we note that the operator A_0 is linear and transforms $[x_n]$ onto $[x_n: n \leqslant 0]$. Its continuity is a direct consequence of Lemma 6 because of independence of the random variables $A_0 x$ and $x - A_0 x$. Further, the conditions (i) and (iii) are obvious. In order to prove (ii) suppose that $x \in [x_n]$ and for every $y \in [x_n: n \leqslant 0]$ the random variables x and y are independent. Hence, in particular, it follows that the random variables x and $A_0 x$ are independent. Since $x - A_0 x$ and $A_0 x$ are also independent and $x = (x - A_0 x) + A_0 x$, we infer, by a simple reasoning, that $A_0 x$ is a constant random variable. But 0 is the only constant random variable belonging to $[x_n]$, which implies $A_0 x = 0$. Thus condition (ii) is also fulfilled and, consequently, A_0 is the predictor for $\{x_n\}$ based on the past up to time $n = 0$. Finally, for x given by (25) we obtain, in view of (1) and (26), the equation

$$A_{-k} x = T^{-k} A_0 T^k x = \sum_{n=-\infty}^{-k} \beta_n v_n,$$

which implies $\lim_{k\to\infty} A_{-k} x = 0$. Thus the sequence $\{x_n\}$ is completely non-deterministic, which completes the proof of the Theorem.

THEOREM 3. *Let $\{x_n\}$ be a stationary sequence admitting a prediction. Then there exists a norm $\| \ \|_0$ on $[x_n]$ invariant under the shift transformation induced by $\{x_n\}$ and such that the convergence in the norm $\| \ \|_0$ is equivalent to the convergence in probability. Moreover, for every $x \in [x_n]$ and for every predictor A_k $(k = 0, \pm 1, \pm 2, \ldots)$ the formula*

$$\|x - A_k x\|_0 = \inf\{\|x - y\|_0 : y \in [x_n: n \leqslant k]\} \tag{27}$$

holds.

Proof. By Theorem 1 the sequence $\{x_n\}$ is the sum of two independent components $\{x'_n\}$ and $\{x''_n\}$, where $\{x'_n\}$ is a deterministic stationary sequence and $\{x''_n\}$ completely non-deterministic stationary sequence. Moreover, $[x_n] = [x'_n] \oplus [x''_n]$. Thus each element x belonging to $[x_n]$ has a unique representation $x = x' + x''$, where $x' \in [x'_n]$ and $x'' \in [x''_n]$. Further, by Theorem 2, there exists a sequence $\{v_n\}$ of independent identically distributed random variables belonging to $[x''_n]$ such that $[x''_n]$ consists of all series $\sum_{n=-\infty}^{\infty} \beta_n v_n$ which, by Lemma 5, converge with probability 1 regardless of the order of summation. Moreover, for every $x \in [x_n]$ we have the formula

$$A_0 x = x' + \sum_{n=-\infty}^{0} \beta_n v_n, \tag{28}$$

where

$$x = x' + \sum_{n=-\infty}^{\infty} \beta_n v_n \quad (x' \in [x'_n]). \tag{29}$$

Put

$$\|x\|_0 = \|x'\| + \sup_N \Big\| \sum_{n\in N} \beta_n v_n \Big\|, \tag{30}$$

where x is represented by (29), $\| \ \|$ denotes the Fréchet norm and the supremum is extended over all subsets N of integers. We note that all series $\sum_{n\in N} \beta_n v_n$ converge with probability 1 regardless of the order of summation (see [2], Corollary 1, p. 118) and, of course, their sums are independent of the order of summation. It is clear that $\| \ \|_0$ is a norm on $[x_n]$ and $\|x\|_0 \geqslant \|x\|$ for all $x \in [x_n]$. Thus the convergence in the norm $\| \ \|_0$ implies the convergence in probability. Now we shall prove the converse implication. Let us assume the contrary, that is, we assume the existence of a sequence $\{y_k\}$ in $[x_n]$ which tends to 0 in probability and $\|y_k\|_0 > c$ $(k = 1, 2, \ldots)$, where c is a positive constant. Representing y_k in the form

$$y_k = y_k' + \sum_{n=-\infty}^{\infty} \gamma_n^{(k)} v_n,$$

where $y_k' \in [x_n']$ and taking into account the formula $y_k' = A_{-\infty} y_k$, we have, by continuity of the operator $A_{-\infty}$, $\lim_{k\to\infty} \|y_k'\| = 0$. Thus, by (30), there exists a sequence $N_1, N_2, \ldots$ of subsets of integers such that

$$\lim_{k\to\infty} \Big\| \sum_{n\in N_k} \gamma_n^{(k)} v_n \Big\| > 0. \tag{31}$$

The random variables $\sum_{n\in N_k} \gamma_n^{(k)} v_n$ and $y_k - \sum_{n\in N_k} \gamma_n^{(k)} v_n$ are independent and their sum, being equal to y_k, tends to 0 in probability as $k\to\infty$. Consequently, by Lemma 6, $\sum_{n\in N_k} \gamma_n^{(k)} v_n$ tends to 0 in probability as $k\to\infty$, which contradicts (31). Thus the convergence in probability implies the convergence in the norm $\| \ \|_0$.

Since both subspaces, $[x_n']$ and $[x_n'']$, and the Fréchet norm are invariant under the shift transformation induced by the sequence $\{x_n\}$, the norm $\| \ \|_0$ is also invariant. Hence and from (1) it follows that to prove (27) for all k it suffices to prove it for $k = 0$.

From (28), (29) and (30) we obtain the formula

$$\|x - A_0 x\|_0 = \sup_N{}^{+} \Big\| \sum_{n\in N} \beta_n v_n \Big\|, \tag{32}$$

where $\sup^+$ denotes the supremum extended over all subsets N of positive integers. Further, each element y from $[x_n\colon n \leqslant 0]$ is of the form

$$y = y' + \sum_{n=-\infty}^{0} \delta_n v_n,$$

where $y' \in [x'_n]$. Thus, by (30) and (32),

$$\|x-y\|_0 \geqslant \|x'-y'\| + \sup_N{}^+ \Big\| \sum_{n\in N} \beta_n v_n \Big\| \geqslant \|x - A_0 x\|_0,$$

which implies the equation

$$\|x - A_0 x\|_0 = \sup\{\|x-y\|_0 \colon y \in [x_n \colon n \leqslant 0]\}.$$

The Theorem is thus established.

REFERENCES

[1] S. Banach, *Théorie des opérations linéaires*, New York 1955.

[2] J. L. Doob, *Stochastic processes*, New York-London 1953.

[3] B. V. Gnedenko and A. N. Kolmogorov, *Limit distributions for sums of independent random variables*, Cambridge 1954.

MATHEMATICAL INSTITUTE OF THE POLISH ACADEMY OF SCIENCES
INSTITUTE OF MATHEMATICS, WROCŁAW UNIVERSITY

Reçu par la Rédaction le 8. 4. 1963