Stationary Processes as Shifts of Functions of Independent Random Variables

M. ROSENBLATT*

1. Summary. Let x_n, $n = 0, \pm 1, \pm 2, \cdots$, be a strictly stationary process. Two closely related problems are posed with respect to the structure of strictly stationary processes. In the first problem we ask whether one can construct a random variable $\xi_n = g(x_n, x_{n-1}, \cdots)$, a function of $x_n, x_{n-1}, \cdots$, that is independent of the past, that is, independent of $x_{n-1}, x_{n-2}, \cdots$. Such a sequence of random variables $\{\xi_n\}$ is a sequence of independent and identically distributed random variables. Further, given such a construction, is x_n a function of ξ_n, $\xi_{n-1}, \cdots$. Necessary and sufficient conditions for such a representation are obtained in the case where x_n is a finite state Markov chain with the positive transition probabilities in any row of the transition probability matrix $P = (p_{ij})$ of x_n distinct (Section 3). Such a representation is comparatively rare for a finite state Markov chain. In the second problem, the assumption that the independent and identically distributed ξ_n's be functions of $x_n, x_{n-1}, \cdots$ is removed. We ask whether for some such family $\{\xi_n\}$ there is a process $\{y_n\}$, $y_n = g(\xi_n, \xi_{n-1}, \cdots)$, with the same probability structure as $\{x_n\}$. This is shown to be the case for every ergodic finite state Markov chain with nonperiodic states (Section 4). Sufficient conditions for such representations in the case of a general strictly stationary process are obtained in Section 5.

2. Introduction. There are problems on the representation of strictly stationary processes that are motivated by certain results in the linear prediction problem and remarks of P. Lévy.

In the linear prediction problem one considers a weakly stationary process x_n, $n = 0, \pm 1, \cdots$, that is, a process x_n with

$$(1) \qquad \mathrm{E}x_n^2 < \infty, \qquad \mathrm{E}x_n x_m = r_{n-m}, \qquad \mathrm{E}x_n \equiv 0.$$

Under certain conditions on the spectrum of the process (the spectral distribution function $F(\lambda)$ absolutely continuous with $\log F'(\lambda)$ integrable), the process x_n can be represented as a one-sided weighted moving average of orthonormal random variables ξ_n defined on the present and past of the process. More explicitly,

---

ξ_n is an element of the closed linear manifold (closure in mean square) generated by x_m, $m \leq n$, with

(2) $$\mathrm{E}\xi_n \equiv 0, \qquad \mathrm{E}\xi_n\xi_m = \delta_{n,m}$$

and x_n has a representation

(3) $$x_n = \sum_{k=0}^{\infty} a_k\xi_{n-k}, \qquad \sum a_k^2 < \infty$$

(see [1]). This suggests a corresponding representation problem with orthogonality replaced by independence and linear relationship by functional dependence.

Before going on to present the problem, it should be noted that any statement made in the paper is to be understood modulo sets of measure zero, where the measure is the probability measure implicit in the problem at hand. Thus if $g(\cdot)$ is said to be measurable with respect to a Borel field of sets $\mathcal{B}$ we understand that it is equal almost everywhere to a function strictly measurable with respect to $\mathcal{B}$. Further, two Borel fields $\mathcal{B}$, $\mathcal{B}'$ will be said to be the same if given any set $B \, \varepsilon \, \mathcal{B}$ there is a set $B' \, \varepsilon \, \mathcal{B}'$ with the measure of the symmetric difference of B and B' zero, and if $\mathcal{B}'$ has the same property relative to $\mathcal{B}$. Clearly the two Borel fields are not the same strictly, but they are the same modulo sets of measure zero and that is of interest to us.

Let x_n, $n = 0, \pm 1, \cdots$, be a strictly stationary process with T the corresponding one-step shift operator. $\mathcal{B}_n$ is the Borel field of sets generated by x_n, x_{n-1}, $\cdots$.

Problem 1. (a) *When can one find a random variable ξ_0 measurable with respect to $\mathcal{B}_0$, independent of x_{-1}, x_{-2}, $\cdots$, and such that $\mathcal{B}_0 = \mathcal{B}_{-1} \times \mathcal{C}_0$ (product field) where $\mathcal{C}_0$ is the Borel field generated by ξ_0?*

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

A second problem is suggested by some remarks by P. Lévy [5]. Let $x_1, x_2, \cdots$ be a sequence of random variables. Under rather broad conditions P. Lévy has shown that one can find a sequence of independent, uniformly distributed random variables $\xi_1, \xi_2, \cdots$ and Borel functions $y_1 = g_1(\xi_1), y_2 = g_2(\xi_2, \xi_1), \cdots$ of $\xi_1, \xi_2, \cdots$ such that $y_1, y_2, \cdots$ have the same probability structure as $x_1, x_2, \cdots$.

Problem 2. Let x_n, $n = 0, \pm 1, \cdots$ be a strictly stationary process. *Can one find a sequence of independent, uniformly distributed (on $[0, 1]$) random variables ξ_n, $n = 0, \pm 1, \cdots$, and a Borel function $g(\xi) = g(\xi_0, \xi_{-1}, \cdots)$, $\xi = (\cdots, \xi_{-1}, \xi_0, \xi_1, \cdots)$, such that*

(4) $$y_n = g(T^n\xi), \qquad n = 0, \pm 1, \pm 2, \cdots,$$

has the same probability structure as x_n, $n = 0, \pm 1, \cdots$? Here T is the shift operator, that is,

(5) $T\xi = T(\cdots, \xi_{-1}, \xi_0, \xi_1, \cdots) = (\cdots, \xi_0, \xi_1, \xi_2, \cdots).$

Notice that g is a one-sided function of the sequence of random variables ξ_n .

If problem 1 is soluble, problem 2 is soluble. The converse is, however, not true. Notice that the solutions to these problems have nothing to do with the existence or nonexistence of moments of the process considered. They are problems about the Borel fields $\mathcal{B}_n$ generated by the process and the probability measure of the process on these fields. *Let $\cap \mathcal{B}_n$ be the Borel field of sets common to all the Borel fields $\mathcal{B}_n$, $n = 0, \pm 1, \cdots$.*

Lemma 1. *If problem 2 is soluble, the only functions measurable with respect to $\cap \mathcal{B}_n$ are the constant functions.*

Consider any random variable z, $E|z| < \infty$, measurable with respect to the Borel field generated by a stationary process $\{y_n\}$ for which problem 2 is soluble. We can assume that the random variables y_n are bounded. Just consider any monotone one-one function $h(\cdot)$ mapping $(-\infty, \infty)$ into $(0, 1)$. The new process $\{\bar{y}_n\}$, $\bar{y}_n = h(y_n)$, is bounded and has the same Borel fields $\mathcal{B}_n$ as $\{y_n\}$.

Since the random variables y_n are bounded, given any fixed $\epsilon > 0$ we can find a polynomial $p(y_n, \cdots, y_{n-k})$ in an appropriately large but finite number of random variables $y_n, \cdots, y_{n-k}$ such that

(6) $E|z - p| < \epsilon.$

Since the y_n's are one-sided functions of the ξ_n's (independent and uniformly distributed [0, 1]), there is a Borel function $q(\xi_n, \cdots, \xi_{n-j})$ in $\xi_n, \cdots, \xi_{n-j}$ for j sufficiently large such that

(7) $E|q - p| < \epsilon.$

But then

(8) $E|z - q| < 2\epsilon.$

Now

(9) $E(q|\cap\mathcal{B}_n) = E(q|\mathcal{B}_m) = E(q|\xi_m, \xi_{m-1}, \cdots) = Eq$

for $m < n - j$. But $E(z \mid \mathcal{B}_m) \to E(z \mid \cap \mathcal{B}_n)$ as $m \to -\infty$ in absolute mean by a martingale convergence theorem [1], [6]. Further,

(10) $E|E(z|\mathcal{B}_m) - E(q|\mathcal{B}_m)| \leqq E|z - q| < 2\epsilon.$

This implies that

(11) $E|E(z|\cap\mathcal{B}_n) - E(z)| < 4\epsilon.$

Since this is true for every $\epsilon > 0$, $E(z \mid \cap \mathcal{B}_n) = E(z)$ with probability one. Now consider any function w measurable with respect to $\cap \mathcal{B}_n$. We can always assume that $E|w| < \infty$, since otherwise one could find a one-one monotone function $h(\cdot)$ such that $z = h(w)$ has finite first moment. If z is a constant

almost everywhere so is w. But clearly z is a constant almost everywhere since $E(z \mid \cap \mathfrak{B}_n) = Ez$. The proof is complete.

It is an open question as to whether the converse of Lemma 1 is true. We shall look at this question in a few special but interesting cases. The necessary condition for the representation of problem 2 obtained in Lemma 1 is, of course, a necessary condition for the representation of problem 1. *Let* $\{x_n\}$ *be a strictly stationary process with the only functions measurable with respect to* $\cap$ $\mathfrak{B}_n$ *the constant functions.* It is natural to *call such processes purely nondeterministic processes.*

From this point on *assume that the conditional probability distribution function of the process* $\{x_n\}$,

$$(12) \qquad F(a_n \mid a_{n-1}, a_{n-2}, \cdots) = P[x_n \leqq a_n \mid x_{n-1} = a_{n-1}, x_{n-2} = a_{n-2}, \cdots],$$

is a Borel measurable function of the real variables a_n, a_{n-1}, $\cdots$. Given this weak assumption, we shall look for conditions on the conditional probability distribution that imply that part (a) of problem 1 is satisfied. The following convenient notion of equivalence of two distribution functions F, G is introduced. *F and G are equivalent if there is a one-one function mapping the jumps of F onto the jumps of G and preserving the size of the jumps.* Notice that any two continuous distribution functions F, G are equivalent according to this definition.

Lemma 2. *There is a random variable* ξ_n *satisfying part* (a) *of problem 1 if and only if the distribution functions* $F(a_n \mid a_{n-1}, a_{n-2}, \cdots)$ *(as functions of* a_n*) are equivalent for almost all* a_{n-1}, a_{n-2}, $\cdots$ *(with respect to the probability measure of the process* $\{x_n\}$*).*

Since ξ_n is independent of x_{n-1}, x_{n-2}, $\cdots$ the conditional probability distribution of ξ_n given x_{n-1}, x_{n-2}, $\cdots$ is independent of x_{n-1}, x_{n-2}, $\cdots$. Moreover, since $\mathfrak{B}_n = \mathfrak{B}_{n-1} \times \mathfrak{A}_n$ the conditional probability distribution of ξ_n given x_{n-1}, x_{n-2}, $\cdots$ must with probability one be equivalent to the conditional probability distribution of x_n given x_{n-1}, x_{n-2}, $\cdots$. The probability distribution functions $F(a_n \mid a_{n-1}, \cdots)$ must therefore be equivalent for almost all a_{n-1}, a_{n-2}, $\cdots$.

Conversely, assuming the equivalence of the distribution functions $F(a_n \mid a_{n-1}, a_{n-2}, \cdots)$ for almost all a_{n-1}, a_{n-2}, $\cdots$ we must show how to construct a random variable ξ_n with the desired properties. If the distribution functions $F(a_n \mid a_{n-1}, \cdots)$ are continuous for almost all a_{n-1}, a_{n-2}, $\cdots$ there is no difficulty. Simply set

$$(13) \qquad \xi_n = F(x_n \mid x_{n-1}, x_{n-2}, \cdots),$$

and an argument of P. Lévy [5] indicates that ξ_n is uniformly distributed $[0, 1]$ and independent of x_{n-1}, x_{n-2}, $\cdots$. Further,

$$(14) \qquad x_n = F^{-1}(\xi_n \mid x_{n-1}, x_{n-2}, \cdots)$$

with probability one. Let us therefore consider what happens if the distribution

functions $F(a_n \mid a_{n-1}, a_{n-2}, \cdots)$ have with probability one a discrete part. Let the jumps in the discrete part, ordered in terms of magnitude, be $p_1 \geqq p_2 \geqq \cdots$. For convenience assume there is strict inequality among the p_i. An obvious elaboration of the argument given in the case of strict inequality holds in the more general context with equality allowed. The magnitudes of the jumps $p_1, p_2, \cdots$ with probability one do not depend on $a_{n-1}, a_{n-2}, \cdots$ because of the equivalence of the conditional probability distributions. Let A_j be the set of points

$$(15) \qquad A_j = \{(a_n, a_{n-1}, \cdots) | \Delta F(a_n | a_{n-1}, \cdots) = p_j\}$$

where ΔF denotes the magnitude of the jump (if there is any) of the conditional distribution function at the argument indicated. We associate a set of integers $S((a_n, a_{n-1}, \cdots))$ with every point $(a_n, a_{n-1}, \cdots) \notin \cup A_j$. The integer $j \in S((a_n, a_{n-1}, \cdots))$ if there is a set A_j and a point $(a_n', a_{n-1}, a_{n-2}, \cdots) \in A_j$ with $a_n' < a_n$. Of course, the set $S((a_n, a_{n-1}, \cdots))$ may be the empty set. Set

$$(16) \qquad \xi_n = \begin{cases} j & \text{if } (x_n, x_{n-1}, \cdots) \in A_j \\ F(x_n | x_{n-1}, \cdots) - \sum_{i \in S((x_n, x_{n-1}, \cdots))} p_i & \text{if } (x_n, x_{n-1}, \cdots) \notin \cup A_j. \end{cases}$$

Then ξ_n is independent of $x_{n-1}, x_{n-2}, \cdots$. The random variable ξ_n is uniformly distributed on $(0, 1 - \sum p_i)$ with density function one and assumes the integral value j with probability p_j. Further, the construction indicates that x_n is given with probability one by the knowledge of $\xi_n, x_{n-1}, x_{n-2}, \cdots$.

There is a case in which part (a) of problem 1 is satisfied trivially and yet there is no hope of having part (b) satisfied. Suppose $F(a_n \mid a_{n-1}, \cdots)$ is for almost all $a_{n-1}, a_{n-2}, \cdots$ a pure jump function with a jump of size one. Then x_n is determined with probability one given $x_{n-1}, x_{n-2}, \cdots$ so that $\mathcal{B}_n = \mathcal{B}_{n-1} = \cdots$ for all n. It is natural to *call such a process a purely deterministic process.* Such a process will satisfy the necessary condition of Lemma 1 only if the random variables x_n are a constant c with probability one. Since we are interested in problems 1 and 2, the purely deterministic processes are to be avoided.

3. Markov Chains and Problem 1.

We shall not be able to obtain very broad results in the general context spoken of in Section 2. The discussion in the general context of strictly stationary processes is relegated to a later section (Section 5) because more detailed and enlightening results can be obtained for strictly stationary Markov chains with a finite number of states.

Let $\{x_n\}$ be a strictly stationary Markov chain with transition probability matrix

$$(17) \qquad \mathrm{P} = (p_{ij}), \qquad p_{ij} = \mathrm{P}[x_{n+1} = j | x_n = i], \qquad i, j = 1, 2, \cdots,$$

and stationary vector of probabilities

$$(18) \qquad p_i = \mathrm{P}[x_n = i] > 0.$$

It is already clear from Lemma 2 that we shall not be able to construct a random variable ξ_n measurable with respect to $\mathfrak{G}_n$ satisfying the conditions of part (a) of problem 1 unless the probability distributions $\{p_{ij} \; ; j = 1, 2, \cdots\}$ indexed by i are equivalent. An alternative way of stating this is as follows. Given any i, i' there is a permutation $M(i, i')$ of the integers j such that

$$(19) \qquad p_{ij} = p_{i', M(i, i')j}$$

for all j. Such Markov chains are admittedly rather special. Nonetheless they are interesting and give some insight into problem 1. We shall *call such Markov chains uniform chains* because the row probability distributions are the same except for permutation. Let the positive probability masses common to the distributions $\{p_{ij}, j = 1, 2, \cdots\}$ be $q_1 \geqq q_2 \geqq \cdots > 0$ in order of magnitude, $\sum q_i = 1$. When the q_i's are all distinct one can characterize the random variable ξ_n simply.

Lemma 3. *If x_n is a stationary uniform Markov chain with q_i's distinct, ξ_n is uniquely determined (up to a one-one transformation modulo sets of measure zero) and is of the form*

$$(20) \qquad \xi_n = \xi(x_n, x_{n-1}) = k \quad for \quad (x_n, x_{n-1}) \quad such \; that \quad p_{x_{n-1}, x_n} = q_k .$$

Without too much abuse of notation we can write $\xi_n = \xi(x_n, x_{n-1}, \cdots)$. It is already clear that ξ must take m distinct values ($m \leqq \infty$ the number of distinct q's) with probability one, and we label the values $1, 2, \cdots$ for convenience. Let the inverse function of ξ, as a function of x_n for fixed $x_{n-1}, x_{n-2}, \cdots$, be $\eta(\xi, x_{n-1}, x_{n-2}, \cdots)$. Since ξ_n is independent of $x_{n-1}, x_{n-2}, \cdots$

$$(21) \qquad P[\xi_n = j | x_{n-1} = i_1, x_{n-2} = i_2, \cdots]$$
$$= P[\xi_n = j] = p_{i_1, \eta(j, i_1, i_2, \cdots)} \qquad > 0.$$

But the q_i's are distinct, and therefore for fixed i_1, j the function $\eta(j, i_1, i_2, \cdots)$ must be independent of i_2, i_3, $\cdots$. Thus $\eta(j, i_1, i_2, \cdots) = \eta(j, i_1)$ and hence $\xi(x_n, x_{n-1}, \cdots) = \xi(x_n, x_{n-1})$. It is clear that ξ must be constant on the set

$$(22) \qquad \{(i, j) | p_{ij} = q_k\},$$

and for convenience we set $\xi = k$ on this set. Any other ξ satisfying the conditions of part (a) of problem 1 must be a one-one function of this ξ.

The random variable ξ_n was seen to be essentially uniquely determined when the q_i's are distinct. It is easy to see that it is not uniquely determined if the q_i's are not distinct. For example, if the transition probability matrix is

$$(23) \qquad P = \begin{vmatrix} a & a & b \\ a & b & a \\ b & a & a \end{vmatrix}, \qquad a, b > 0, \qquad a \neq b,$$

we can take

$$(24) \qquad \xi(x_n , x_{n-1}) = \begin{cases} 0 & \text{if} \quad (x_n , x_{n-1}) = (0,0),\, (0,1),\, (1,2) \\ 1 & \text{if} \quad (x_n , x_{n-1}) = (1,0),\, (2,1),\, (2,2) \\ 2 & \text{otherwise} \end{cases}$$

or, alternatively,

$$(25) \qquad \xi(x_n , x_{n-1}) = \begin{cases} 0 & \text{if} \quad (x_n , x_{n-1}) = (0,0),\, (2,1),\, (2,2) \\ 1 & \text{if} \quad (x_n , x_{n-1}) = (1,0),\, (0,1),\, (1,2) \\ 2 & \text{otherwise.} \end{cases}$$

In fact, ξ need no longer be a function of only x_n, x_{n-1}, for one can take

$$(26) \qquad \xi(x_n , x_{n-1} , x_{n-2}) = \begin{cases} 0 & \text{if} \quad (x_n , x_{n-1} , x_{n-2}) = (0,0,0),\, (1,0,1), \\ & \qquad (0,0,2),\, (0,1,0),\, (2,1,1),\, (0,1,2), \\ & \qquad (1,2,0),\, (2,2,1),\, (1,2,2) \\ 1 & \text{if} \quad (x_n , x_{n-1} , x_{n-2}) = (1,0,0),\, (0,0,1), \\ & \qquad (1,0,2),\, (2,1,0),\, (0,1,1),\, (2,1,2), \\ & \qquad (2,2,0),\, (1,2,1),\, (2,2,2) \\ 2 & \text{otherwise.} \end{cases}$$

Let us restrict ourselves to uniform Markov chains with distinct q_i's and try to find out under what circumstances the answer to problem 1 is in the affirmative. We have already seen that a positive answer to problem 2, and hence to problem 1 (making use of Lemma 1), can be given only if all the functions measurable with respect to $\cap\; \mathfrak{G}_n$ are the constant functions. This means that *the set of states of the Markov chain must be an irreducible closed set of ergodic* (that is, nonperiodic) *states* [2]. 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

$$(27) \qquad M_k = \{e_{ij}(k)\}$$

where

$$(28) \qquad e_{ij}(k) = \begin{cases} 1 & \text{if} \quad p_{ij} = q_k \\ 0 & \text{otherwise.} \end{cases}$$

Given the Markov chain $\{x_n\}$ there are m corresponding matrices M_k. Each matrix M_k has precisely one element equal to one in each row and all other elements zero. No two M_k's have a one in the same entry. Every product of a finite number of the M_k's also has these properties. The family of matrices $\{M_k\}$ corresponds to a family of mappings of the set of states of the process into the same set of states induced by the process $\{x_n\}$. Consider the semigroup

(under multiplication) generated by the matrices M_k . This semigroup may or may not contain a matrix with a column of ones. If it does, we *call the semigroup point collapsing.*

Theorem 1. *Let* $\{x_n\}$ *be a finite state stationary uniform Markov chain with* q_i's *distinct. Let* $\{M_k\}$ *be the family of mapping matrices generated by* $\{x_n\}$. *The answer to problem 1 is positive if and only if the semigroup generated by* $\{M_k\}$ *is point collapsing.*

First assume the semigroup generated by the family of matrices $\{M_k\}$ is point collapsing. There is then a finite product $M_{k_1} \cdots M_{k_s}$ of matrices of this family (with possible duplication) which has all the elements in the ith column one. Now the random variable ξ_n satisfying the conditions of part (a) of problem 1 is uniquely determined by Lemma 3 and is given by

$$(29) \qquad \xi_n = \xi(x_n , x_{n-1}) = k \quad \text{if} \quad e_{x_{n-1}, x_n}(k) = 1.$$

Consider an event A_i of the form

$$(30) \qquad A_i = \{\xi_i = k_1 , \cdots , \xi_{i+s} = k_s\}.$$

If A_i occurs, it is clear that $x_{i+s} = i$ with probability one. The events A_{n-s} , A_{n-2s} , $\cdots$ are independent, and hence by the lemma of Borel-Cantelli at least one of the events will occur with probability one. Thus at least one of the events A_i , $j \leqq n - s + 1$, must occur with probability one. Let T_k be the mapping induced by M_k , that is, $T_k i = j$ where j is the unique j such that $e_{ij}(k) = 1$. We now construct a function $g(\xi_n , \xi_{n-1} , \cdots)$ such that $x_n = g(\xi_n , \xi_{n-1} , \cdots)$ with probability one. Consider the sequences $(\cdots , \xi_{n-1} , \xi_n)$. Let S_i be the set of sequences $(\cdots , \xi_{n-1} , \xi_n)$ for which some A_i , $i \leqq n - s + 1$, occurs and A_i is the A_i with largest index, $i \leqq n - s + 1$, that occurs. The sets S_i , $j \leqq n - s + 1$, are disjoint and $P(\cup S_i) = 0$ ($\bar{S}$ denotes the complement of S) as already remarked above. Let

$$(31) \qquad g(\xi_n , \xi_{n-1} , \cdots) = \begin{cases} T_{\xi_n} \cdots T_{\xi_{i+s+1}} i & \text{if} \quad (\cdots , \xi_{n-1} , \xi_n) \, \varepsilon \, S_i \\ 0 & \text{otherwise.} \end{cases}$$

It is then clear that $x_n = g(\xi_n , \xi_{n-1} , \cdots)$ with probability one since $(\cdots , \xi_{n-1} , \xi_n) \, \varepsilon \, S_i$ for some $j \leqq n - s + 1$ with probability one. Thus the answer to problem 1 is positive.

Now, on the contrary, assume that the semigroup generated by the family of matrices $\{M_k\}$ is not point collapsing. Every matrix of the semigroup has a certain number of distinct columns containing a one as a column entry. The minimal number of distinct columns among all matrices of the semigroup is certainly greater than one (let us say $r > 1$) since the semigroup is not point collapsing. But this means that knowledge of all the ξ_i's, $i \leqq n$, can tell no more than that x_n takes one of r distinct values. The answer to problem 1 must therefore be negative.

One might hope for a positive answer to problem 1 if there is a random variable ξ_n satisfying part (a) of the problem and the necessary condition of Lemma 1 is satisfied. However, by making use of Theorem 1 a large number of processes satisfying these conditions for which the answer to problem 1 is negative can be constructed.

We give the simplest example exhibiting this interesting sort of pathology. Let $\{x_n\}$ be the strictly stationary Markov chain with the two states 1, 2, the stationary instantaneous distribution

$$(32) \qquad \qquad P[x_n = 1] = P[x_n = 2] = \tfrac{1}{2}$$

and transition probability matrix

$$P = \begin{bmatrix} p & q \\ q & p \end{bmatrix}, \qquad 0 < p, q = 1 - p < 1, \qquad p \neq q.$$

Now

$$(33) \qquad \qquad M_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad M_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

so that

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

is measurable with respect to $\mathfrak{B}_n$ and yet independent of x_{n-1}, x_{n-2}, $\cdots$. By Lemma 3 it is essentially the only random variable with these properties. Further, $\{x_n\}$ is purely nondeterministic. Nonetheless x_n is not a function of ξ_n, ξ_{n-1}, $\cdots$ since $E(x_n \mid \xi_n, \xi_{n-1}, \cdots) = Ex_n$.

One would like to give a neat characterization of the point collapsing semigroups generated by the matrices $\{M_k\}$ induced by a uniform Markov chain. This does not appear to be possible. However, one can easily sketch the details of a computational procedure that would determine whether or not such a semigroup is point collapsing. Let $N < \infty$ be the number of states of the finite state uniform Markov chain $\{x_n\}$. Let $\{T_k\}$ be the family of mappings of the integers $1, \cdots, N$ into the integers $1, \cdots, N$ determined by the matrices $\{M_k\}$. If the semigroup is point collapsing some T_k must map the N states onto $N_1 < N$ states. Given N_1 states, consider the ensemble of all distinct N_1-tuples (including the original one) that can be obtained from this N_1-tuple by applying a finite number of the T_k mappings. Some one of these N_1-tuples must be collapsed further onto $N_2 < N_1$ states by some T_k mapping if the semigroup is point collapsing. Continue in this manner until the original set of states is collapsed onto one state. If, at some stage in the construction, all the distinct N_j-tuples $(N_j > 1)$ obtained cannot be contracted, the semigroup is not point collapsing. If N is not small, this is admittedly a very tedious procedure.

4. Markov Chains and Problem 2. The answer to problem 1 is generally negative, even in the case of a finite state purely nondeterministic Markov chain. However, the answer to problem 2 for such Markov chains will be shown to be positive. The following theorem will be proved.

Theorem 2. *Let* $\{x_n\}$ *be a stationary purely nondeterministic Markov chain. Let there be an* $\epsilon > 0$ *and a finite number of states* M *such that*

$$(34) \qquad \sum_{j \,\epsilon\, M} p_{ij} = p_{i,M} > \epsilon$$

for all i. *A representation of the type mentioned in problem 2 holds for such a process.*

Notice that this result holds for all ergodic finite state stationary Markov chains with nonperiodic states. Actually Theorem 2 is proved by enriching the probability space of the $\{x_n\}$ process in an appropriate way.

Introduce the Markov process y_n, $1 < y_n < \infty$, with stationary probability density

$$(35) \qquad p(y|y') = p_{ij} \quad \text{if} \quad y' \,\epsilon\, (i, i+1], \qquad y \,\epsilon\, (j, j+1], \qquad i, j = 1, 2, \cdots,$$

and stationary instantaneous probability distribution

$$(36) \qquad P[y_n \,\epsilon\, A] = \sum p_i m(A \cap (i, i+1]),$$

where $p_i = P[x_n = i]$. We shall construct a process on the probability space of the y_n process with the same probability structure as that of the x_n process. In spite of a possible abuse of notation we shall call this new process the x_n process. The process x_n can be taken as $x_n = g(y_n)$ where $g(y) = j$ if $y \,\epsilon\, (j, j+1]$ (see [7]). The probability space of the y_n process is rich enough so that one can construct random variables $\alpha_n = G(y_n \mid y_{n-1})$ independent of y_{n-1}, y_{n-2}, $\cdots$ with α_n uniformly distributed on $(0, 1)$. $G(y \mid y')$ is a Borel function of y, y' and is so constructed that, for fixed y_{n-1}, $G(y_n \mid y_{n-1})$ is with probability one a one-one function of y_n. We can therefore write $y_n = G^{-1}(\alpha_n \mid y_{n-1})$. It will be convenient to introduce the associated family of transformations T_α, $0 < \alpha < 1$, where

$$(37) \qquad T_\alpha y = G^{-1}(\alpha|y).$$

Notice that it is equivalent to give either $G(y \mid y')$ or the corresponding family of transformations $\{T_\alpha\}$ due to the essentially one-one character of G. Such a family of random variables α_n can be introduced, for example, by taking $G(y|y') = F(y \mid y')$ where $F(y \mid y')$ is the conditional distribution function of y_n given y_{n-1}. Note that in this case T_α for each α, $0 < \alpha < 1$, maps each interval $(j, j+1]$ into a single point. There are other families of transformations T_α, associated with random variables α_n of the above form, that have this property. We shall construct a convenient family T_α for our purposes in the proof of Theorem 2.

By (34) for some fixed $\epsilon > 0$ there is a finite set of states which we can take as $1, \cdots, M$ (by relabeling the states if necessary) such that $\sum_{j=1}^{M} p_{ij} > \epsilon$

for all i. Since the Markov chain $\{x_n\}$ is irreducible and the states are nonperiodic, there must be two paths from state one to state one of positive probability with lengths that are relatively prime. Add all the states passed through in these paths to the states $1, \cdots, M$. Further, add all the other states passed through in a path of minimal length and positive probability from state i to state j, i, $j = 1, \cdots, M$. Call the collection of states included thus far S. We can then connect any two states i, j ε S by a path of positive probability containing only states of S. If the states of S are $N(S)$ in number it may be convenient to relabel the states of the process so that they appear as the states $1, 2, \cdots, N = N(S)$. Notice that the matrix of transition probabilities of the process $\{x_n\}$, as restricted to the states of S, is irreducible and aperiodic, that is, some sufficiently high power of the restricted matrix has all its elements positive.

Let S' be any subset of S and $N(S')$ the number of states in S'. *Call S' a contracting set if there is a set of states $S'' \subset S$ with $N(S'') < N(S')$ such that for each i ε S' there is a j ε S'' with $p_{ij} > 0$.* The set $S'' \subset S$ can be reached from $S' \subset S$ if for each i ε S' there is a j ε S'' with $p_{ij} > 0$. *S'' is accessible from S' if for some k there is a sequence of subsets (of S) $S_1 = S'$, S_2, $\cdots$, $S_k = S''$ such that S_j can be reached from S_{j-1}, $j = 2, \cdots, k$.* If $N(S') > 1$, $S' \subset S$, there is a set $S'' \subset S$ with $N(S'') < N(S')$ that is accessible from S'. For suppose there were not. Every m-tuple of states of S is accessible from every other m-tuple of states of S because of the irreducible and nonperiodic character of the transition matrix when restricted to the states of S. If there were no set of states S'' with the desired property no m-tuple with $m = N(S') > 1$ could be contracted. Take any two distinct states i, j ε S. Since they are in some m-tuple and no m-tuple is contracting, there is no state k ε S such that p_{ik}, $p_{jk} > 0$. This means that for any i ε S there is one and only one $j = j(i)$ such that $p_{ij} > 0$. But this implies that the matrix of transition probabilities, as restricted to the states of S, is periodic, a contradiction.

Start with the whole set of states $S_0 = S = \{1, 2, \cdots, N\}$. This set is contracting. Let $S_1 \subset S$ be a set of states with $N(S_1) < N(S)$ that is accessible from S_0 in a minimal number of steps. Let $S_2 \subset S$ be a set with $N(S_2) < N(S_1)$ that is accessible from S_1 in a minimal number of steps. Continue the construction until a set, say S_k, with one element is obtained. Let $M_1, \cdots, M_a$ be sets of states with $M_1 = S_0$, $M_2 = S_1$, $\cdots$, $M_{2+r_1} = S_2$, $\cdots$, $M_{2+r_1+r_2} = S_3$, $\cdots$, $M_a = S_k$ where $a = 2 + r_1 + \cdots + r_k$ and M_{j+1} can be reached from M_j. In other words, the M_j's are the sets of states passed through in going from S_0 to S_k. The number of steps in going from S_j to S_{j+1} can be at most $N^2 = N(S)^2$ in number (see [8]). Therefore the total number of steps from M_1 to M_a can be at most N^3. Let the minimal positive transition probability p_{ij}, i, j ε S, be δ (> 0). Given any state i ε M_j, let $l = f_j(i)$ be the state of lowest index in M_{j+1} for which $p_{il} > 0$. Further, given any state i ε S. there is a state $l = f(i)$ ε S such that $p_{il} > \epsilon/N$. Let $\delta' = \min(\epsilon/N, \delta)$.

Now consider the Markov process $\{y_n\}$, constructed at the beginning of our proof, and the Markov chain $x_n = g(y_n)$. We now set up random variables

$\alpha_n = G(y_n \mid y_{n-1})$, uniformly distributed and independent of y_{n-1}, y_{n-2}, $\cdots$ with the associated family of one-one transformations T_α satisfying conditions now to be specified. Consider disjoint subintervals (γ_i, δ_i), $i = 0, 1, \cdots, a$, of $(0, 1)$ with $0 < \delta_i - \gamma_i < \delta'/(N^3 + 1)$. The transformations T_α are to be set up (and can be set up) so that they satisfy the following conditions. Let

$$(38) \qquad \begin{aligned} T_\alpha(i, i+1] &\subset (f(i), f(i) + 1] \quad \text{if} \quad i \notin S, \\ T_\alpha(i, i+1] &\subset (1, N+1] \qquad \text{if} \quad i \in S \end{aligned}$$

when $\gamma_0 < \alpha < \delta_0$. Thus if $\gamma_0 < \alpha_n < \delta_0$ we know that $1 < y_n \leqq N + 1$ (or equivalently $x_n \in S$). Further, let

$$(39) \qquad T_\alpha(i, i+1] \subset (f_i(i), f_i(i) + 1]$$

for all $i \in M_j$ if $\gamma_j < \alpha < \delta_j$. Moreover, let the family of transformations T_α, $0 < \alpha < 1$, map each interval $(i, i+1]$, $i = 1, 2, \cdots$, into a single point. The point will generally depend on α. Now if the event A_j occurs,

$$(40) \qquad A_j = \{\gamma_0 < \alpha_j < \delta_0, \cdots, \gamma_a < \alpha_{j+a} < \delta_a, \gamma_0 < \alpha_{j+a+1} < \delta_0\},$$

we know with probability one the value of y_{j+a+1}. For the first $a + 1$ conditions imply that $i < y_{j+a} \leqq i + 1$ where i is the single state of M_a. Thus $y_{j+a+1} = T_{\alpha_{j+a+1}}(i, i+1]$. The random variables $\{\alpha_j\}$ are independent uniformly distributed random variables. Just as in the proof of Theorem 1, the Borel-Cantelli lemma implies that at least one of the events A_j, $j \leqq n - a - 1$, occurs with probability one. Let E_j be the set of sequences $(\cdots, \alpha_{n-1}, \alpha_n)$ for which some A_i, $i \leqq n - a - 1$, occurs and A_j is the A_i with largest index, $i \leqq n - a - 1$, that occurs. The events E_j are disjoint, and with probability one every sequence $(\cdots, \alpha_{n-1}, \alpha_n)$ lies in some E_j, $j \leqq n - a - 1$. Let

$$(41) \qquad h(\alpha_n, \alpha_{n-1}, \cdots) = \begin{cases} T_{\alpha_n} \cdots T_{\alpha_{j+a+1}}(i, i+1] & \text{if} \quad (\cdots, \alpha_{n-1}, \alpha_n) \in E_j \\ 0 & \text{otherwise.} \end{cases}$$

Then $y_n = h(\alpha_n, \alpha_{n-1}, \cdots)$ with probability one and hence

$$x_n = g(h(\alpha_n, \alpha_{n-1}, \cdots))$$

with probability one. The proof of Theorem 2 is complete.

5. Stochastic Processes. We now consider some conditions sufficient (but not necessary) for a positive answer to problem 1, and hence problem 2 in the case of a general strictly stationary stochastic process $\{x_n\}$. There are several conditions that will be introduced to avoid unduly tedious measure theoretic complications. We have already seen that there are difficulties in obtaining random variables ξ_n satisfying the conditions of part (a) of problem 1 if the conditional distribution function $F(a_n \mid a_{n-1}, \cdots)$ has a discrete component with positive probability. To avoid this assume that $F(a_n \mid a_{n-1}, \cdots)$ *is continuous in a_n with probability one.* Further, *let the random variables x_n of the process be*

bounded between zero and one. This last assumption is on restriction since it can always be brought about by considering a process $z_n = h(x_n)$ where h is a one-one monotone function mapping $(-\infty, \infty)$ into $(0, 1)$. Further, *let* $F(a_n \mid a_{n-1}, \cdots)$ *be a strictly increasing function of* a_n, $0 < a_n < 1$, *with probability one.* By Lemma 2 there is a random variable ξ_n satisfying part (a) of problem 1. Such a random variable is, in fact, given by

$$(42) \qquad \xi_n = F(x_n | x_{n-1}, x_{n-2}, \cdots).$$

This is not the only possible construction of such a random variable. The conditions to be specified will be enough to ensure that the process x_n is of the form

$$(43) \qquad x_n = g(\xi_n, \xi_{n-1}, \cdots);$$

in other words problem 1 has a positive answer with this particular construction (see formula (42)) of such a random variable. In a report KALLIANPUR & WIENER [3] discuss some work in this direction. The conjectured result is that a representation of the form (43) holds under the above conditions and under the assumption that $\{x_n\}$ is purely nondeterministic. The argument given implies that the conjectured representation can be established in terms of the specific random variables (42). We shall later give an example for which a representation of the form (43) is valid, but not in terms of the random variables (42). This indicates that generally one has to consider constructing random variables ξ_n satisfying part (a) of problem 1 which are not of the form (42).

Let μ be the measure on half infinite sequences $(a_n, a_{n-1}, \cdots)$, $0 < a_i < 1$, induced by the probability measure of the process $\{x_n\}$. Let T_α, $0 < \alpha < 1$, be the family of transformations of the set of sequences $(a_n, a_{n-1}, \cdots)$ into the real numbers a_{n+1}, $0 < a_{n+1} < 1$, induced by the conditional distribution function, that is,

$$(44) \qquad T_\alpha(a_n, a_{n-1}, \cdots) = F^{-1}(\alpha | a_n, a_{n-1}, \cdots).$$

Notice that

$$(45) \qquad x_n = T_{\xi_n}(x_{n-1}, x_{n-2}, \cdots)$$

with probability one where ξ_n is the uniformly distributed random variable

$$(46) \qquad \xi_n = F(x_n | x_{n-1}, \cdots)$$

independent of $x_{n-1}, x_{n-2}, \cdots$. If $x_n = g(\xi_n, \xi_{n-1}, \cdots)$ then clearly $x_n = E(x_n \mid \xi_n, \xi_{n-1}, \cdots)$ and hence

$$(47) \qquad x_n - E(x_n | \xi_n, \xi_{n-1}, \cdots, \xi_{n-k}) \to 0$$

in probability as $k \to \infty$ or equivalently

$$(48) \qquad E |x_n - E(x_n | \xi_n, \cdots, \xi_{n-k})| \to 0$$

because of the boundedness of the random variables x_n. Because of the stationarity of the $\{x_n\}$ process, condition (48) could equally well have been written

(49) $$\mathrm{E}\,|x_n - \mathrm{E}(x_n|\xi_n\,,\,\cdots\,,\xi_1)| \to 0$$

as $n \to \infty$. To avoid cumbersome formulas we introduce the following notation. Let

$$\bar{x}_m = (x_m\,,\,x_{m-1}\,,\,\cdots),$$

(50) $$T_{\xi_m}\bar{x}_{m-1} = T_{\xi_m}(x_{m-1}\,,\,x_{m-2}\,,\,\cdots),$$

$$\{T_{\xi_m}\bar{x}_{m-1}\} = (T_{\xi_m}\bar{x}_{m-1}\,,\,x_{m-1}\,,\,x_{m-2}\,,\,\cdots).$$

Theorem 3. *The condition*

(51) $$\mathrm{E}\,|x_n - \mathrm{E}(x_n|\xi_n\,,\,\cdots\,,\xi_1)| \to 0 \quad as \quad n \to \infty$$

is satisfied if and only if there is a set S of $\bar{y}_0$ points of μ measure one such that

(52) $$T_{\xi_n}\{T_{\xi_{n-1}}\{\cdots\{T_{\xi_1}\bar{y}_0\}\cdots\}\} - T_{\xi_n}\{T_{\xi_{n-1}}\{\cdots\{T_{\xi_1}\bar{y}_0'\}\cdots\}\} \to 0$$

in $\xi_1\,,\,\cdots\,,\,\xi_n\,,\,\cdots$ measure for all $\bar{y}_0\,,\,\bar{y}_0' \,\varepsilon\, S$ as $n \to \infty$.

Suppose there is a set S of μ measure one satisfying condition (52). Let $\bar{z}_0\,\varepsilon\, S$. Then for all $\bar{x}_0\,\varepsilon\, S$ the measure

$$m\{((\xi_1\,,\,\cdots\,,\xi_n)||T_{\xi_n}\{\cdots\{T_{\xi_1}\bar{x}_0\}\cdots\} - T_{\xi_n}\{\cdots\{T_{\xi_1}\bar{z}_0\}\cdots\}| > \epsilon)\} \to 0$$

as $n \to \infty$ for any fixed $\epsilon > 0$. But then

$$\mathrm{P}\{|x_n - T_{\xi_n}\{\cdots\{T_{\xi_1}\bar{z}_0\}\cdots\}| > \epsilon\} = \int m\{((\xi_1\,,\,\cdots\,,\xi_n)||T_{\xi_n}\{\cdots\{T_{\xi_1}\bar{x}_0\}\cdots\}$$

$$- T_{\xi_n}\{\cdots\{T_{\xi_1}\bar{z}_0\}\cdots\}| > \epsilon)\}\,d\mu(\bar{x}_0) \to 0$$

as $n \to \infty$ and hence

$$x_n - T_{\xi_n}\{\cdots\{T_{\xi_1}\bar{z}_0\}\cdots\} \to 0$$

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

$$\mathrm{E}\,|\mathrm{E}(x_n|\xi_n\,,\,\cdots\,,\xi_1) - T_{\xi_n}\{\cdots\{T_{\xi_1}\bar{z}_0\}\cdots\}|$$

$$= \mathrm{E}\,|\mathrm{E}(T_{\xi_n}\{\cdots\{T_{\xi_1}\bar{x}_0\}\cdots\} - T_{\xi_n}\{\cdots\{T_{\xi_1}\bar{z}_0\}\cdots\}|\xi_n\,,\,\cdots\,,\xi_1)|$$

$$\leqq \mathrm{E}(\mathrm{E}(|T_{\xi_n}\{\cdots\{T_{\xi_1}\bar{x}_0\}\cdots\} - T_{\xi_n}\{\cdots\{T_{\xi_1}\bar{z}_0\}\cdots\}||\xi_n\,,\,\cdots\,,\xi_1))$$

$$= \mathrm{E}\,|x_n - T_{\xi_n}\{\cdots\{T_{\xi_1}\bar{z}_0\}\cdots\}| \to 0.$$

But then $x_n - \mathrm{E}(x_n|\xi_n\,,\,\cdots\,,\,\xi_1) \to 0$ as $n \to \infty$ in probability.

Now suppose that $x_n - \mathrm{E}(x_n|\xi_n\,,\,\cdots\,,\,\xi_1) \to 0$ in probability as $n \to \infty$ and condition (52) is not satisfied. Then given a sufficiently small $\epsilon > 0$ there are sets $S_1\,,\,S_2$ with $\mu(S_1),\,\mu(S_2) > \delta(\epsilon) > 0$ such that for $\bar{x}_0\,\varepsilon\, S_1\,,\,\bar{x}_0'\,\varepsilon\, S_2$

$$m\{((\xi_1\,,\,\cdots\,,\xi_n)||T_{\xi_n}\{\cdots\{T_{\xi_1}\bar{x}_0\}\cdots\} - T_{\xi_n}\{\cdots\{T_{\xi_1}\bar{x}_0'\}\cdots\}| > \epsilon)\} > \delta(\epsilon)$$

for an infinite number of values of n. However,

$$x_n - \mathrm{E}(x_n | \xi_n , \cdots , \xi_1) \to 0$$

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

$$T_{\xi_n}\{\cdots\{T_{\xi_1}\bar{x}_0\}\cdots\} - \mathrm{E}(T_{\xi_n}\{\cdots\{T_{\xi_1}\bar{x}_0'\}\cdots\}|\xi_n , \cdots , \xi_1) \to 0$$

in $\xi_1 , \cdots , \xi_n , \cdots$ measure, a contradiction.

Corollary 1. *Let the family of transformations* T_α , $0 < \alpha < 1$, *be such that, given any two* $\bar{x}_0$, $\bar{x}_0'$ *with* $0 < x_{-j} , x_{-j}' < 1$,

$$(53) \qquad \mathrm{E}\,|T_{\xi_1}\bar{x}_0 - T_{\xi_1}\bar{x}_0'| \leqq \sum_{j=0}^{\infty} \beta_j \,|x_{-j} - x'_{-j}|$$

for some sequence of positive constants β_j *with* $\sum \beta_j < 1$. *Then* $\{x_n\}$ *has a representation of form* (43). *The expectation in formula* (53) *is understood to be an average with respect to the uniformly distributed random variable* ξ.

It is clear that

$$\mathrm{E}\,|T_{\xi_n}\{\cdots\{T_{\xi_1}\bar{x}_0\}\cdots\} - T_{\xi_n}\{\cdots\{T_{\xi_1}\bar{x}_0'\}\cdots\}| \leqq (\sum \beta_j)^n$$

so that

$$T_{\xi_n}\{\cdots\{T_{\xi_1}\bar{x}_0\}\cdots\} - T_{\xi_n}\{\cdots\{T_{\xi_1}\bar{x}_0'\}\cdots\} \to 0$$

as $n \to \infty$ in probability for all $\bar{x}_0$, $\bar{x}_0'$ and hence the hypothesis of Theorem 3 is satisfied.

Corollary 2. *Suppose there is a unique ergodic distribution corresponding to the transition probability measure generated by the conditional probability distribution* $F(a_n \mid a_{n-1} , \cdots)$ *and that the Doeblin condition* (D) ([1], [6]) *is satisfied by this transition probability measure. Further, given any two points* $\bar{x}_0$, $\bar{x}_0'$ ε S *(see Theorem 3) with* $0 < x_{-j} < x_{-j}' < 1$ *for all* j, *let*

$$(54) \qquad\qquad T_\alpha \bar{x}_0 < T_\alpha \bar{x}_0'$$

for all α, $0 < \alpha < 1$. *Then* $\{x_n\}$ *has a representation of the form* (43).

Notice that

$$x_n - x_n' = T_{\xi_n}\{\cdots\{T_{\xi_1}\bar{x}_0\}\cdots\} - T_{\xi_n}\{\cdots\{T_{\xi_1}\bar{x}_0'\}\cdots\} \leqq 0$$

if $\bar{x}_0 < \bar{x}_0'$, that is, if $x_{-j} < x_{-j}'$ for all j ($\leqq 0$). However, the hypothesis of the corollary implies that x_n , x_n' have a common limiting distribution, say $F(a)$. But

$$\mathrm{P}\{x_n' > a, x_n \leqq a\} + \mathrm{P}\{x_n' \leqq a\} = \mathrm{P}\{x_n' \leqq a\}$$

so that

$$\mathrm{P}\{x_n' > a, x_n \leqq a\} \to 0$$

at every point a ($F(a)$ is continuous since $F(a_n \mid a_{n-1} , \cdots)$ is continuous in a_n).

Given any $\epsilon > 0$ we can find $k(\epsilon)$ points $b_1 < \cdots < b_k$ such that

$$\bigcup_{j=1}^{k} \{x_n \leqq b_j \,, x_n' > b_j\} \supset \{x_n' - x_n > \epsilon\}.$$

But then

$$\mathrm{P}\{x_n' - x_n > \epsilon\} \to 0$$

for every $\epsilon > 0$ as $n \to \infty$, and hence $x_n' - x_n \to 0$ in probability as $n \to \infty$. This argument has been carried out for any two points $\bar{x}_0$, $\bar{x}_0'$ with $\bar{x}_0 < \bar{x}_0'$. However, given any two points $\bar{x}_0$, $\bar{x}_0'$ (without such an order relation) we can find a third point $\bar{x}_0''$ such that $\bar{x}_0'' < \bar{x}_0$, $\bar{x}_0'$ and hence $x_n' - x_n \to 0$ for x_n', x_n derived from any two points $\bar{x}_0'$, $\bar{x}_0$. Now simply apply Theorem 3.

We now construct the simple example mentioned at the beginning of this section to show that even though a strictly stationary process $\{x_n\}$ may not be a one-sided function $x_n = g(\xi_n \,, \xi_{n-1} \,, \cdots)$ of the random variables (42), it may have such a representation in terms of random variables ξ_n' satisfying part (a) of problem 1 and defined in a different and more suitable manner. Set up a sequence of points b_i, $i = 0, \pm 1, \cdots$, with $0 < \cdots < b_i < b_{i+1} < \cdots < 1$. Let $S \subset (0, 1)$ be

$$(55) \qquad S = \bigcup_{i=-\infty}^{\infty} \{y \mid b_{2i} < y \leqq b_{2i+1}\}$$

and $\bar{S} = (0, 1) - S$. Let $\{x_n\}$ be the strictly stationary Markov process with transition probability density

$$(56) \qquad p(y|y') = \begin{cases} p_i/(b_{i+1} - b_i) & \text{if} \quad b_i < y \leqq b_{i+1} \,, \quad y' \,\varepsilon\, S \,, \\ p_i/(b_i - b_{i-1}) & \text{if} \quad b_{i-1} < y \leqq b_i \,, \quad y' \,\varepsilon\, \bar{S} \,, \end{cases}$$

where the p_i's are distinct positive numbers with $\sum p_i = 1$. Consider the family of transformations T_α generated by the conditional distribution function of the process $\{x_n\}$. If $\sum_{j=-\infty}^{i-1} p_j < \alpha \leqq \sum_{j=-\infty}^{i} p_j$, then $b_i < T_\alpha x \leqq b_{i+1}$ when $x \,\varepsilon\, S$ and $b_{i-1} < T_\alpha x \leqq b_i$ when $x \,\varepsilon\, \bar{S}$. Thus if we start out with two points x, x' with $x \,\varepsilon\, S$, $x' \,\varepsilon\, \bar{S}$ their transforms $T_{\xi_n} \cdots T_{\xi_1} x$, $T_{\xi_n} \cdots T_{\xi_1} x'$ will never be both in S or both in $\bar{S}$. Moreover, at any stage n we can separate the transforms by a quantity bounded below by a positive number by specifying an appropriate condition of the form $\beta < \xi_n < \beta'$. Thus the condition of Theorem 3 is not satisfied and hence $\{x_n\}$ does not have a representation of form (43) with the ξ_n's given by (42). Nonetheless the process $\{x_n\}$ is of the same type as the process $\{y_n\}$ constructed in the proof of Theorem 2. By using the construction given in Theorem 2 we can generate random variables ξ_n' satisfying the conditions of part (a) of problem 1 and such that $x_n = g(\xi_n' \,, \xi_{n-1}' \,, \cdots)$.

In conclusion, it is worthwhile noting that problem 1 can be rephrased in the terminology of information theory [4]. The original process is passed through a deterministic channel that manufactures independent random variables ξ_n

independent of the past, that is, independent of x_{n-1}, x_{n-2}, $\cdots$. We ask under what circumstances the original process $\{x_n\}$ can be expressed in terms of the output of the channel $\{\xi_n\}$ with the time sense preserved, that is, when $x_n = g(\xi_n, \xi_{n-1}, \cdots)$ for some Borel function g.

REFERENCES

[1] J. L. DOOB, *Stochastic Processes*, New York (1953).

[2] W. FELLER, *An Introduction to Probability Theory and its Applications*, New York (1950).

[3] G. KALLIANPUR & N. WEINER, *Non-linear prediction*, Technical Report No. 1 (1956), Office of Naval Research, Cu-2-56-Nonr-266, (39)-CIRMIP, Project NR-047-015.

[4] A. I. KHINCHIN, *Mathematical Foundations of Information Theory*, New York (1957).

[5] P. LÉVY, *Theorie de l'addition des variables aleatoires*, Paris (1937).

[6] M. LOEVE, *Probability Theory*, New York (1955).

[7] M. ROSENBLATT, Functions of a Markov process that are Markovian, *Journal of Mathematics and Mechanics*, **8** (1959), pp. 585–596.

[8] H. WIELANDT, Unzerlegbare, nicht negative Matrizen, *Math. Zeit.*, **52** (1950), pp. 642–648.

Indiana University*
Bloomington, Indiana

*The author is now located in the Division of Applied Mathematics, Brown University.