

INDEPENDENCE AND DEPENDENCE

M. ROSENBLATT

BROWN UNIVERSITY

1. Introduction

A stochastic process is commonly used as a model in studying the behavior of a random system through time. It will be convenient for us to take the stochastic process $\{x_t\}$ as *discrete in time* $t = \dots, -1, 0, 1, \dots$. Processes of independent random variables are the simplest and most completely understood. It is, however, clear that these are extremely limited in scope as models and one must have recourse to dependent processes (the random variables x_t not independent) in order to have any power in description. For simplicity, *let us further restrict ourselves to* processes that are stable through time, *stationary processes*. For such processes the probabilities of events shifted through time remain the same, that is, the probability

$$(1) \quad P\{x_{t_1+h} \leq a_1, \dots, x_{t_n+h} \leq a_n\}$$

is independent of h . Such models occur fairly often in the physical sciences. If mean properties of the process are to be capable of being estimated reasonably well from part of a realization of the process, some form of asymptotic independence for blocks of random variables of the process that are widely separated must be satisfied. This is, in effect, the gist of many of the results in ergodic theory. Two types of interesting problems are posed. *The first* of these is *concerned with reasonable notions of asymptotic independence and what types of processes satisfy them. The second is that of characterizing those processes $\{x_t\}$ that can be constructed out of independent processes by a function and its shifts*, that is,

$$(2) \quad x_t = f(\dots, \xi_{t-1}, \xi_t, \xi_{t+1}, \dots)$$

where $\{\xi_t\}$ is a process of independent random variables. Neither of these questions have elicited satisfactory answers. However, there are some small results that do give insights into the problems. The object of this paper is a presentation and discussion of a few of these limited results.

2. Mixing

Ergodicity itself might be thought of as a form of asymptotic independence. However, the most obvious formulation of asymptotic independence is the

This research was supported by the Office of Naval Research.

stronger *mixing condition*. Suppose T is the shift (in time) operator associated with the stationary process. The process is said to be mixing if

$$(3) \quad P\{A \cap T^n B\} \rightarrow P\{A\}P\{B\}$$

as $n \rightarrow \infty$ for any two events A, B . The normal stationary processes are the simplest stationary processes and Maruyama [6] has obtained a simple and intuitive necessary and sufficient condition for such a process to be mixing. To verify the mixing property, it is enough to verify property (3) for any two events A, B determined by conditions on a finite number of the random variables x_t .

MARUYAMA'S THEOREM. *A necessary and sufficient condition for a normal stationary process to be mixing is that the covariance $r_t = \text{Cov}(x_{\tau+t}, x_\tau)$ tend to zero as $|t| \rightarrow \infty$.*

It is clear that no generality is lost by assuming that the mean of the process is zero. The necessity of the condition that $r_t \rightarrow 0$ as $|t| \rightarrow \infty$ is then almost completely apparent. We approximate the random variables $x_{\tau+t}, x_\tau$ by step functions in the following way. Let

$$(4) \quad A(k, \epsilon) = \{\omega | k\epsilon \leq x_\tau(\omega) < (k + 1)\epsilon\}$$

and

$$(5) \quad c_{A(k, \epsilon)}(\omega) = \begin{cases} 1 & \text{if } \omega \in A(k, \epsilon) \\ 0 & \text{otherwise.} \end{cases}$$

Set

$$(6) \quad x_\tau^{(N, \epsilon)}(\omega) = \sum_{k=-N}^{K=N} k\epsilon c_{A(k, \epsilon)}(\omega).$$

Then for sufficiently large N

$$(7) \quad \begin{aligned} E|x_\tau^{(N, \epsilon)}(\omega) - x_\tau(\omega)| &= E|x_\tau^{(N, \epsilon)}(T^t\omega) - x_{t+\tau}(\omega)| < \delta, \\ |r_t - E x_\tau^{(N, \epsilon)}(T^t\omega)x_\tau^{(N, \epsilon)}(\omega)| &\leq E|x_{t+\tau}(\omega)x_\tau(\omega) - x_\tau^{(N, \epsilon)}(T^t\omega)x_\tau^{(N, \epsilon)}(\omega)| < \delta. \end{aligned}$$

Since the mean of the process is zero, for N sufficiently large

$$(8) \quad |E x_\tau^{(N, \epsilon)}(\omega)| = |E x_\tau^{(N, \epsilon)}(T^t\omega)| < \delta.$$

However by the mixing condition

$$(9) \quad \begin{aligned} E x_\tau^{(N, \epsilon)}(T^t\omega)x_\tau^{(N, \epsilon)}(\omega) &= \sum_{k, j=-N}^N k\epsilon j\epsilon E c_{A(k, \epsilon)}(T^t\omega)c_{A(j, \epsilon)}(\omega) \\ &= \sum_{k, j=-N}^N k j \epsilon^2 P[A(j, \epsilon) \cap T^t A(k, \epsilon)] \\ &\rightarrow \sum_{k, j=-N}^N k j \epsilon^2 P[A(j, \epsilon)]P[A(k, \epsilon)] \\ &= [E x_\tau^{(N, \epsilon)}(\omega)]^2. \end{aligned}$$

But this limit relation coupled with the inequalities (7) indicate that

$$(10) \quad \overline{\lim}_{|t| \rightarrow \infty} |r_t| < 3\delta$$

and since this is true for any δ , we have $\lim_{|t| \rightarrow \infty} r_t = 0$.

The sufficiency of this condition follows just as easily. Let A, B be determined by conditions on blocks of k successive random variables (these could in fact be taken to be the same blocks if k is sufficiently large). Then

$$(11) \quad P\{A \cap T^n B\} = (2\pi)^{-k} |M|^{-1/2} \int_A \int_B \exp \left\{ -\frac{1}{2} (u, v) M^{-1} (u, v)' \right\} du dv,$$

where u, v are two row k -vectors and (u, v) the $2k$ -vector formed by joining them. The matrix M is the covariance matrix of two blocks of k successive random variables of the process separated by $n - k$ random variables. Write M in the form

$$(12) \quad M = \begin{pmatrix} R_1 & R_2 \\ R_2' & R_1 \end{pmatrix}$$

where R_1 is the $k \times k$ covariance matrix of a block of k successive random variables and R_2 the $k \times k$ matrix of covariances of the one block of k with the other block removed from it by $n - k$ random variables (R_2' is the transpose of R_2). The fact that $r_t \rightarrow 0$ as $|t| \rightarrow \infty$ implies that M^{-1} exists. It is in fact given by

$$(13) \quad M^{-1} = \begin{pmatrix} A & B \\ B' & C \end{pmatrix}$$

with $A = (R_1 - R_2 R_1^{-1} R_2')^{-1}$, $C = (R_1 - R_2' R_1^{-1} R_2)^{-1}$, $B = -R_1^{-1} R_2 C = -A R_2 R_1^{-1}$. The matrix R_1 is fixed since k is fixed. But R_2 is a $k \times k$ matrix whose entries are the covariances r_t with $|t| > n - k$. They approach zero uniformly as $n \rightarrow \infty$ so that R_2 approaches the null matrix. Thus

$$(14) \quad M^{-1} \rightarrow \begin{pmatrix} R_1^{-1} & 0 \\ 0 & R_1^{-1} \end{pmatrix}$$

$$(15) \quad P\{A \cap T^n B\} \rightarrow (2\pi)^{-k} |R_1|^{-1} \int_A \int_B \exp \left\{ -\frac{1}{2} [u R_1^{-1} u' + v R_1^{-1} v'] \right\} du dv \\ = P\{A\} P\{B\}.$$

Stronger mixing conditions are sometimes invoked in deriving results like central limit theorems for dependent processes. The following condition is an example of such a strong mixing condition: Let A, B be any events determined by conditions on the random variables x_k , $k \leq m$, and x_k , $k \geq n$, respectively with $n > m$. The process $\{x_k\}$ is said to satisfy the strong mixing condition M if

$$(16) \quad |P\{A \cap B\} - P\{A\}P\{B\}| \leq d(n - m)$$

for all such events A, B and some function d where d is a function on the positive integers n that decreases to zero as $n \rightarrow \infty$. Notice that such a mixing condition is a uniform mixing condition in what we might call the distance $n - m$ between the events A, B .

Let $\{x_k\}$ be a stationary process with mean zero, $E x_k \equiv 0$, that satisfies the strong mixing condition M . Further let the process satisfy the two moment conditions

$$(17) \quad E \left| \sum_{j=1}^n x_j \right|^2 = \sum_{k=-n}^n (n - |k|) r_k = h(n) \rightarrow \infty \quad \text{as } n \rightarrow \infty,$$

$$(18) \quad E \left| \sum_{j=1}^n x_j \right|^4 = \sum_{j_1=1}^n E x_{j_1} x_{j_2} x_{j_3} x_{j_4}$$

$$= \sum_{j_1=1}^n [r_{j_1-j_2} r_{j_3-j_4} + r_{j_1-j_3} r_{j_2-j_4} + r_{j_1-j_4} r_{j_2-j_3} + P_{j_2-j_1, j_3-j_1, j_4-j_1}]$$

$$= 3[h(n)]^2 + \sum_{|a_i| \leq n} P_{a_1, a_2, a_3} N(a_1, a_2, a_3)$$

$$= O[h(n)]^2.$$

The sums $\sum_{j=1}^n x_j$ can then be shown to be asymptotically and nontrivially normally distributed when suitably normed as $n \rightarrow \infty$. Of course, the statement of nontrivialness means that the limiting normal distribution does not have variance zero. The proof of this limit theorem can be found in [7]. The following two examples in part at least indicate the limits of the central limit theorem cited above.

In the first example the process $\{x_k\}$ is taken to be

$$(19) \quad x_k = \epsilon_k - \epsilon_{k-1}$$

where the ϵ_k are independent, identically distributed nonnormal random variables with finite 4th moment and vanishing 1st moment. Since $\sum_{k=1}^n x_k = \epsilon_n - \epsilon_0$, it is readily seen that $h(n)$ does not diverge as $n \rightarrow \infty$. Thus, we get a nonnormal limiting distribution, namely that of $\epsilon_1 - \epsilon_0$, without norming the sum at all. The process certainly satisfies the strong mixing condition and moment condition (18) but not moment condition (17). The difficulty arises because the process $\{x_k\}$ has too little spectral mass in the vicinity of $\lambda = 0$. Its spectral density is $f(\lambda) = (1/2\pi)|1 - \exp(-i\lambda)|^2$.

The second example has greater interest. Consider a normal process $\{y_k\}$ with mean zero and covariance sequence $r_k = (1 + k^2)^{-\gamma}$ with $\gamma > 0$. The spectral density of the process $\{y_k\}$ is

$$(20) \quad a(\lambda) = \sum_{k=-\infty}^{\infty} b(\lambda + 2k\pi)$$

where

$$(21) \quad b(\lambda) = b(|\lambda|) = \{\Gamma(\gamma)\}^{-2} e^{-|\lambda|} \int_0^{\infty} \mu^{\gamma-1} (\mu + |\lambda|)^{\gamma-1} e^{-2\mu} d\mu.$$

The spectral density $a(\lambda)$ is continuous and bounded away from zero if $|\lambda| > \epsilon > 0$. If $\gamma < 1/2$, it has a singularity of the form $|\lambda|^{2\gamma-1}$ in the vicinity of $\lambda = 0$. However, it should be noted that the process is in any case, purely non-deterministic and therefore mixing in the ordinary sense. Let the process $\{x_k\}$ be of the form $x_k = y_k^2 - 1$. The covariance sequence of the process $\{x_k\}$ is $r_k = 2(1 + k^2)^{-2\gamma}$. We shall be interested in processes for which $\gamma < 1/4$. The spectral density $f(\lambda)$ of $\{x_k\}$ has a singularity of the form $|\lambda|^{4\gamma-1}$ in the neighborhood of $\lambda = 0$ as can be seen from (21). We shall see that $n^{-1+2\gamma} \sum_{k=1}^n x_k$ has a

limiting nonnormal distribution as $n \rightarrow \infty$. For the characteristic function of $n^{-1+2\gamma} \sum_{k=1}^n x_k$ is

$$(22) \quad |I - 2itn^{-1+2\gamma}R|^{-1/2} \exp \{-in^2\gamma tr_0\} = \exp \left\{ \sum_{k=2}^{\infty} (2itn^{-1+2\gamma})^k \text{sp}(R^k)/k \right\}$$

and

$$(23) \quad \begin{aligned} &(n^{-1+2\gamma})^k \text{sp}(R^k) \\ &= (n^{-1+2\gamma})^k \sum_{i_j=1}^n r_{i_1-i_2} r_{i_2-i_3} \cdots r_{i_k-i_1} \\ &\rightarrow \int \underbrace{\cdots}_{k} \int |x_1 - x_2|^{-2\gamma} |x_2 - x_3|^{-2\gamma} \cdots |x_k - x_1|^{-2\gamma} dx_1 \cdots dx_k \\ &= c_k > 0, \end{aligned}$$

as $n \rightarrow \infty$ (here $\text{sp}(R)$ denotes the trace of the matrix R). The characteristic function of the limiting distribution is

$$(24) \quad \exp \left\{ \sum_{k=2}^{\infty} (2it)^k \frac{c_k}{k} \right\}$$

the characteristic function of a nonnormal distribution. The curious aspect of this example is that moment conditions 1 and 2 are satisfied. We must therefore conclude that the strong mixing condition is not satisfied by $\{x_k\}$. But this means that the normal process $\{y_k\}$ does not satisfy a strong mixing condition, since $\{x_k\}$ is derived from it by an instantaneous nonlinear filter, even though $\{y_k\}$ is purely nondeterministic and hence mixing. *This is rather surprising.* It certainly indicates that *there would be a great deal of interest in a result that would characterize the Gaussian stationary processes that satisfy the strong mixing condition discussed in this section.*

3. Stationary processes generated by independent random variables

Most of the stationary processes that arise in practice are generated from independent identically distribution random variables by a linear or nonlinear operation and its shifts. Let $\eta = (\cdots, \eta_{-1}, \eta_0, \eta_1, \cdots)$ be a doubly infinite sequence of independent random variables with a common distribution function. Take T as the shift operator defined by

$$(25) \quad T\eta = \eta' = (\cdots, \eta_0, \eta_1, \eta_2, \cdots),$$

that is, the n th component of $\eta', (T\eta)'_n = \eta_{n+1}$. The processes we are thinking of are generated in the following way. Let g be a Borel function of an infinite number of variables such that the random variable $g(\eta)$ has finite second moment $E|g(\eta)|^2 < \infty$. A stationary process $\{x_n\}$ is generated by taking

$$(26) \quad x_n = g(T^n\eta), \quad n = 0, \pm 1 \cdots,$$

We can take $E\{x_n\} \equiv 0$ without any loss of generality. In this section we shall prove a result of Doob and Leibler [3] indicating that *the spectral distribution function of any process of this type is absolutely continuous*. The form of the proof given is that in [2].

In proving this result it is enough to assume that the η_j are uniformly distributed on the unit interval. For if η is not uniformly distributed, one can find a monotone function φ such that $\varphi(\xi)$ has the same distribution as η if ξ is uniformly distributed on the unit interval.

For each integer $m = 1, 2, \dots$ we define a process $\{x_{n,m}\}$ by

$$(27) \quad x_{n,m} = E\{x_n | \eta_{n-m+1}, \dots, \eta_{n+m-1}\}.$$

For fixed n , $\{x_{n,m}\}$ is a martingale. It follows from well-known theorems on martingales that $x_{n,m}$ converges to x_n in mean square and with probability one as $m \rightarrow \infty$.

Let $\theta_j(u_j), j = 0, 1, \dots$, be a sequence of orthonormal polynomials on the unit interval $[0, 1]$ with $\theta_j(u)$ of degree j precisely. It is well-known that the functions $\{\theta_j(u)\}$ are complete in $L^2[0, 1]$. The family of products $\{\theta_{j_1}(u_1) \dots \theta_{j_k}(u_k)\}$ is a complete orthonormal system for $L^2[U_k]$ where U_k is the k -dimensional unit cube. To get a complete orthonormal system for the infinite-dimensional cube U we take the function identically one and the system of products $\{\theta_{j_1}(u_{m_1}) \dots \theta_{j_k}(u_{m_k})\}$ where $m_1 < m_2 < \dots < m_k, j_1, \dots, j_k \geq 1$ and $k = 1, 2, \dots$. Let k be an arbitrary positive integer. From the discussion given above it follows that $x_{0,k}$ can be expanded in the form

$$(28) \quad x_{0,k} = \sum_{j=1}^{2k-1} \sum_{-k+1 \leq m_1 < \dots < m_j \leq k-1} \sum_{n_1, \dots, n_j=1}^{\infty} C(\bar{n}_j, \bar{m}_j) \varphi_{\bar{n}_j}(\eta_{\bar{m}_j})$$

where $\bar{m}_j = (m_1, \dots, m_j), \bar{n}_j = (n_1, \dots, n_j)$ and $\varphi_{\bar{n}_j}(\eta_{\bar{m}_j}) = \theta_{n_1}(\eta_{m_1}) \dots \theta_{n_j}(\eta_{m_j})$. Since $E\{|x_{0,k}|^2\} \leq E\{|x_0|^2\} < \infty$ it follows that

$$(29) \quad \sum_{j=1}^{\infty} \sum_{m_1 < m_2 < \dots < m_j} \sum_{n_1, \dots, n_j=1}^{\infty} C(\bar{n}_j, \bar{m}_j)^2 < \infty.$$

Since $x_0 = \lim_{k \rightarrow \infty} x_{0,k}$ with probability one it follows that

$$(30) \quad x_0 = \sum_{j=1}^{\infty} \sum_{m_1 < \dots < m_j} \sum_{n_1, \dots, n_j=1}^{\infty} C(\bar{n}_j, \bar{m}_j) \varphi_{\bar{n}_j}(\eta_{\bar{m}_j})$$

with probability one. The identity

$$(31) \quad x_n = \sum_{j=1}^{\infty} \sum_{m_1 < \dots < m_j} \sum_{n_1, \dots, n_j=1}^{\infty} C(\bar{n}_j, \bar{m}_j - n) \varphi_{\bar{n}_j}(\eta_{\bar{m}_j})$$

with $\bar{m}_j - n = (m_1 - n, \dots, m_j - n)$ follows in a similar manner.

Let k be a positive integer, $\bar{n} = (n_1, \dots, n_k)$ a k -tuple of positive integers and $\bar{m} = (m_1, \dots, m_k)$ a k -tuple of integers with $m_1 < \dots < m_k$. For any integer t let $x_t(\bar{m}, \bar{n})$ be given by

$$(32) \quad x_t(\bar{m}, \bar{n}) = \sum_{j=-\infty}^{\infty} C(\bar{n}, \bar{m} + j - t) \varphi_{\bar{n}}(\eta_{\bar{m}+j})$$

where $\bar{m} + l = (m_1 + l, \dots, m_k + l)$. $\{x_t(\bar{m}, \bar{n})\}$ is a moving average and such processes have an absolutely continuous spectrum. We shall distinguish (\bar{m}', \bar{n}') from (\bar{m}, \bar{n}) if $\bar{n} \neq \bar{n}'$ or $\bar{m} - \bar{m}'$ is not a vector with equal coordinates. Then the processes $x_t(\bar{m}, \bar{n})$ and $x_t(\bar{m}', \bar{n}')$ are orthogonal if (\bar{m}, \bar{n}) and (\bar{m}', \bar{n}') are distinct in the sense indicated above. There is an enumerable collection of such processes. Let $\{x_i^{(i)}\}$ be the i th process with $f^{(i)}(\lambda)$ its spectral density. Because of the orthogonality of the processes $\{x_i^{(i)}\}$ the spectral density $f_n(\lambda)$ of $\sum_{i=1}^n x_i^{(i)}$ is given by $\sum_{i=1}^n f^{(i)}(\lambda)$. The sequence $f_n(\lambda)$ is nondecreasing in n since the $f^{(i)}(\lambda)$ are non-negative. Let $f(\lambda) = \lim_{n \rightarrow \infty} f_n(\lambda)$. Then

$$(33) \quad \int_{-\pi}^{\pi} f(\lambda) d\lambda = \int_{-\pi}^{\pi} \lim_{n \rightarrow \infty} f_n(\lambda) d\lambda \leq \lim_{n \rightarrow \infty} \int_{-\pi}^{\pi} f_n(\lambda) d\lambda \\ = \lim_{n \rightarrow \infty} E \left\{ \sum_{i=1}^n x_i^{(i)} \right\}^2 = E\{|g|^2\} < \infty.$$

The function $f(\lambda)$ is finite almost everywhere and integrable. An easy computation indicates that $f(\lambda)$ is the spectral density of the process $\{x_t\}$ and the desired result is established. Analogous results for continuous parameter processes have been considered by Anzai [1], Kakutani [5] and Itô [4]. The importance of such processes has been stressed in the work of Wiener [10].

It follows from the result obtained above that if h is any square integrable function defined on the probability space of $x = (\dots, x_{-1}, x_0, x_1, \dots)$ and if $\{V_n\}$ is the process given by $V_n = h(T^n x)$, then $\{V_n\}$ has an absolutely continuous spectrum. One would like to know whether one can characterize such processes in terms of independent random variables with a representation like that of (26). At present this question is unresolved. In fact, not much work has been done on the question of determining those processes that can be given a representation of the form (26). Some insight into this question is given by the discussion in the remainder of this paper.

4. Structural questions for stationary processes

We now pose two interesting structural questions for stationary processes. Let $\{x_t\}$, with $t = 0, \pm 1, \dots$, be a stationary process and T the corresponding one-step shift operator. Further, let \mathfrak{B}_n be the Borel field generated by x_n, x_{n-1}, \dots . The first problem can be conveniently split into two parts.

PROBLEM 1. (a) *When can one find a random variable ξ_0 measurable with respect to \mathfrak{B}_0 , independent of x_{-1}, x_{-2}, \dots and such that $\mathfrak{B}_0 = \mathfrak{B}_{-1} \times \mathfrak{A}_0$ (product field) where \mathfrak{A}_0 is the Borel field generated by ξ_0 ? As we shall see, the answer to this part is almost immediate and rather simple. The second part is the nontrivial question.*

(b) *Notice that if (a) is satisfied, the random variables $\xi_n = T^n \xi_0, n = 0, \pm 1, \dots$ form a sequence of independent, identically distributed random variables. Let \mathfrak{A}_n be the Borel field generated by ξ_n . Is x_n measurable with respect to $\dots \times \mathfrak{A}_{n-1} \times \mathfrak{A}_n$?*

This problem can be easily rephrased in the language of information theory. The random variable ξ_n can be regarded as the output of a noiseless channel ξ that $\{x_n\}$ is sent through, that is, $\xi_n = \xi(x_n, x_{n-1}, \dots)$. In part (a) we ask that the output at time n , ξ_n , be independent of the past of the input $(x_{n-1}, x_{n-2}, \dots)$ and yet that ξ_n coupled with knowledge of the past of the input be enough to determine the input x_n at present. Thus we can write $x_n = f(\xi_n, x_{n-1}, x_{n-2}, \dots)$ and by iteration x_n as a function of $\xi_n, \xi_{n-1}, \dots, \xi_{n-k}, x_{n-k-1}, x_{n-k-2}, \dots$ for any finite k . In part (b) one asks whether one can go to the limit so that x_n can be considered a function of ξ_n, ξ_{n-1}, \dots alone.

Essentially the second question can be regarded as a reformulation of the first question where the condition that ξ be a noiseless channel is relaxed.

PROBLEM 2. Can one find a sequence of independent uniformly distributed (on $[0, 1]$) random variables ξ_n , with $n = 0, \pm 1, \dots$ and a Borel function $g(\xi) = g(\xi_0, \xi_{-1}, \dots)$, $\xi = (\dots, \xi_{-1}, \xi_0, \xi_1, \dots)$ such that

$$(34) \quad y_n = g(T^n \xi) \quad n = 0, \pm 1, \pm 2, \dots$$

has the same probability structure as x_n , with $n = 0, \pm 1, \dots$? Here T is the shift operator on ξ . Of course, a positive answer to problem 1 implies that problem 2 can be answered in the affirmative.

Let $\bigcap \mathcal{B}_n$ be the Borel field of sets common to all the Borel fields \mathcal{B}_n , with $n = 0, \pm 1, \dots$. Call a stationary process $\{x_i\}$ purely nondeterministic if the only functions measurable with respect to $\bigcap \mathcal{B}_n$ are the constant functions. This means that there is no information in the infinite past of the process. A little reflection indicates that if problem 1 or problem 2 are soluble for $\{x_n\}$, it must be a purely nondeterministic process (see [8]).

As remarked before, the answer to part (a) of problem 1 is not difficult. Let

$$(35) \quad F(a_n | a_{n-1}, a_{n-2}, \dots) = P\{x_n \leq a_n | x_{n-1} = a_{n-1}, x_{n-2} = a_{n-2}, \dots\}$$

be the conditional probability distribution function of $\{x_n\}$. Assume that (35) is a Borel measure function of the real variables a_n, a_{n-1}, \dots . Given two distribution functions F and G , we shall say they are equivalent if there is a one-one function mapping the jumps of F and G onto each other and preserving the size of the jumps. In [8] it is shown that there is a random variable ξ_n satisfying part (a) of problem 1 if and only if the distribution functions $F(a_n | a_{n-1}, \dots)$ (as functions of a_n) are equivalent for almost all a_{n-1}, a_{n-2}, \dots (with respect to the probability measure of the process $\{x_n\}$).

Let us assume that part (a) of problem 1 is satisfied and inquire under what conditions part (b) is fulfilled. The remarks of section 3 indicate that there is no loss in generality in assuming the processes $\{x_n\}$ and $\{\xi_n\}$ bounded. Naturally take $\xi(x_n, x_{n-1}, \dots)$ as a Borel measurable function of its arguments. Since $\mathcal{B}_0 = \mathcal{B}_{-1} \times \mathcal{G}_0$ it is clear that there is a family of transformations S_α on half infinite sequences into the real numbers induced by ξ such that

$$(36) \quad x_n = S_{\xi_n}(x_{n-1}, x_{n-2}, \dots)$$

with probability one where ξ_n is independent of x_{n-1}, x_{n-2}, \dots and is given

by $\xi_n = \xi(x_n, x_{n-1}, \dots)$. To say that x_n is measurable with respect to $\dots \times \mathcal{A}_{n-1} \times \mathcal{A}_n$ is equivalent to stating that $x_n = g(\xi_n, \xi_{n-1}, \dots)$ where g is a Borel measurable function. But this will hold if and only if

$$(37) \quad x_n - E(x_n | \xi_n, \xi_{n-1}, \dots, \xi_{n-k}) \rightarrow 0$$

in probability as $k \rightarrow \infty$ making use of the boundedness of the $\{x_n\}$ process. Let μ be the measure on half infinite sequences induced by the probability measure of the process $\{x_n\}$. For convenience we introduce the following notation. Let

$$(38) \quad \begin{aligned} \bar{x}_m &= (x_m, x_{m-1}, \dots) \\ S_{\xi_m} \bar{x}_{m-1} &= S_{\xi_m}(x_{m-1}, x_{m-2}, \dots) \\ \{S_{\xi_m} \bar{x}_{m-1}\} &= (S_{\xi_m} \bar{x}_{m-1}, x_{m-1}, x_{m-2}, \dots). \end{aligned}$$

We now give an argument proving the following theorem.

THEOREM. *The condition that*

$$(39) \quad x_n - E(x_n | \xi_n, \dots, \xi_1) \rightarrow 0$$

in probability as $n \rightarrow \infty$ is satisfied if and only if there is a set of points M of \bar{y}_0 points of μ measure one such that

$$(40) \quad S_{\xi_n} \{S_{\xi_{n-1}} \{ \dots \{S_{\xi_1} \bar{y}_0\} \dots \}\} - S_{\xi_n} \{S_{\xi_{n-1}} \{ \dots \{S_{\xi_1} \bar{y}'_0\} \dots \}\} \rightarrow 0$$

in $\xi_1, \dots, \xi_n, \dots$ measure for all \bar{y}_0, \bar{y}'_0 in M as $n \rightarrow \infty$.

Assume there is a set M of μ measure one satisfying (40). Take $\bar{z}_0 \in M$. For all $\bar{x}_0 \in M$ and any fixed $\epsilon > 0$ the measure

$$(41) \quad m\{(\xi_1, \dots, \xi_n) \mid |S_{\xi_n} \{ \dots \{S_{\xi_1} \bar{x}_0\} \dots \} - S_{\xi_n} \{ \dots \{S_{\xi_1} \bar{z}_0\} \dots \} | > \epsilon\} \rightarrow 0$$

as $n \rightarrow \infty$. This implies that

$$(42) \quad \begin{aligned} P\{|x_n - S_{\xi_n} \{ \dots \{S_{\xi_1} \bar{z}_0\} \dots \} | > \epsilon\} \\ = \int m\{(\xi_1, \dots, \xi_n) \mid |S_{\xi_n} \{ \dots \{S_{\xi_1} \bar{x}_0\} \dots \} \\ - S_{\xi_n} \{ \dots \{S_{\xi_1} \bar{z}_0\} \dots \} | > \epsilon\} d\mu(\bar{x}_0) \rightarrow 0 \end{aligned}$$

as $n \rightarrow \infty$ so that

$$(43) \quad x_n - S_{\xi_n} \{ \dots \{S_{\xi_1} \bar{z}_0\} \dots \} \rightarrow 0$$

in probability as $n \rightarrow \infty$. On the other hand

$$(44) \quad \begin{aligned} E|E(x_n | \xi_n, \dots, \xi_1) - S_{\xi_n} \{ \dots \{S_{\xi_1} \bar{z}_0\} \dots \}| \\ = E|E(S_{\xi_n} \{ \dots \{S_{\xi_1} \bar{x}_0\} \dots \} - S_{\xi_n} \{ \dots \{S_{\xi_1} \bar{z}_0\} \dots \} | \xi_n, \dots, \xi_1)| \\ \leq E[E(|S_{\xi_n} \{ \dots \{S_{\xi_1} \bar{x}_0\} \dots \} - S_{\xi_n} \{ \dots \{S_{\xi_1} \bar{z}_0\} \dots \} | \xi_n, \dots, \xi_1)] \\ = E|x_n - S_{\xi_n} \{ \dots \{S_{\xi_1} \bar{z}_0\} \dots \}| \rightarrow 0. \end{aligned}$$

Therefore $x_n - E(x_n | \xi_n, \dots, \xi_1) \rightarrow 0$ as $n \rightarrow \infty$ in probability.

Now assume that $x_n - E(x_n | \xi_n, \dots, \xi_1) \rightarrow 0$ in probability as $n \rightarrow \infty$ but that (40) is not satisfied. Given a sufficiently small $\epsilon > 0$ there are then sets M_1, M_2 with $\mu(M_1), \mu(M_2) > \delta(\epsilon) > 0$ such that for $\bar{x}_0 \in M_1, \bar{x}'_0 \in M_2$

(45) $m\{(\xi_1, \dots, \xi_n) \mid |S_{\xi_n}\{\dots\{S_{\xi_1}\bar{x}_0\}\dots\} - S_{\xi_n}\{\dots\{S_{\xi_1}\bar{x}'_0\}\dots\}| > \epsilon\} > \delta(\epsilon)$
 for an infinite number of values of n . However

$$(46) \quad x_n - E(x_n \mid \xi_n, \dots, \xi_1) \rightarrow 0$$

implies that for almost all \bar{x}_0 (μ measure)

$$(47) \quad S_{\xi_n}\{\dots\{S_{\xi_1}\bar{x}_0\}\dots\} - E(S_{\xi_n}\{\dots\{S_{\xi_1}\bar{x}_0\}\dots\} \mid \xi_n, \dots, \xi_1) \rightarrow 0$$

in ξ_1, \dots, ξ_n measure, a contradiction.

What results exist on these two problems are rather scattered. The most coherent set of results are those on Markov chains. Let us therefore consider problem 1 for Markov chains. Let the transition probability matrix of the chain be $P = (p_{ij})$. It has already been remarked that part (a) of problem 1 has an affirmative answer if and only if the conditional distribution functions are equivalent for almost every past. This means that the row probability distributions in the transition matrix must be the same except for permutation. Let the positive probability masses common to the distributions $\{p_{ij}, j = 1, 2, \dots\}$ be $q_1 \geq q_2 \geq \dots > 0$ in order of magnitude, $\sum q_i = 1$. We call such Markov chains uniform chains. When the q_i are all distinct one can characterize the random variable ξ_n simply. *If $\{x_n\}$ is a stationary uniform Markov chain with the q_i distinct, ξ_n is uniquely determined (up to a one-one transformation modulo sets of measure zero) and is of the form*

$$(48) \quad \xi_n = \xi(x_n, x_{n-1}) = k \quad \text{for } (x_n, x_{n-1}) \text{ such that } p_{x_{n-1}, x_n} = q_k.$$

This is implied by the following argument. Write $\xi_n = \xi(x_n, x_{n-1}, \dots)$. Now ξ must take m distinct values (where $m \leq \infty$ is the number of distinct q) with probability one. Label the values 1, 2, \dots for convenience. Let the inverse function of ξ as a function of x_n for fixed x_{n-1}, x_{n-2}, \dots be $\eta(\xi, x_{n-1}, x_{n-2}, \dots)$. Since ξ_n is independent of x_{n-1}, x_{n-2}, \dots

$$(49) \quad P\{\xi_n = j \mid x_{n-1} = i_1, x_{n-2} = i_2, \dots\} = P\{\xi_n = j\} = \rho_{i_1, \eta(j, i_1, i_2, \dots)}.$$

Since the q_i are distinct, the function $\eta(j, i_1, i_2, \dots)$ for fixed i_1, j must be independent of i_2, i_3, \dots . Therefore $\eta(j, i_1, i_2, \dots) = \eta(j, i_1)$ so that $\xi(x_n, x_{n-1}, \dots) = \xi(x_n, x_{n-1})$. The function ξ must be constant on the set

$$(50) \quad \{(i, j) \mid p_{ij} = q_k\}$$

and for convenience we set $\xi = k$ on this set.

The remarks just made indicate that the answer to part (a) of problem 1 is rarely in the affirmative in the case of Markov chains even though the chain is purely nondeterministic. Nonetheless it is of some interest to find out *when part (b) of problem 1 is satisfied for stationary uniform Markov chains with the q_i distinct*. We have already noted that processes for which problem 1 (or 2) has a positive answer must be purely nondeterministic. *In the case of Markov chains this means that the set of all states must be an irreducible closed set of persistent states*. An interesting necessary and sufficient condition for a positive answer to problem 1 can be given in algebraic terms. Let M_k be the matrix

$$(51) \quad M_k = \{e_{ij}(k)\}$$

where

$$(52) \quad e_{ij}(k) = \begin{cases} 1 & \text{if } p_{ij} = q_k \\ 0 & \text{otherwise.} \end{cases}$$

There are m such distinct matrices M_k . Each such matrix M_k has precisely one element equal to one in each row and all other elements zero. Notice that every product of a finite number of the M_k also has these properties. No two distinct M_k have a one in the same location. The matrices $\{M_k\}$ correspond to a family of mappings of the set of states of the process into itself induced by the transition probability matrix P . First consider the semigroup generated (by finite operations) from the $\{M_k\}$. Notice that the matrices of this semigroup are of the same type, that is, have precisely one element equal to one in each row and all the rest zero. Complete this semigroup by adding to it all limit matrices. Call this semigroup the *completed semigroup generated by $\{M_k\}$* . We shall say that this semigroup is *point collapsing* if it has as an element a matrix with all the elements in some column one. A simple application of the theorem of this section indicates that a *purely nondeterministic stationary uniform Markov chain with distinct q_i has a representation of the type indicated in problem 1 if and only if the completed semigroup generated by the $\{M_k\}$ is point collapsing*. Since the proof is almost immediate we do not go through the details.

It is easy to give examples of purely nondeterministic Markov chains for which 1(a) is satisfied and 1(b) is not. The simplest such example is the Markov chain with the two states 1, 2 the stationary instantaneous distribution

$$(53) \quad P\{x_n = 1\} = P\{x_n = 2\} = \frac{1}{2}$$

and the transition probability matrix

$$(54) \quad P = \begin{pmatrix} p & q \\ q & p \end{pmatrix}, \quad 0 < p, q = 1 - p < 1, p \neq q.$$

Here

$$(55) \quad M_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad M_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

so that

$$(56) \quad \xi_n = \xi(x_n, x_{n-1}) = \begin{cases} 1 & \text{if } (x_{n-1}, x_n) = (1, 1), (2, 2) \\ 2 & \text{otherwise.} \end{cases}$$

Notice that x_n is not a function of ξ_n, ξ_{n-1}, \dots since $E(x_n | \xi_n, \xi_{n-1}, \dots) = Ex_n$.

It is generally a tedious computation even in the case of a finite state uniform Markov chain to find out whether the completed semigroup generated by the $\{M_k\}$ is point collapsing. One might hope for a simple criterion that would provide either a necessary or a sufficient condition for a positive answer to problem 1 in terms perhaps of the invariant vector (stationary instantaneous distribution) of the Markov chain. In the case of the simple example given

above, we note that the stationary distribution is uniform. Perhaps this would provide such a criterion. We give two simple examples to show that a satisfactory criterion must be somewhat more elaborate than a simple statement about uniformity of the invariant distribution. Consider a Markov chain with transition probability matrix

$$(57) \quad P = \begin{pmatrix} q_1 & q_2 & q_3 & q_4 \\ q_1 & q_2 & q_4 & q_3 \\ q_2 & q_1 & q_3 & q_4 \\ q_2 & q_1 & q_4 & q_3 \end{pmatrix}$$

with the q_i positive and distinct and $q_1 + q_2 = q_3 + q_4 = 1/2$. The invariant distribution is clearly uniform. Nonetheless, a simple computation indicates that the semigroup generated by the $\{M_k\}$ is point collapsing so that the answer to problem 1 is positive. A second example is provided by the Markov chain with transition matrix

$$(58) \quad P = \begin{pmatrix} q_1 & q_3 & q_2 & q_4 \\ q_1 & q_3 & q_2 & q_4 \\ q_2 & q_4 & q_3 & q_1 \\ q_2 & q_4 & q_3 & q_1 \end{pmatrix}$$

with the q_i positive and distinct. If the q_i are appropriately chosen, the invariant distribution will have distinct elements. Nonetheless the answer to problem 1 is negative since the semigroup generated by the $\{M_k\}$ is not point collapsing.

For reasons of clarity it is worthwhile seeing what one can say about the problem in terms of the concepts associated with the coding problem of information theory, even though it is trivial. Let us consider the capacity of the channel $\xi = \xi(x_n, x_{n-1})$ assuming that we are dealing with finite state processes (let us say m states). Since the channel is a noiseless channel with finite memory and essentially one-one given knowledge of x_{n-1} , it follows that the average entropy of an input, the input and output jointly, and the output alone are all the same. But this implies that the channel capacity is the maximum of average entropies for all possible inputs. Thus, the channel capacity is $m \log m$. Nonetheless, as we have noted there are inputs $\{x_n\}$ with average entropy smaller than $m \log m$ that cannot be reconstructed from the output. On reflection, the reason is clear. Here we do not allow coding. The channel ξ is determined so as to generate independent random variables from the input where the input is to be directly transmitted without coding. Given ξ_n, \dots, ξ_{n-k} the knowledge lacking to reproduce x_n is that of x_{n-k-1} . For some processes we recover this information as $k \rightarrow \infty$ and for others we don't. Problem 1 is that of determining the processes for which this information is recoverable.

The second problem has a more natural answer in the case of Markov chains. In fact, in [9] it is shown that the purely nondeterministic stationary Markov chains are precisely the Markov chains for which a representation of the form (19) is possible. In the case of Markov chains this is equivalent to a positive

answer to problem 2. Clearly this is the best one can hope for. It is an interesting and open question as to whether the same is true of all purely nondeterministic stationary processes with respect to problem 2. Essentially the proof proceeds by first passing the Markov chain $\{x_n\}$ through a channel with noise and without memory so as to enlarge the probability space. The output of this channel is a Markov process $\{y_n\}$ from which one can immediately reproduce $\{x_n\}$. A deterministic channel ξ is then determined in terms of $\{y_n\}$. This channel manufactures independent random variables and it is in terms of this output that $\{x_n\}$ is constructed.

In obtaining the results cited above, the simplicity of the state space of the process (the fact that it is countable) plays an important role. In the case of a Markov process with a conditional probability distribution $F(a_n|a_{n-1})$ continuous in a_n for almost every past, the state space is continuous and one can immediately construct a noiseless channel satisfying part (a) of problem 1 by setting

$$(59) \quad \xi_n = \xi(x_n, x_{n-1}) = F(x_n|x_{n-1}).$$

Here ξ_n is independent of the past x_{n-1}, x_{n-2}, \dots and yet ξ_n and x_{n-1} determine x_n . Nonetheless, even if $\{x_n\}$ is purely nondeterministic, it is generally not true that problem 1 has an affirmative answer for this channel. In fact there may be some other channel $\bar{\xi}$ satisfying 1(a) for which problem 1 has an affirmative answer even though the answer is negative for ξ (see the example on page 680 of [8]). Therefore, one must generally take care in choosing the proper ξ . A discussion of some scattered results in the case of continuous state space stationary processes can be found in [8].

I would like to thank D. Slepian and S. P. Lloyd for their helpful comments.

REFERENCES

- [1] H. ANZAI, "A remark on spectral measures of the flow of Brownian motion," *Osaka Math. J.*, Vol. 1 (1949), pp. 95-97.
- [2] J. R. BLUM and M. ROSENBLATT, "A class of stationary processes and a central limit theorem," *Duke Math. J.*, Vol. 24 (1957), pp. 73-78.
- [3] J. L. DOOB and R. A. LEIBLER, "On the spectral analysis of a certain transformation," *Amer. J. Math.*, Vol. 65 (1943), pp. 263-272.
- [4] K. ITÔ, "Spectral type of the shift transformation of differential processes with stationary increments," *Trans. Amer. Math. Soc.*, Vol. 81 (1956), pp. 253-263.
- [5] S. KAKUTANI, "Determination of the spectrum of the flow of Brownian motion," *Proc. Nat. Acad. Sci. U.S.A.*, Vol. 36 (1950), pp. 319-323.
- [6] G. MARUYAMA, "The harmonic analysis of stationary stochastic processes," *Mem. Fac. Sci. Kyushu Univ., Ser. A*, Vol. 4 (1949), pp. 45-106.
- [7] M. ROSENBLATT, "A central limit theorem and a mixing condition," *Proc. Nat. Acad. Sci. U.S.A.*, Vol. 42 (1956), pp. 43-47.
- [8] ———, "Stationary processes as shifts of functions of independent random variables," *J. Math. Mech.*, Vol. 8 (1959), pp. 665-682.
- [9] ———, "Stationary Markov chains and independent random variables," *J. Math. Mech.*, Vol. 9 (1960), pp. 945-950.
- [10] N. WIENER, "The homogeneous chaos," *Amer. J. Math.*, Vol. 60 (1939), pp. 897-936.