MEMOIRS OF THE COLLEGE OF SCIENCE, UNIVERSITY OF KYOTO, SERIES A
Vol. XXXIII, Mathematics No. 2, 1960.

The prediction theory of stationary random distributions

By

K. BALAGANGADHARAN

(Received August 20, 1960)

1. Introduction

The main results of the present paper are a theorem on the existence of a 'backward moving average representation' for purely non-deterministic stationary random distributions (Theorem 5.1), and another, giving a spectral criterion ensuring the existence of such a representation (Theorem 6.2). We follow the method of Hanner [2] for establishing the former; the tool for proving the latter turns out to be a slight generalization (Lemma 6.1) of a well-known theorem of Paley and Wiener [4] on the Fourier transforms of functions vanishing on a half-line.

At a time when work on this paper was substantially complete, it was brought to the author's notice that the prediction problem for stationary random distributions has been solved by G. Maruyama, and earlier still, presumably, by Rozanov [5]. Maruyama's results have not yet appeared in print but were presented recently to a seminar at Kyushu University; his technique consists essentially in reducing a random distribution to a stochastic process by smoothing. A similar idea has been exploited by Rozanov. The corresponding problem for stationary n^{th} order increments has been solved earlier by Yaglom [8].

The author is heavily indebted to Professor K. Itô for generous help during the preparation of this paper, in particular for the proof of the crucial Lemma 4.4., and to Mr. N. Ikeda for a critical discussion. Thanks are due to Professor K. Chandrasekharan for kind encouragement all through.

2. Generalities

We summarize, for ready reference, a few facts relating to stationary random distributions; details may be found in [3]. We employ the terminology and notations $\mathfrak{D}$, $\mathfrak{D}'$, $\mathcal{S}$, $\mathcal{S}'$, $*$, $\vee$, $\sim$ etc. as standardized in [6].

A *random distribution* X is a continuous linear functional on the space $\mathfrak{D}\equiv\mathfrak{D}(R^1)$ [6] with values in a Hilbert space $L^2(\Omega)$, $\Omega\equiv(\Omega, \mathfrak{B}, P)$ a probability space. We suppose throughout that $EX(\varphi)\equiv 0$, $\varphi\in\mathfrak{D}$, where E denotes the expectation. We consider only the case in which X is (*weakly*) *stationary*, i.e. for every pair φ, $\psi\in\mathfrak{D}$,

$$(\tau_h X(\varphi),\ \tau_h X(\psi)) = (X(\varphi),\ X(\psi))\,;$$

here $(\cdot\,,\,\cdot)$ denotes the inner product in $L^2(\Omega)$ amd τ_h is the translation operator on X defined by

$$(\tau_h X)(\varphi) = X(\tau_{-h}\varphi), \quad (\tau_h\varphi)(x) = \varphi(x+h), \quad \varphi\in\mathfrak{D},\ h \text{ real.} \tag{2.1}$$

The *covariance distribution* ρ of X, defined by

$$(X(\varphi),\ X(\psi)) = \rho(\varphi*\tilde{\psi})\,,$$

can be represented by a formula of Khintchine type:

$$\rho(\varphi) = \int_{-\infty}^{\infty}(\mathfrak{F}\varphi)(\lambda)\,d\mu(\lambda)\,, \tag{2.2}$$

where $d\mu$, the *spectral measure* of X, is a non-negative measure of slow growth on the real Borel sets. In analogy with a standard formula in the theory of stationary processes, we have here

$$X(\varphi) = \int_{-\infty}^{\infty}(\mathfrak{F}\varphi)(\lambda)\,dM(\lambda)\,, \tag{2.3}$$

where M is a random measure with $||dM(\lambda)||^2=\mu(d\lambda)$.

Formula (2.2), in particular, will be of some consequence to us. Notice, in the first place, that it leads to the relation

$$(X(\varphi),\ X(\psi)) = \int_{-\infty}^{\infty}\mathfrak{F}(\varphi*\tilde{\psi})(\lambda)\,d\mu(\lambda) = \int_{-\infty}^{\infty}(\mathfrak{F}\varphi)(\lambda)\overline{(\mathfrak{F}\psi)(\lambda)}\,d\mu(\lambda)\,,$$

which shows that the correspondence $X(\varphi)\rightarrow\mathfrak{F}\varphi$ establishes an isometry between a dense subset in $L^2(X)$, the subspace spanned in $L^2(\Omega)$ by $\{X(\varphi),\ \varphi\in\mathfrak{D}\}$, and one in $L^2(-\infty,\ \infty\,;\ d\mu)$, and as such can be extended into an isomorphism between $L^2(X)$ and

$L^2(d\mu) \equiv L^2(-\infty, \infty; d\mu)$. It follows that $L^2(X)$ is a *separable* Hilbert space.

A further consequence of (2. 2) is the following basic lemma, presumed new, which will be of crucial significance to us in § 5.

LEMMA 2. 1. If $d\mu$ is absolutely continuous, then $L^2(X)$ is generated by translates of $X(\varphi)$ for a single fixed $\varphi \in \mathfrak{D}$.

More explicitly, if $d\mu$ is absolutely continuous (relative to Lebesgue measure), then for arbitrary $\psi \in \mathfrak{D}$ and a given fixed $\varphi \in \mathfrak{D}$, we can find a finite number n of real constants h_ν and complex constants c_ν such that

$$||X(\psi) - \sum_{\nu=1}^{n} c_\nu X(\tau_{h_\nu} \varphi)||$$

is arbitrarily small. And by the isomorphism $X(\varphi) \to \mathfrak{F}\varphi$, this is equivalent to the statement that for arbitrary $\varepsilon > 0$, φ, $\psi \in \mathfrak{D}$, there exist suitable constants n, c_ν, h_ν such that

$$\int_{-\infty}^{\infty} |(\mathfrak{F}\psi)(x) - (\mathfrak{F}\varphi)(x) \sum_{\nu=1}^{n} c_\nu e^{2\pi i h_\nu x}|^2 d\mu(x) < \varepsilon .$$

This fact turns out to be an extension of the well-known L^2-closure theorem of Wiener [7] and can be proved by an appropriate modification of Wiener's argument. We content ourselves with the remark that each $\mathfrak{F}\varphi$ is a rapidly decreasing function which can be extended to the whole complex plane as an entire function of exponential type; hence $\mathfrak{F}\varphi$ has at most a countable number of (real) zeros which form a Lebesgue null set and by assumption of the lemma, a $d\mu$-null set. The facts that $\mathfrak{F}\varphi$ is rapidly decreasing and that $d\mu$ is absolutely continuous enable us to push Wiener's argument through.

3. The Wold decomposition

The prediction problem for X consists in the following: denote by $L^2(X; a)$ the closed linear subspace of $L^2(X)$ generated by $\{X(\varphi), \varphi \in \mathfrak{D}_a\}$ where $\mathfrak{D}_a = \{\varphi : \varphi \in \mathfrak{D}, \operatorname{supp} \varphi \subset (-\infty, a]\}$; given $L^2(X; a)$ we ask for the best estimate, in the sense of minimum mean square error, of $X(\varphi_0)$ for given $\varphi_0 \in \mathfrak{D}$. Obviously there exists a unique *linear predictor* $\tilde{X}(\varphi_0) \equiv (X(\varphi_0)/L^2(X; a))$; the symbol $(\xi/\mathfrak{M})$ denotes the projection of the element ξ on the subspace $\mathfrak{M}$.

Let us observe first that the shift-transformations τ_h induce a one-parameter abelian group of unitary operators T_h on $L^2(X)$, the definition of T_h being the natural one:

$$T_h(\sum_{\nu=1}^{n} c_\nu X(\varphi_\nu)) = \sum_{\nu=1}^{n} c_\nu X(\tau_{-h}\varphi_\nu), \quad c_\nu \in C, \ \varphi_\nu \in \mathfrak{D}, \qquad n = 1, 2, \cdots .$$

We then have

$$T_h L^2(X ; a) = L^2(X ; a+h) . \tag{3.1}$$

The $L^2(X ; a)$'s form a non-decreasing family of subspaces:

$$L^2(X ; a) \subset L^2(X ; b), \quad a < b . \tag{3.2}$$

Two extreme cases may be envisaged in this situation, leading to the following

DEFINITION. If $\bigcap_a L^2(X ; a) = L^2(X)$, then X is called *deterministic*; if $\bigcap_a L^2(X;a) = \{0\}$, then X is called *purely non-deterministic.*

It is easily seen that the former alternative is equivalent to $L^2(X ; a) = L^2(X)$ for each a; the latter to the fact that as $a \to -\infty$ $\|(X(\varphi)/L^2(X ; a)\| \to 0$ for each $\varphi \in \mathfrak{D}$.

The first step in prediction is to obtain the Wold decomposition of X embodied in the following

THEOREM 2.1. A stationary random distribution X can be expressed uniquely in the form $X = Y + Z$, where Y and Z are stationary random distributions, mutually orthogonal in the sense that $Y(\varphi) \perp Z(\varphi)$ for every $\varphi \in \mathfrak{D}$. Besides, one of these is deterministic while the other is purely non-deterministic.

We omit the proof which is the same as that of the corresponding theorem for processes [2, Theorem 1].

Since the deterministic component always survives as it is on projection to any $L^2(X ; a)$, we shall hereafter *assume that X is purely non-deterministic.* We shall show in the next section that this assumption entails the absolute continuity of the spectral measure.

4. The absolute continuity of the spectral measure

For every pair of real numbers a, b, $a < b$, we consider the space $L^2(X ; a, b) = L^2(X ; b) \ominus L^2(X ; a)$, $\ominus$ denoting orthogonal complementation. For any $\varphi \in \mathfrak{D}_b$, $(X(\varphi)/L^2(X ; a, b))$ is orthogonal to $L^2(X ; a)$ and may be interpreted as the part of $X(\varphi)$ which is not determined by the distribution up to time a. Since we are in the purely non-deterministic case, each $L^2(X;a,b)$ contains elements of positive norm. We have evidently

$$T_h L^2(X;a,b) = L^2(X;a+h,\ b+h). \tag{4.1}$$

Following Hanner [2, § 3] we proceed to construct a random measure B, i.e. an $L^2(X)$-valued set function $B(\cdot)$ on the ring of Borel sets in R^1 with finite Lebesgue measure m, having the property $(B(S_1),\ B(S_2)) = m(S_1 \cap S_2)$. We choose a fixed positive number u and an arbitrary element $z \in L^2(X;0,u)$ and set for each interval $(a,b]$

$$B((a,b]) = \left(\int_A^B T_h z\, dh / L^2(X;a,b)\right), \tag{4.2}$$

$A \leqslant a-u$, $B \geqslant b$. The integral exists as a Riemann-Stieltjes integral, *a la* Cramér, and is independent of A and B, and the definition along with (4.1) leads to the following properties: if $a<b<c$, then

$$\text{(i)}\quad B((a,b]) + B((b,c]) = B((a,c]); \tag{4.3}$$

$$\text{(ii)}\quad B((a,b]) \perp B((b,c]); \tag{4.4}$$

$$\text{(iii)}\quad T_h B((a,b]) = B((a+h,\ b+h]). \tag{4.5}$$

By a routine procedure then B is extended to all Borel sets and we observe that $\|B(\cdot)\|^2$ defines a measure on linear Borel sets which is translation-invariant and hence

$$\|B((a,b])\|^2 = K(b-a),\quad K = K(z) \geqslant 0.$$

Now a minor modification of Hanner's argument [2, Prop. C] shows that there exists at least one $z \in L^2(X;0,u)$ such that $K>0$. We then normalize B by taking $K=1$ and call it the Brownian random measure.

With respect to B we define in a standard way a stochastic integral

$$\zeta = \int_{-\infty}^{\infty} g(u)\, dB(u)$$

for functions $g \in L^2(-\infty,\infty)$ and observe that if g, h belong to $L^2(-\infty,\infty)$, then

$$\left(\int_{-\infty}^{\infty} g(u)\, dB(u),\ \int_{-\infty}^{\infty} h(u)\, dB(u)\right) = \int_{-\infty}^{\infty} g(u)\overline{h(u)}\, du. \tag{4.6}$$

Thus, if $L^2(B)$ denotes the closed linear subspace in $L^2(X)$ generated by $\{B((a, b]), -\infty < a < b < \infty\}$, then the spaces $L^2(B)$ and $L^2(-\infty, \infty)$ are isomorphic. We wish to show eventually that $L^2(B) = L^2(X)$.

If we now write

$$X_1(\varphi) = (X(\varphi)/L^2(B)), \quad \varphi \in \mathfrak{D}, \tag{4.7}$$

then X_1 is itself a stationary random distribution. For each real t,

$$L^2(B) = L^2(B((a, b]), \quad a < b \leqslant t) \oplus L_2(B((a, b]), \quad t \leqslant a < b).$$

If $t(\varphi) = \sup\{s : \varphi(s) \neq 0, \ \varphi \in \mathfrak{D}\}$, then by the definition of B we have

$$X(\varphi) \perp L^2((B(a, b]) ; t(\varphi) \leqslant a < b),$$
$$B((a, b]) \subset L^2(X ; a, b) \subset L^2(X ; b).$$

Hence

$$X_1(\varphi) \in L^2((B(a, b]), \quad a < b \leqslant t(\varphi)) \subset L^2(\mathrm{X} ; t(\varphi)). \tag{4.8}$$

From this it can be seen that X_1 is purely non-deterministic. On the other hand $X_1(\varphi) \not\equiv 0$, or $X(\varphi) \perp L^2(B)$ for every $\varphi \in \mathfrak{D}$ and so $L^2(X) \perp L^2(B)$; but this would contradict the fact that $L^2(B) \subset L^2(X)$ and contains elements of positive norm.

Now $X_1(\varphi) \in L^2(B)$ and so for some $g \equiv g_\varphi \in L^2(-\infty, \infty)$ we have

$$X_1(\varphi) = \int_{-\infty}^{t(\varphi)} g_\varphi(u)\, dB(u); \tag{4.9}$$

that $g_\varphi(u)$ vanisnes for almost all $u > t(\varphi)$ following from (4.8). This representation has some interesting consequences which are indicated in the lemmas which follow.

LEMMA 4.1. g_φ commutes with translations, in the sense that

$$g_{\tau_h \varphi} = \tau_h g_\varphi. \tag{4.10}$$

PROOF: We have, on the one hand,

$$X_1(\tau_h \varphi) = \int_{-\infty}^{t(\varphi)-h} g_{\tau_h \varphi}(u)\, dB(u),$$

and on the other,

$$X_1(\tau_h \varphi) = T_{-h} X_1(\varphi) = T_{-h} \int_{-\infty}^{t(\varphi)} g_\varphi(u)\, dB(u) = \int_{-\infty}^{t(\varphi)-h} \tau_h g_\varphi(u)\, dB(u),$$

by definition of the integral and (4.4). Hence $g_{\tau_h \varphi} = \tau_h g_\varphi$.

LEMMA 4.2. There exists a (Schwartz) distribution G such that

$$g_\varphi = G*\varphi . \tag{4.11}$$

Further $G\in\mathcal{S}'$ and $\mathfrak{F}G$ is a function which is the product of a polynomial and a square-summable function.

PROOF: $g_\varphi\in L^2(-\infty,\infty)$ and hence $g_\varphi\in\mathfrak{D}'$. (4.9) shows that the correspondence $\varphi\to g_\varphi$ is a linear mapping of $\mathfrak{D}$ into $\mathfrak{D}'$. It is continuous because if $\varphi_n\to\varphi$ in $\mathfrak{D}$, then

$$\begin{aligned}||X(\varphi_n)-X(\varphi)||^2 &= \left\|\int_{-\infty}^{t(\varphi_n)} g_{\varphi_n}(u)\,dB(u)-\int_{-\infty}^{t(\varphi)} g_\varphi(u)\,dB(u)\right\|^2\\ &= \left\|\int_{-\infty}^{a} (g_{\varphi_n}(u)-g_\varphi(u))\,dB(u)\right\|^2\end{aligned}$$

for some a, recalling the definition of convergence in $\mathfrak{D}$, and by (4.6) the last espression is equal to

$$\int_{-\infty}^{a} |g_{\varphi_n}(u)-g_\varphi(u)|^2du ,$$

and so $g_{\varphi_n}\to g_\varphi$ in L^2 and therefore in $\mathfrak{D}'$. On the other hand by Lemma 4.1, the mapping commutes with translations. Hence by a well-known theorem of Schwartz [6, II, p. 53], the representation (4.11) holds for some $G\in\mathfrak{D}'$.

Next, since $\varphi\in\mathfrak{D}$, $G*\varphi=g_\varphi\in C^\infty$. But $g_\varphi^{(p)}\in L^2(-\infty,\infty)$ for each $p=0,1,2,\cdots$, so $g_\varphi\in\mathfrak{D}_{L^2}$. Since for every $\varphi\in\mathfrak{D}$, $G*\varphi\in\mathfrak{D}_{L^2}$, again by a theorem of Schwartz [6, II, p. 56], $G\in\mathfrak{D}'_{L^2}$. On the other hand a Fourier transformation gives $\mathfrak{F}G\cdot\mathfrak{F}\varphi=\mathfrak{F}g_\varphi\in L^2(-\infty,\infty)$, so that $\mathfrak{F}G$ is a function—recall that $\mathfrak{F}\varphi$ has at most a countable number of zeros. But $G\in\mathfrak{D}'_{L^2}$. Therefore [6, II, p. 126] $\mathfrak{F}G$ is the product of a polynominal and a square-summable function. This completes the proof of the lemma.

We have thus shown that X_1 defined by (4.7) has the representation

$$X_1(\varphi) = \int (G*\varphi)(u)\,dB(u) . \tag{4.12}$$

(We shall sometimes omit the limits of integration; in any case they may be taken to be $-\infty$ and ∞.) A consequence of this is

LEMMA 4.3. X_1 has a spectral measure $d\mu_1$ which is absolutely continuous relative to Lebesgue measure.

PROOF: A routine computation using (2.2), (4.12), (4.6) and the Parseval relation shows that

$$\int \mathfrak{F}(\varphi*\tilde{\psi})(\lambda)\,d\mu_1(\lambda) = \int \mathfrak{F}(\varphi*\tilde{\psi})(\lambda)\,|\mathfrak{F}G(\lambda)|^2 d\lambda\,.$$

And since elements of the form $\varphi*\tilde{\psi}$ are dense in $\mathcal{S}$, we get

$$d\mu_1(\lambda) = |\mathfrak{F}G(\lambda)|^2 d\lambda\,.$$

Thus $d\mu_1$ is absolutely conttnuous.

This enables us to prove the following crucial

LEMMA 4. 4. $d\mu$ is absolutely continuous.

PROOF: Let $d\mu = d\mu_c + d\mu_s$ be the Lebesgue decomposition of $d\mu$ into its absolutely continuous and singular parts respectively. By (2. 3), if dM_c and dM_s are the random measures corresponding to $d\mu_c$ and $d\mu_s$, then we have

$$X(\varphi) = \int (\mathfrak{F}\varphi)(\lambda)\,dM(\lambda) = \int (\mathfrak{F}\varphi)(\lambda)\,dM_c(\lambda) + \int (\mathfrak{F}\varphi)(\lambda)\,dM_s(\lambda)$$
$$\equiv X_c(\varphi) + X_s(\varphi)\,,$$

where X_c and X_s now are clearly mutually orthogonal stationary random distributions, both purely non-deterministic in view of the relations

$$X_c(\varphi) = (X(\varphi)\,|\,L^2(M_c)), \quad X_s(\varphi) = (X(\varphi)\,|\,L^2(M_s))\,,$$

$L^2(M_c)$ and $L^2(M_s)$ having the obvious significations. Now consider X_s and write

$$X_s(\varphi) = X_s^{(1)}(\varphi) + X_s^{(2)}(\varphi)\,,$$

where

$$X_s^{(1)}(\varphi) = (X_s(\varphi)\,|\,L^1(B_s)), \quad X_s^{(2)}(\varphi) = X_s(\varphi) \ominus (X_s(\varphi)\,|\,L^2(B_s))\,,$$

B_s being the Brownian random measure corsesponding to X_s, constructed according to the procedure we have described for X. Let $X_s^{(i)}$ have spectral measure $d\mu_s^{(i)}$, $i=1, 2$. Now suppose, if possible, that $X_s(\varphi) \not\equiv 0$; by an earlier argument $X_s^{(1)}(\varphi) \not\equiv 0$ and $X_s^{(1)}$ has a representation of the type (4. 12). Consequently, by Lemma 4. 3, $d\mu_s^{(1)}$ is absolutely continuous. The $X_s^{(i)}$ being mutually orthogonal, the spectral measures add up: $d\mu_s = d\mu_s^{(1)} + d\mu_s^{(2)}$. But this equation states that the singular measure $d\mu_s$ has a non-trivial absolutely continuous component. This contradiction shows that $X_s(\varphi) \equiv 0$, i.e. that $d\mu$ is absolutely continous.

A familiar argument [1, p. 532], suitably altered, enables us to prove the following partial converse to Lemma 4. 4; it extends to random distributions a theorem of Kolmogoroff and Karhunen [1, p. 532] on stationary processes.

LEMMA 4. 5. If a stationary random distribution (not necessarily purely non-deterministic) has an absolutely continuous spectral measure, then it admits a (possibly two-sided) moving average representation

$$X(\varphi) = \int (T*\varphi)(u)\,dB(u)$$

where T is a tempered distribution.

5. The backward moving average representation

We wish to uphold the representation (4. 12) not only for X_1 but also for X. For this it would suffice to show that $X_1 = X$; equivalently, that $L^2(B) = L^2(X)$. This we proceed to do by an adaptation of Hanner's argument [2, Prop. D], using Lemma 4. 4 in conjunction with Lemma 2. 1.

LEMMA 5. 1. $X_1 = X$.

PROOF: Let

$$Y(\varphi) = X(\varphi) \ominus X_1(\varphi)\,;$$

then Y is a stationary random disiribution and from (4. 8) it easily follows that it is purely non-deterministic. Also $Y(\varphi)$ which is equal to $(X(\varphi) | L^2(X) \ominus L^2(B))$ is orthogonal to $L^2(B)$. We have thus a decomposition of X as the sum of two purely non-deterministic random distributions: $X = X_1 + Y$ and $X_1 \not\equiv 0$. We shall show that $Y=0$. If not, we construct for Y, as we did for X, a Brownian random measure B'; and projecting on $L^2(B')$ we get

$$Y(\varphi) = Y_1(\varphi) + Y_2(\varphi)\,,$$

with $Y_1(\varphi) = (Y(\varphi) / L^2(B'))$, $Y_2(\varphi) \perp L^2(B')$ and $Y_1(\varphi) \not\equiv 0$. Then

$$X(\varphi) = X_1(\varphi) + Y_1(\varphi) + Y_2(\varphi)\,.$$

For X_1 we have the representation (4. 9):

$$X_1(\varphi) = \int_{-\infty}^{t(\varphi)} g_\varphi(u)\,dB(u)\,,$$

$g_\varphi(u) = 0$ for $u > t(\varphi)$. Correspondingly, for B', $B'((a, b]) \in L^2(Y)$ and for a suitable $h \equiv h_\varphi \in L^2(-\infty, \infty)$ vanishing for $u > t(\varphi)$, we have

$$Y_1(\varphi) = \int_{-\infty}^{t(\varphi)} h_\varphi(u)\, dB'(u)\,. \tag{5.1}$$

Since $Y_1(\varphi) \perp L^2(B)$ and $L^2(B') \subset L^2(Y)$, we have $L^2(B') \perp L^2(B)$. Also $Y_2(\varphi) \in L^2(Y) \perp L^2(B)$ and $Y_2(\varphi) \perp L^2(B')$. Finally,

$$B'((a, b]) \in L^2(Y\,; b) \subset L^2(X\,; b)\,,$$

so that for each t we have

$$L^2(B'((a, b]), \quad a < b \leqslant t) \subset L^2(X\,; t)\,. \tag{5.2}$$

Now consider, for a fixed $\varphi_0 \in \mathfrak{D}_0$, $X_1(\varphi_0) \neq 0$, and some $k < 0$, the element

$$\xi = \int_k^{t(\varphi_0)} \overline{h_{\varphi_0}(k-u)}\, dB(u) - \int_k^{t(\varphi_0)} \overline{g_{\varphi_0}(k-u)}\, dB'(u) \tag{5.3}$$

In order to complete the proof we need only show that for properly chosen k, ξ represents a non-zero element of $L^2(B)$ which is orthogonal to $L^2(X)$. Now, since $t(\varphi_0) = 0$,

$$\|\xi\|^2 = \int_k^0 |h_{\varphi_0}(k-u)|^2 du + \int_k^0 |g_{\varphi_0}(k-u)|^2 du\,,$$

and this is positive for some negative k since

$$\lim_{k \to -\infty} \int_k^0 |g_{\varphi_0}(k-u)|^2 du = \int_{-\infty}^0 |g_{\varphi_0}(u)|^2 du = \|X(\varphi_0)\|^2 > 0\,.$$

We now check that $\xi \perp L^2(X)$. We observe first that $\xi \perp X(\varphi_0)$; indeed

$$X(\varphi_0) = \int_{-\infty}^0 g_{\varphi_0}(u)\, dB(u) + \int_{-\infty}^0 h_{\varphi_0}(u)\, dB'(u) + Y_2(\varphi_0)\,,$$

and by virtue of (4.6) and the orthogonality relations mentioned above we have

$$(\xi,\ X(\varphi_0)) = \int_{-\infty}^0 g_{\varphi_0}(u)\, h_{\varphi_0}(k-u) du - \int_{-\infty}^0 h_{\varphi_0}(u)\, g_{\varphi_0}(k-u)\, du = 0\,.$$

Next, $\xi \perp X(\tau_l \varphi_0)$ for any l. For, in view of Lemma 4.1 we have

$$X(\tau_l \varphi_0) = \int_{-\infty}^{-l} \tau_l\, g_{\varphi_0}(u)\, dB(u) + \int_{-\infty}^{-l} \tau_l\, h_{\varphi_0}(u)\, dB'(u) + Y_2(\tau_l \varphi_0)$$

and consequently

$$(\xi,\ X(\tau_l \varphi_0)) = \int_{-\infty}^{-l} \tau_l\, g_{\varphi_0}(u)\, h_{\varphi_0}(k-u)\, du - \int_{-\infty}^{-l} \tau_l\, h_{\varphi_0}(u)\, g_{\varphi_0}(k-u)\, du = 0\,,$$

since to translate a convolution we need translate only one of the factors. Thus ξ is orthogonal to the subspace generated by the translates of $X(\varphi_0)$. Finally, by the absolute continuity of $d\mu$ established in Lemma 4. 4, Lemma 2. 1 becomes operative, so that the space generated by the translates of $X(\varphi_0)$ is the entire space $L^2(X)$ and hence $\xi \perp L^2(X)$. This completes the proof.

As a result, equation (4. 12) holds with X_1 replaced by X. Now φ has its support in $(-\infty, t(\varphi)]$ while $g_\varphi = G*\varphi$ has its in $(-\infty, t(\varphi)]$ too. It is well known [6, II, p. 12] that the support of a convolution is contained in the sum of the supports of its factors; hence by the continuity of the mapping φ to g_φ, $\operatorname{supp} G \subset (-\infty, 0]$. We have thus completely proved the following fundamental theorem on the 'backward moving average representation' of a purely non-deterministic stationary random distribution:

THEOREM 5. 1. X has the representation

$$X(\varphi) = \int_{-\infty}^{t(\varphi)} g_\varphi(u)\, dB(u) = \int_{-\infty}^{t(\varphi)} (G*\varphi)(u)\, dB(u)\,, \tag{5. 4}$$

where G is a tempered distribution with support in $(-\infty, 0]$ whose Fourier transform is a function which is the product of a polynomial and a square-summable function, and B is a Brownian random measure.

REMARK. As in Hanner [2, § 6], it can be shown easily that G is unique up to multiplication by a factor of absolute value 1.

(5. 4) taken along with (4. 8) gives us directly

LEMMA 5. 2. $L^2((B(a, b]),\ a < b \leqslant t) = L^2(X; t)$ for each t.

This fact gives us immediately the solution to the prediction problem: the predictor of $X(\varphi)$ relative of $L^2(X; 0)$ is

$$(X(\varphi)\,|\,L^2(X; 0)) = (X(\varphi)\,|\,L^2(B; 0)) = \int_{-\infty}^{0} (G*\varphi)(u)\, dB(u)\,;$$

and the error of prediction σ is given by

$$\sigma^2 = \int_0^{t(\varphi)} |(G*\varphi)(u)|^2\, du\,.$$

6. The spectral criterion

We proceed next to obtain a spectral criterion which would guarantee a representation such as (5. 4). Our point of departure will be the following generalization of a classical theorem of Paley and Wiener [4, Theorem XII].

LEMMA 6.1. A necessary and sufficient condition that a non-trivial function $\mu'(x)$ which is the product of a polynomial and a square-summable function shall be such that $\mu'(x)=|\nu(x)|$ where $\mathfrak{F}\nu\in\mathcal{S}'$ has its support in $(-\infty, 0]$ is that

$$\int_{-\infty}^{\infty}\frac{\log\mu'(x)}{1+x^{2}}dx>-\infty. \tag{6.1}$$

PROOF: The case in which $\mu'\in L^{2}(-\infty, \infty)$ is covered by the Paley-Wiener theorem. We resort to smoothing to deal with the general case.

a) Necessity: Suppose first that $\mu'(x)=|\nu(x)|$, $\operatorname{supp}\mathfrak{F}\nu\subset(-\infty, 0]$. Choose $m\in\mathcal{S}$ such that

$$\int_{-\infty}^{\infty}\frac{\log m(x)}{1+x^{2}}dx>-\infty; \tag{6.2}$$

such an m clearly exists. Since *a fortiori* $m\in L^{2}(-\infty, \infty)$, by the Paley-Wiener theorem we can find n such that

$$m(x)=|n(x)|,\quad \mathfrak{F}n\in L^{2}\subset\mathcal{S}',\quad \operatorname{supp}\mathfrak{F}n\subset(-\infty, 0].$$

Now consider $\mu'(x)m(x)$. This clearly belongs to $L^{2}(-\infty, \infty)$ and $\mu'(x)m(x)=|\nu(x)n(x)|$, while $\mathfrak{F}(\nu n)=\mathfrak{F}\nu*\mathfrak{F}n$, so that $\operatorname{supp}\mathfrak{F}(\nu n)\subset\operatorname{supp}\mathfrak{F}\nu+\operatorname{supp}\mathfrak{F}n\subset(-\infty, 0]$. Hence, again by the Paley-Wiener theorem,

$$\int_{-\infty}^{\infty}\frac{\log\mu'(x)+\log m(x)}{1+x^{2}}dx>-\infty. \tag{6.3}$$

(6.2) and (6.3) together imply (6.1).

b) Sufficiency: Now suppose that $\mu'(x)$ is of the stated order of growth, $\mu'(x)=|\nu(x)|$ and (6.1) holds. We shall prove that $\mathfrak{F}\nu\subset(-\infty, 0]$. Let $m(x)$ be as above. Then $\mu'm\in L^{2}(-\infty, \infty)$, $\mu'(x)m(x)=|\nu(x)n(x)|$, and since (6.1) and (6.3) hold, so does (6.3); another appeal to the Paley-Wiener theorem gives us $\operatorname{supp}\mathfrak{F}(\nu n)\subset(-\infty, 0]$. Again $\operatorname{supp}\mathfrak{F}(\nu n)\subset\operatorname{supp}\mathfrak{F}\nu+\operatorname{supp}\mathfrak{F}n$ and hence $\operatorname{supp}\mathfrak{F}\nu\subset(-\infty, 0]$. This finishes the proof of the lemma.

The spectral criterion we are looking for is now given by

THEOREM 6.1. The representation (5.4) holds if and only if $d\mu$ is absolutely continuous with spectral density $\mu'(x)$ satisfying (6.1). (It follows that $\mu'(x)>0$ a.e.).

PROOF: Suppose first that (5.4) holds. By the computation indicated in the proof of Lemma 4.3, $d\mu$ is absolutely continuous and $\mu'(x)=|(\mathfrak{F}G)(x)|^2$, so that $\sqrt{\mu'(x)}=|(\mathfrak{F}G)(x)|$ is of the order of a polynomial multiplied by a square-summable function and $\operatorname{supp} G \subset (-\infty, 0]$. By Lemma 6.1 then μ' satisfies (6.1).

Conversely, if $d\mu$ is absolutely continuous with (6.1) holding, then by Lemma 4.5 we have a representation of the type (5.4) with the support of G arbitrary, and by the same computation as the one mentioned above, $\mu'(x)=|(\mathfrak{F}G)(x)|^2$. On the other hand, by Lemma 6.1, μ' has just this representation with $\operatorname{supp} G \subset (-\infty, 0]$. Our result now follows from the uniqueness of the Fourier transform. Theorem 6.1 is thus completely proved.

We may finally summarize all our results in the following

THEOREM 6.2. For a stationary random distribution X the three following statements are equivalent:

(i) X is purely non-deterministic;

(ii) X has a backward moving average representation (5.4);

(iii) X has an absolutely continuous spectral measure with density μ' satisfying (6.1).

7. Concluding remarks

A mean-continuous stationary stochastic process $x(t)$ defines a stationary random distribution X by

$$X(\varphi) = \int x(t)\varphi(t)\,dt\,,$$

the integral being interpreted as a Bochner integral. It is easily checked that $L^2(x)=L^2(x)$ the space spanned by $\{x(t)\}$. Also $L^2(X\,;\,a)=L^2(x\,;\,a)$ for every a; the definitions of the terms 'deterministic' and 'purely non-deterministic' coincide whether one looks upon $x(t)$ as a process or as a distribution. The prediction theory of stationary processes is thus, naturally, subsumed under that of stationary random distributions.

In the case of an X associated with a process $x(t)$, the representation (5.4) reduces to the usual backward moving average representation of processes [2]. In fact, $G*\varphi$ now reduces to

$G*\delta_s$, so that the kernel is a translate of a distribution which which would reduce, in this case, to an L^2-function.

Tata Institute of Fundamental
Research, Bombay and
Kyoto University

REFERENCES

[8] J. L. Doob: *Stochastic processes*, New York; 1953.

[2] O. Hanner: Deterministic and non-deterministic stationary random processes, *Arkiv för Mat.* II, 14; 1949; 161-177.

[3] K. Itô: Stationary random distributions, *Mem. Coll. Sci. Univ. Koyoto* A, XXVIII 3; 1953; 209-223.

[4] R.E.A.C. Paley and N. Wiener: *Fourier transforms in the complex domain*, AMS Colloq. XIX, New York; 1934.

[5] Yu. A. Rozanov: On the extrapolation of random distributions (Russian), *Tiiorea Veroyat.* IV; 1959; 465-471.

[6] L. Schwartz: *Theorie des distributions*, Paris; I, 1950; II, 1951.

[7] N. Wiener: *The Fourier integral and certain of its applications*, Cambridge; 1933.

[8] A. M. Yaglom: Correlation theory of processes with random stationary *n*th order increments, *AMS Translations*, 2 (8); 1958; 87-141.