FUNCTIONS OF FINITE MARKOV CHAINS[1]

By S. W. Dharmadhikari

University of California, Berkeley

0. Summary. This paper came out of an attempt to solve the following general problem: Suppose $\{Y_n, n \geqq 1\}$ is a stationary process with a finite state-space J. Under what conditions can we express it as a function of a finite Markov chain? More precisely, when can we find a stationary Markov chain $\{X_n, n \geqq 1\}$ with a finite state-space I and a function f on I onto J such that the process $\{f(X_n)\}$ has the same distribution as $\{Y_n\}$? We do not solve the general problem here but for mixing processes we obtain a theorem which is the best possible in a certain sense.

Suppose ϵ denotes a state of J and suppose s, t denote finite sequences of states of J. If $s = \epsilon_1 \cdots \epsilon_n$, let $p(s) = P[(Y_1, \cdots, Y_n) = s]$. For each ϵ, define $n(\epsilon)$ to be the largest integer n such that we can find s_i, t_i $(i = 1, \cdots, n)$ such that the matrix $\|p(s_i \epsilon t_j)\|$ is nonsingular. Gilbert [4] has shown that if $\{Y_n\}$ is a function f of a finite Markov chain $\{X_n\}$ and if f takes $N(\epsilon)$ states of I into the state ϵ of J, then $n(\epsilon) \leqq N(\epsilon)$. If $n(\epsilon) = N(\epsilon)$, then $\{Y_n\}$ is said to be a regular function of a Markov chain. Thus a necessary condition for $\{Y_n\}$ to be a function of a finite Markov chain is that $\sum n(\epsilon)$ is finite. It is proved here that if $\sum n(\epsilon) < \infty$ and if the process $\{Y_n\}$ is mixing, then there exists a positive integer m^* such that for every $m \geqq m^*$ the process $\{Y_{nm+1}, n \geqq 0\}$ is a function of a Markov chain with $\sum n(\epsilon)$ states. An example is constructed to show that m^* cannot, in general, be brought down to 1. Thus the whole process $\{Y_n, n \geqq 1\}$ may still not be a function of a Markov chain with $\sum n(\epsilon)$ states.

1. Introduction. Let $\{X_n, n \geqq 1\}$ be a stationary Markov chain with a finite state-space I, transition matrix M and initial distribution $\mathbf{m}$. Let f be a function on I onto some finite set J. We can assume that $J = \{0, 1, \cdots, D - 1\}$. Let $N(\epsilon)$ states of I go into the state ϵ of J. Then I can be conveniently represented as $\{\epsilon j \mid j = 1, \cdots, N(\epsilon); \epsilon = 0, 1, \cdots, D - 1\}$. The transition matrix can be partitioned as

$$M = \begin{pmatrix} M_{00} & \cdots & M_{0,D-1} \\ \vdots & & \vdots \\ M_{D-1,0} & \cdots & M_{D-1,D-1} \end{pmatrix},$$

where $M_{\epsilon\mu}$ has order $N(\epsilon) \times N(\mu)$. Finally $\mathbf{m}$ can be written as $\mathbf{m} = (m_0, \cdots, m_{D-1})$, where m_ϵ has $N(\epsilon)$ elements.

Received September 29, 1962.

[1] Prepared with the partial support of the National Science Foundation, Grant G-14648. Part of the author's Ph.D. thesis submitted to the University of California, Berkeley. Work done while on leave from the University of Poona.

Let ϵ_i be in J, $(i = 1, \cdots, n)$. Then

$$(1.1) \qquad P[f(X_i) = \epsilon_i ; i = 1, \cdots, n] = m_{\epsilon_1} M_{\epsilon_1 \epsilon_2} \cdots M_{\epsilon_{n-1}\epsilon_n} e_{\epsilon_n},$$

where e_ϵ is the column vector with all its $N(\epsilon)$ elements equal to unity.

Consider now a stationary process $\{Y_n, n \geqq 1\}$ with state-space J. If $\{Y_n\}$ is the function f of the Markov chain $\{X_n\}$ above, then (1.1) shows that

$$(1.2) \qquad P[Y_i = \epsilon_i ; i = 1, \cdots, n] = m_{\epsilon_1} M_{\epsilon_1 \epsilon_2} \cdots M_{\epsilon_{n-1}\epsilon_n} e_{\epsilon_n}.$$

Conversely, if we can find a transition matrix M and a stationary initial distribution $\mathbf{m}$ such that (1.2) holds, then $\{Y_n\}$ is a function of a finite Markov chain.

(1.2) shows that, as far as computations of the probabilities involving the Y-process are concerned, the nonnegativity of the matrix M does not play any role. The term "pseudo-Markov matrix" will be used to denote a square matrix whose rows add to 1. Gilbert [4] shows that starting with stationary processes satisfying certain conditions it is possible to obtain a class of pseudo-Markov matrices from which the probability structure of the given process can be reproduced through a functional approach represented by (1.2). However, in order to admit interpretation as probabilities, the elements of M must be nonnegative. Our approach will therefore be to construct certain classes of pseudo-Markov matrices and then try to find conditions under which the classes contain nonnegative matrices.

Blackwell and Koopmans [1] and Gilbert [4] have studied the problem of identifying the underlying transition matrix when it is already known that the process $\{Y_n\}$ is a function of a finite Markov chain. To each state ϵ, Gilbert attached a number $n(\epsilon)$ which can be defined completely in terms of the process $\{Y_n\}$ without any reference to an underlying Markov chain. He proved that for a function of a finite Markov chain $\sum n(\epsilon)$ must be finite and conjectured that this condition is also sufficient. Martin Fox has disproved this conjecture, but this result has not yet been published. Some sufficient conditions for $\{Y_n\}$ to be a function of a finite Markov chain have been obtained by this author [2] and by Fox [3].

In this paper we have tried to see how far one can go with the condition $\sum n(\epsilon) < \infty$ if the process $\{Y_n\}$ is mixing. Section 2 studies a matrix which was introduced by Gilbert and which is of fundamental importance in the present work. It is shown that certain mixing conditions characterize the asymptotic behavior of A^n. The main theorem of this paper is proved in Section 3. That this theorem is the best possible in a certain sense is exhibited by means of an example. This example also shows that the condition $\sum n(\epsilon) < \infty$ is not sufficient for $\{Y_n\}$ to be a regular function of a Markov chain. This result has been reported by Fox [3], but our example, which involves 2 states, appears to be simpler than his example, which uses 4 states.

2. The fundamental matrix A. Consider a stationary process $\{Y_n, n \geqq 1\}$ with a finite state-space $J = \{0, 1, \cdots, D - 1\}$. ϵ, μ, δ will denote states of J and

s, t will denote finite sequences of states of J. If $s = \epsilon_1 \cdots \epsilon_n$, let $p(s) = P[(Y_1, \cdots, Y_n) = s]$. Assume that each $p(\epsilon)$ is positive. The empty sequence denoted by $\varnothing$ will satisfy $p(\varnothing) = 1$ and $p(s\varnothing) = p(\varnothing s) = 1$. $P_\epsilon(s_1, \cdots, s_m; t_1, \cdots, t_n)$ will denote the $m \times n$ matrix whose (ij)th element is $p(s_i\epsilon t_j)$. For every ϵ, define the number $n(\epsilon)$ to be the highest integer n such that we can find s_i, t_i $(i = 1, \cdots, n)$, such that the matrix $\|p(s_i\epsilon t_j)\|$ is nonsingular. *It will be assumed throughout this work that $n(\epsilon)$ is finite for every ϵ in J.* This assumption is equivalent to the finiteness of $\sum n(\epsilon)$, because J is finite. Thus for each ϵ there exist $s_{\epsilon i}$, $t_{\epsilon i}$, $(i = 1, \cdots, n(\epsilon))$ such that

$$P_\epsilon(s_{\epsilon 1}, \cdots, s_{\epsilon n(\epsilon)}; t_{\epsilon 1}, \cdots, t_{\epsilon n(\epsilon)})$$

is nonsingular, whereas for every s and t, $P_\epsilon(s_{\epsilon 1}, \cdots, s_{\epsilon n(\epsilon)}, s; t_{\epsilon 1}, \cdots, t_{\epsilon n(\epsilon)}, t)$ is singular. Hence, for each s, there exist unique constants $a_{\epsilon i}(s)$, $(i = 1, \cdots, n(\epsilon))$, such that

$$p(s\epsilon t_{\epsilon j}) = \sum_{i=1}^{n(\epsilon)} a_{\epsilon i}(s) p(s_{\epsilon i}\epsilon t_{\epsilon j}), \qquad (j = 1, \cdots, n(\epsilon)). \tag{2.1}$$

The same constants satisfy, for every t

$$p(s\epsilon t) = \sum_{i=1}^{n(\epsilon)} a_{\epsilon i}(s) p(s_{\epsilon i}\epsilon t). \tag{2.2}$$

Similarly there are, for every t, unique constants $a_{\epsilon i}^*(t)$, $(i = 1, \cdots, n(\epsilon))$, such that, for every s

$$p(s\epsilon t) = \sum_{i=1}^{n(\epsilon)} a_{\epsilon i}^*(t) p(s\epsilon t_{\epsilon i}). \tag{2.3}$$

For every ϵ and μ, write $a_{\epsilon i,\mu j} = a_{\mu j}(s_{\epsilon i}\epsilon)$, $(i = 1, \cdots, n(\epsilon); j = 1, \cdots, n(\mu))$. Let $A_{\epsilon\mu}$ denote the $n(\epsilon) \times n(\mu)$ whose (ij)th element is $a_{\epsilon i,\mu j}$. Then from (2.2)

$$\begin{aligned} P_\mu(s_{\mu 1}, \cdots, s_{\mu n(\mu)}; \epsilon t) &= P_\epsilon(s_{\mu 1}\mu, \cdots, s_{\mu n(\mu)}\mu; t) \\ &= A_{\mu\epsilon} P_\epsilon(s_{\epsilon 1}, \cdots, s_{\epsilon n(\epsilon)}; t). \end{aligned} \tag{2.4}$$

By induction, we get

$$\begin{aligned} &P_\mu(s_{\mu 1}, \cdots, s_{\mu n(\mu)}; \epsilon_1 \cdots \epsilon_n \epsilon) \\ &\qquad = A_{\mu\epsilon_1} A_{\epsilon_1\epsilon_2} \cdots A_{\epsilon_n\epsilon} P_\epsilon(s_{\epsilon 1}, \cdots, s_{\epsilon n(\epsilon)}; \varnothing). \end{aligned} \tag{2.5}$$

Let $N = \sum n(\epsilon)$ and let A denote the following $N \times N$ matrix.

$$A = \begin{pmatrix} A_{00} & \cdots & A_{0,D-1} \\ \vdots & & \vdots \\ A_{D-1,0} & \cdots & A_{D-1,D-1} \end{pmatrix}. \tag{2.6}$$

We can partition every $N \times N$ matrix U exactly in the same way as A is. We will denote by $U_{\epsilon\mu}$ the submatrix of U corresponding to $A_{\epsilon\mu}$. $u_{\epsilon i,\mu j}$ will denote the (ij)th element of $U_{\epsilon\mu}$. Similarly a vector v of N elements will be written as $v = (v_0, \cdots, v_{D-1})$, where $v_\epsilon = (v_{\epsilon 1}, \cdots, v_{\epsilon n(\epsilon)})$.

For future use we want equations of a slightly more general form than (2.4) and (2.5). For $s = \epsilon_1 \cdots e_m$ and $t = \mu_1 \cdots \mu_n$, we will write $p(s\varnothing^k t) = P[(Y_1, \cdots, Y_m) = s, (Y_{m+k+1}, \cdots, Y_{m+k+n}) = t]$. Using (2.4) we get

$$\begin{aligned} P_\mu(s_{\mu 1}, \cdots, s_{\mu n(\mu)}\,; \varnothing\epsilon t) \\ (2.7) \qquad &= \sum_{\delta=0}^{D-1} P_\mu(s_{\mu 1}, \cdots, s_{\mu n(\mu)}\,; \delta\epsilon t) \\ &= \sum_\delta A_{\mu\delta}A_{\delta\epsilon}P_\epsilon(s_{\epsilon 1}, \cdots, s_{\epsilon n(\epsilon)}\,; t) = (A^2)_{\mu\epsilon}P_\epsilon(s_{\epsilon 1}, \cdots, s_{\epsilon n(\epsilon)}\,; t). \end{aligned}$$

By induction

$$\begin{aligned} (2.8) \qquad P_\mu(s_{\mu 1}, \cdots, s_{\mu n(\mu)}\,; \varnothing^{k_0-1}\epsilon_1\varnothing^{k_1-1}\epsilon_2 \cdots \epsilon_n\varnothing^{k_n-1}\epsilon) \\ = (A^{k_0})_{\mu\epsilon_1}(A^{k_1})_{\epsilon_1\epsilon_2} \cdots (A^{k_n})_{\epsilon_n\epsilon}P_\epsilon(s_{\epsilon 1}, \cdots, s_{\epsilon n(\epsilon)}\,; \varnothing). \end{aligned}$$

For convenience in stating certain results, we need to choose the $s_{\epsilon i}$ in such a way that $s_{\epsilon 1} = \varnothing$. That this can be done will follow from the following lemma whose proof is immediate.

Lemma 2.1. *Let ξ_i, $(i = 1, \cdots, k)$ be a basis for E^k, the k-dimensional Euclidean space. Let η be an arbitrary nonzero vector of E^k. Then η, together with some set of $(k-1)$ ξ's forms a basis for E^k.*

In our problem we have the nonsingular matrix $P_\epsilon(s_{\epsilon 1}, \cdots, s_{\epsilon n(\epsilon)}\,; t_{\epsilon 1}, \cdots, t_{\epsilon n(\epsilon)})$. Its rows therefore form a basis for $E^{n(\epsilon)}$. The vector $(p(\varnothing\epsilon t_{\epsilon 1}), \cdots, p(\varnothing\epsilon t_{\epsilon n(\epsilon)}))$ must be nonzero. If it were zero, then (2.1) implies that $a_{\epsilon i}(\varnothing) = 0$, for all i. Then (2.2) implies that $p(\varnothing\epsilon t) = 0$, for all t. Putting $t = \varnothing$, it follows that $p(\epsilon) = 0$. But $p(\epsilon)$ is assumed to be positive. Hence the above vector is nonzero. From the lemma we therefore conclude that we can omit some row of the matrix $P_\epsilon(s_{\epsilon 1}, \cdots, s_{\epsilon n(\epsilon)}\,; t_{\epsilon 1}, \cdots, t_{\epsilon n(\epsilon)})$, replace it by the nonzero vector above and still keep the resulting matrix nonsingular. Thus one of the $s_{\epsilon i}$ can be taken to be $\varnothing$. Since the ordering is immaterial, we can conveniently take $s_{\epsilon 1} = \varnothing$. Similarly we can take $t_{\epsilon 1} = \varnothing$.

We now proceed to study the connection between the limiting behavior of A^n and the probability structure of the process $\{Y_n\}$. We need to introduce two matrices. F will denote the $N \times N$ matrix such that for each $F_{\epsilon\mu}$, the first column consists of 1's and the remaining columns vanish. S will denote the $N \times N$ diagonal matrix such that the (ii)th element in $S_{\epsilon\epsilon}$ is $p(s_{\epsilon i}\epsilon)$. Observe that if $p(s_{\epsilon i}\epsilon) = 0$, then the ith row of the matrix $P_\epsilon(s_{\epsilon 1}, \cdots, s_{\epsilon n(\epsilon)}\,; t_{\epsilon 1}, \cdots, t_{\epsilon n(\epsilon)})$ vanishes. But this matrix is nonsingular. Thus each $p(s_{\epsilon i}\epsilon) > 0$. This implies that S is nonsingular.

Lemma 2.2.

(i) *$p(s\varnothing^n t)$ converges as $n \to \infty$ for every s and t if, and only if, $p(s_{\epsilon i}\epsilon\varnothing^n\mu t_{\mu j})$ converges as $n \to \infty$, for every ϵ, μ, i and j.*

(ii) *$p(s\varnothing^n t) \to p(s)\cdot p(t)$ as $n \to \infty$, for every s and t if, and only if, $p(s_{\epsilon i}\epsilon\varnothing^n\mu t_{\mu j}) \to p(s_{\epsilon i}\epsilon)\cdot p(\mu t_{\mu j})$ as $n \to \infty$ for every ϵ, μ, i and j.*

Proof. The "only if" parts of both (i) and (ii) are immediate. For the "if" parts, observe that there is nothing to prove if either s or t or both are empty. So let s and t be both nonempty. We can then write $s = s'\epsilon$ and $t = \mu t'$ for some s', t', ϵ and μ.

(i) Let $p(s_{\epsilon i}\epsilon\varnothing^n\mu t_{\mu i})$ converge as $n \to \infty$ for every ϵ, μ, i and j. Then from (2.2)

$$P(s\varnothing^n\mu t_{\mu j}) = p(s'\epsilon\varnothing^n\mu t_{\mu j}) = \sum_{i=1}^{n(\epsilon)} a_{\epsilon i}(s')p(s_{\epsilon i}\epsilon\varnothing^n\mu t_{\mu j}), \tag{2.9}$$

which converges because of the hypothesis. Now using (2.3) we have

$$p(s\varnothing^n t) = p(s'\epsilon\varnothing^n\mu t') = \sum_{j=1}^{n(\mu)} a^*_{\mu j}(t')p(s'\epsilon\varnothing^n\mu t_{\mu j}), \tag{2.10}$$

which converges because (2.9) does. This proves (i).

(ii) Let $p(s_{\epsilon i}\epsilon\varnothing^n\mu t_{\mu j}) \to p(s_{\epsilon i}\epsilon)\cdot p(\mu t_{\mu j})$ as $n \to \infty$, for every ϵ, μ, i and j. Then (2.9) implies

$$p(s\varnothing^n\mu t_{\mu j}) \to p(\mu t_{\mu j}) \sum_{i=1}^{n(\epsilon)} a_{\epsilon i}(s')p(s_{\epsilon i}\epsilon) = p(s'\epsilon)\cdot p(\mu t_{\mu j}).$$

Then (2.10) gives

$$p(s\varnothing^n t) \to p(s) \sum_{j=1}^{n(\mu)} a^*_{\mu j}(t')p(\mu t_{\mu j}) = p(s)\cdot p(\mu t') = p(s)\cdot p(t).$$

This proves (*ii*) and completes the proof of the lemma.

THEOREM 2.1.

(i) *A^n converges as $n \to \infty$ if, and only if $p(s\varnothing^n t)$ converges as $n \to \infty$ for every s and t.*

(ii) *$A^n \to SF$ as $n \to \infty$ if and only if, $p(s\varnothing^n t) \to p(s)\cdot p(t)$ as $n \to \infty$ for every s and t.*

PROOF. Let P be the $N \times N$ matrix such that

$$P_{\epsilon\mu} = P_\epsilon(s_{\epsilon 1}, \cdots, s_{\epsilon n(\epsilon)}; t_{\epsilon 1}, \cdots, t_{\epsilon n(\epsilon)}), \quad \text{if } \epsilon = \mu,$$
$$= 0, \quad \text{if } \epsilon \neq \mu.$$

The choice of the $s_{\epsilon i}$ and $t_{\epsilon j}$ implies that P is nonsingular. Using (2.4) whenever necessary we get

$$(AP)_{\epsilon\mu} = \sum_\delta A_{\epsilon\delta}P_{\delta\mu} = A_{\epsilon\mu}P_{\mu\mu} = A_{\epsilon\mu}P_\mu(s_{\mu 1}, \cdots, s_{\mu n(\mu)}; t_{\mu 1}, \cdots, t_{\mu n(\mu)})$$
$$= P_\epsilon(s_{\epsilon 1}, \cdots, s_{\epsilon n(\epsilon)}; \mu t_{\mu 1}, \cdots, \mu t_{\mu n(\mu)}).$$

$$(A^2P)_{\epsilon\mu} = \sum_\delta A_{\epsilon\delta}(AP)_{\delta\mu} = \sum_\delta A_{\epsilon\delta}P_\delta(s_{\delta 1}, \cdots, s_{\delta n(\delta)}; \mu t_{\mu 1}, \cdots, \mu t_{\mu n(\mu)}) \tag{2.11}$$
$$= \sum_\delta P_\epsilon(s_{\epsilon 1}, \cdots, s_{\epsilon n(\epsilon)}; \delta\mu t_{\mu 1}, \cdots, \delta\mu t_{\mu n(\mu)})$$
$$= P_\epsilon(s_{\epsilon 1}, \cdots, s_{\epsilon n(\epsilon)}; \varnothing\mu t_{\mu 1}, \cdots, \varnothing\mu t_{\mu n(\mu)}).$$

In general,

$$(A^{n+1}P)_{\epsilon\mu} = P_\epsilon(s_{\epsilon 1}, \cdots, s_{\epsilon n(\epsilon)}; \varnothing^n\mu t_{\mu 1}, \cdots, \varnothing^n \mu t_{\mu n(\mu)}).$$

Since P is nonsingular,

$$\begin{aligned} A^n \text{ converges} &\Leftrightarrow A^nP \text{ converges;}\\ &\Leftrightarrow (A^nP)_{\epsilon\mu} \text{ converges for each } \epsilon \text{ and } \mu;\\ &\Leftrightarrow p(s_{\epsilon i}\epsilon\varnothing^n\mu t_{\mu j}) \text{ converges for each } \epsilon, \mu, i \text{ and } j;\\ &\Leftrightarrow p(s\varnothing^n t) \text{ converges for every } s \text{ and } t;\end{aligned}$$

where the last step follows from Lemma 2.2. This proves (i).

Since $S_{\epsilon\mu}$ and $P_{\epsilon\mu}$ vanish for $\epsilon \neq \mu$,

$$(SFP)_{\epsilon\mu} = S_{\epsilon\epsilon}F_{\epsilon\mu}P_{\mu\mu} = \begin{pmatrix} p(s_{\epsilon 1}\,\epsilon) & 0\cdots 0\\ \vdots & \vdots\\ p(s_{\epsilon n(\epsilon)}\,\epsilon) & 0\cdots 0\end{pmatrix} P_\mu(s_{\mu 1}, \cdots, s_{\mu n(\mu)}; t_{\mu 1}, \cdots, t_{\mu n(\mu)}).$$

Hence the (ij)th term in $(SFP)_{\epsilon\mu}$ is $p(s_{\epsilon i}\epsilon)p(\mu t_{\mu j})$, using the fact that $s_{\mu 1} = \varnothing$. Thus

$$\begin{aligned} A^n \to SF &\Leftrightarrow (A^nP)_{\epsilon\mu} \to (SFP)_{\epsilon\mu}; \text{ for every } \epsilon \text{ and } \mu;\\ &\Leftrightarrow p(s_{\epsilon i}\epsilon\varnothing^n\mu t_{\mu j}) \to p(s_{\epsilon i}\epsilon)p(\mu t_{\mu j}), \text{ for every } \epsilon, \mu, i \text{ and } j;\\ &\Leftrightarrow p(s\varnothing^n t) \to p(s)p(t), \text{ for every } s \text{ and } t;\end{aligned}$$

where the last step again follows by Lemma 2.2. This completes the proof of the theorem.

3. A theorem for mixing processes. We start with a definition of "mixing."

DEFINITION. *The process* $\{Y_n\}$ *is said to be mixing if for every* s *and* t, $p(s\varnothing^n t) \to p(s)\cdot p(t)$ *as* $n \to \infty$.

Thus part (ii) of Theorem 2.1 says that $A^n \to SF$ if, and only if, the process is mixing. The importance of the matrix A is exhibited by the following lemma. The lemma is essentially contained in Theorem 2 of Gilbert [4]. But since it contains the crux of the arguments of this paper, its proof will also be given. Recall that $N = \sum n(\epsilon)$.

LEMMA 3.1.

I. *Let* U *be an* $N \times N$ *nonsingular matrix satisfying*

(i) $U_{\epsilon\mu} = 0$ *if* $\epsilon \neq \mu$, *and*

(ii) $\sum_{j=1}^{n(\epsilon)} u_{\epsilon i, \epsilon j} = p(s_{\epsilon i}\epsilon)$, *for every* ϵ *and for* $i = 1, \cdots, n(\epsilon)$.

Let $M = U^{-1}AU$. Then

(a) *M is a pseudo-Markov matrix; that is, every row of M adds up to* 1;

(b) *if v_ϵ is the first row of $U_{\epsilon\epsilon}$ and if $v = (v_0, \cdots, v_{D-1})$, then v is a stationary initial distribution for M; that is, $vM = v$;*

$$(3.1) \quad (c) \qquad p(\mu\epsilon_1 \cdots \epsilon_n\epsilon) = v_\mu M_{\mu\epsilon_1}M_{\epsilon_1\epsilon_2}\cdots M_{\epsilon_n\epsilon}e_\epsilon,$$

where e_ϵ is the $n(\epsilon) \times 1$ matrix consisting of 1's.

II. *Conversely, let M be a pseudo-Markov matrix of order N and let v be a stationary initial distribution for M such that* (c) *holds. Then there is a nonsingular U such that* (i) *and* (ii) *are satisfied, v_ϵ is the first row of $U_{\epsilon\epsilon}$ and $M = U^{-1}AU$.*

PROOF. (I) will be proved first. It is clear that

$$M_{\epsilon\mu} = U_{\epsilon\epsilon}^{-1}A_{\epsilon\mu}U_{\mu\mu}, \qquad v_\epsilon U_{\epsilon\epsilon}^{-1} = (1, 0, \cdots, 0),$$

$$U_{\mu\mu}e_\mu = P_\mu(s_{\mu 1}, \cdots, s_{\mu n(\mu)}; \varnothing), \text{ and } \qquad e_\mu = U_{\mu\mu}^{-1}P_\mu(s_{\mu 1}, \cdots, s_{\mu n(\mu)}; \varnothing).$$

Thus

$$\begin{aligned}\sum_\mu M_{\epsilon\mu}e_\mu &= U_{\epsilon\epsilon}^{-1}\sum_\mu A_{\epsilon\mu}P_\mu(s_{\mu 1}, \cdots, s_{\mu n(\mu)}; \varnothing)\\ &= U_{\epsilon\epsilon}^{-1}\sum_\mu P_\epsilon(s_{\epsilon 1}, \cdots, s_{\epsilon n(\epsilon)}; \mu) \qquad \text{[see (2.4)]}\\ &= U_{\epsilon\epsilon}^{-1}P_\epsilon(s_{\epsilon 1}, \cdots, s_{\epsilon n(\epsilon)}; \varnothing) = e_\epsilon,\end{aligned}$$

which proves (a). Observe that $\sum_\epsilon v_\epsilon M_{\epsilon\mu} = [\sum_\epsilon (1, 0, \cdots, 0)A_{\epsilon\mu}]U_{\mu\mu}$. Since $s_{\epsilon 1} = \varnothing$, the general term in the bracketed matrix is

$$\begin{aligned}\sum_\epsilon a_{\mu j}(\epsilon) = a_{\mu j}(\varnothing) &= 1 \qquad \text{if } j = 1\\ &= 0 \qquad \text{if } j \neq 1.\end{aligned}$$

Thus $\sum_\epsilon v_\epsilon M_{\epsilon\mu} = (1, 0, \cdots, 0)U_{\mu\mu} = v_\mu$, which proves (b). Finally

$$\begin{aligned}v_\mu M_{\mu\epsilon_1} \cdots M_{\epsilon_n\epsilon}e_\epsilon &= (1, 0, \cdots, 0)A_{\mu\epsilon_1} \cdots A_{\epsilon_n\epsilon}P_\epsilon(s_{\epsilon 1}, \cdots, s_{\epsilon n(\epsilon)}; \varnothing)\\ &= (1, 0 \cdots, 0)P_\mu(s_{\mu 1}, \cdots, s_{\mu n(\mu)}; \epsilon_1 \cdots \epsilon_n\epsilon)\\ &= p(\mu\epsilon_1 \cdots \epsilon_n\epsilon), \qquad \text{(since } s_{\mu 1} = \varnothing).\end{aligned}$$

This proves (c) and completes the proof of I.

To prove II, let $q_\epsilon(\varnothing) = v_\epsilon$, $r_\epsilon(\varnothing) = e_\epsilon$ and for $s = \epsilon_1 \cdots \epsilon_n$ and $t = \mu_1 \cdots \mu_n$, let $q_\epsilon(s) = v_{\epsilon_1}M_{\epsilon_1\epsilon_2} \cdots M_{\epsilon_n\epsilon}$, and $r_\mu(t) = M_{\mu\mu_1} \cdots M_{\mu_{n-1}\mu_n}e_{\mu_n}$. Let Q_ϵ be the $n(\epsilon) \times n(\epsilon)$ matrix whose ith row is $q_\epsilon(s_{\epsilon i})$ and let R_ϵ be the $n(\epsilon) \times n(\epsilon)$ matrix whose jth column is $r_\mu(t_{\mu j})$. Then, since (c) holds,

$$P_\epsilon(s_{\epsilon 1}, \cdots, s_{\epsilon n(\epsilon)}; t_{\epsilon 1}, \cdots, t_{\epsilon n(\epsilon)}) = Q_\epsilon R_\epsilon.$$

This shows that Q_ϵ and R_ϵ are nonsingular. Again, since (c) holds,

$$\begin{aligned}Q_\epsilon M_{\epsilon\mu}R_\mu &= P_\epsilon(s_{\epsilon 1}, \cdots, s_{\epsilon n(\epsilon)}; \mu t_{\mu 1}, \cdots, \mu t_{\mu n(\mu)})\\ &= A_{\epsilon\mu}P_\mu(s_{\mu 1}, \cdots, s_{\mu n(\mu)}; t_{\mu 1}, \cdots, t_{\mu n(\mu)}) = A_{\epsilon\mu}Q_\mu R_\mu.\end{aligned}$$

Thus $M_{\epsilon\mu} = Q_\epsilon^{-1}A_{\epsilon\mu}Q_\mu$.

Let $U_{\epsilon\epsilon} = Q_\epsilon$ and let U be the direct sum of the $U_{\epsilon\epsilon}$. Then (i) is satisfied. Since $q_\epsilon(s)e_\epsilon = p(s\epsilon)$, it follows that (ii) is satisfied. The first row of $U_{\epsilon\epsilon}$ is $q_\epsilon(\varnothing) = v_\epsilon$. Finally it is clear that $M = U^{-1}AU$. This completes the proof of the lemma.

Comparing (3.1) with (1.2) we see that if we can find a matrix U, such that the first row of each $U_{\epsilon\epsilon}$ is nonnegative and such that $M = U^{-1}AU$ is nonnegative, then we would be able to say that $\{Y_n\}$ is a function of a Markov chain with $\sum n(\epsilon)$ states. That this is in general impossible will be shown later by means of an example. The situation is the same even if it is assumed that the process is mixing. It is, however, of interest to find out how far we can go for processes which are mixing. First we state a lemma which follows immediately from Lemma 3.1.

LEMMA 3.2. *Let U, M, v and e_ϵ be the same as in part* I *of Lemma* 3.1. *Then*

$$(3.2)\quad p(\mu\varnothing^{k_0-1}\epsilon_1\varnothing^{k_1-1}\epsilon_2\cdots\epsilon_n\varnothing^{k_n-1}\epsilon) = v_\mu(M^{k_0})_{\mu\epsilon_1}(M^{k_1})_{\epsilon_1\epsilon_2}\cdots(M^{k_n})_{\epsilon_n\epsilon}e_\epsilon .$$

Lemma 3.1 says that v is a stationary initial distribution for M. The next lemma gives a sufficient condition for v to be the only such distribution.

LEMMA 3.3. *Let U, M, v be as in Lemma* 3.1. *Let W be an $N \times N$ matrix all of whose rows are equal to w. If $M^n \to W$, then $w = v$.*

PROOF. Observe that $s_{\epsilon 1} = \varnothing$ implies that the elements of v_ϵ add to $p(\epsilon)$. Hence the elements of v add to 1. Therefore $vM = v \Leftrightarrow vM^n = v \Leftrightarrow vW = v \Leftrightarrow w = v$.

This lemma gives us the only matrix we can obtain as a limit of M^n if we want all the limiting rows to be identical. The next lemma gives necessary and sufficient conditions that this should happen. We recall that (a) S denotes the $N \times N$ diagonal matrix such that the (ii)th term in $S_{\epsilon\epsilon}$ is $p(s_{\epsilon i}\epsilon)$; (b) S is nonsingular; (c) F denotes the $N \times N$ matrix such that for each $F_{\epsilon\mu}$, the first column consists of 1's and the remaining columns vanish.

LEMMA 3.4. *Let U, M, v be as in Lemma* 3.1. *Let V be the $N \times N$ matrix all of whose rows are equal to v. Then $M^n \to V$ if, and only if, the process is mixing.*

PROOF. U can be written as $U = SW$, where

(a) $W_{\epsilon\mu} = 0$, if $\epsilon \neq \mu$;
(b) each row of W adds to 1;
(c) W is nonsingular.

(b) and (c) imply that $WF = F = W^{-1}F$. From Theorem 2.1,

$$\text{The process is mixing} \Leftrightarrow A^n \to SF;$$

$$\Leftrightarrow M^n \to U^{-1}SFU = W^{-1}FU = FU;$$

$$\Leftrightarrow \text{for every } \epsilon \text{ and } \mu;$$

$$(M^n)_{\epsilon\mu} \to F_{\epsilon\mu}U_{\mu\mu} = (v_\mu', \cdots, v_\mu')';$$

$$\Leftrightarrow M^n \to V;$$

which proves the lemma.

We are now ready to state the main theorem of this paper. Let I be a set of $N = \sum n(\epsilon)$ points which can be conveniently enumerated as

$$\{\epsilon i \mid i = 1, \cdots, n(\epsilon), \epsilon = 0, \cdots, D-1\}.$$

Let f be the function on I onto J defined by

$$f(\epsilon i) = \epsilon. \tag{3.3}$$

THEOREM 3.1. *Let $\{Y_n, n \geqq 1\}$ be a stationary process with a finite state-space J. Let $\sum n(\epsilon) < \infty$ and let the process be mixing. Let I be a set of $N = \sum n(\epsilon)$ elements and f the function on I onto J defined by (3.3). Then there exists a positive integer m^* such that for every $m \geqq m^*$, the process $\{Y_{nm+1}, n \geqq 0\}$ is the function f of a stationary Markov chain with state-space I.*

PROOF. Let U, M, v, V and e_ϵ be as in Lemmas 3.1 and 3.4, with the additional restriction that each entry of each v_ϵ is positive.

(a) Let $M^m \geqq 0$ and let $\{Z_n, n \geqq 1\}$ be a stationary Markov chain with state-space I, transition matrix M^m and initial distribution v. Then, using Lemma 3.2, we have

$$\begin{aligned} P[f(Z_1), \cdots, f(Z_{n+2})) = \mu\epsilon_1 \cdots, \epsilon_n\epsilon] &= v_\mu(M^m)_{\mu\epsilon_1} \cdots (M^m)_{\epsilon_n\epsilon}e_\epsilon \\ &= p(\mu\varnothing^{m-1}\epsilon_1 \cdots \epsilon_n\varnothing^{m-1}\epsilon) \\ &= P[(Y_1, Y_{m+1}, \cdots, Y_{(n+1)m+1}) = \mu\epsilon_1 \cdots \epsilon_n\epsilon]. \end{aligned}$$

That is $\{f(Z_n), n \geqq 1\}$ has the same distribution as $\{Y_{nm+1}, n \geqq 0\}$.

(b) Since $\{Y_n\}$ is mixing, Lemma 3.4 implies that $M^n \to V$, which has all its entries positive. Therefore there is a positive integer m^* such that for every $m \geqq m^*$ each entry of M^m is nonnegative. The theorem now follows from (a).

The rest of this section will be devoted to showing that the above theorem is the best possible in a certain sense. Precisely, we intend to show by means of an example that under the conditions imposed the whole process $\{Y_n\}$ may still not be a function of a Markov chain with $\sum n(\epsilon)$ states—that is, a regular function of a Markov chain. Using Lemma 3.1 we see that it will be enough to show that no matter how we choose U satisfying conditions (i) and (ii), $M = U^{-1}AU$ will have at least one negative entry. Because $U_{\epsilon\mu} = 0$ for $\epsilon \neq \mu$, it will be enough to exhibit that, for some ϵ, $A_{\epsilon\epsilon}$ cannot be similar to a nonnegative matrix. Since similar matrices have the same trace, it will suffice to construct a process for which the trace of some $A_{\epsilon\epsilon}$ is strictly negative. We proceed to do this now. Consider the pseudo-Markov matrix

$$M = \begin{pmatrix} \lambda_1 & 0 & 0 & 1 - \lambda_1 \\ 0 & -\lambda_2 & 0 & 1 + \lambda_2 \\ 0 & 0 & -\lambda_3 & 1 + \lambda_3 \\ 1 - \lambda_1, & c(1 + \lambda_2), & -c(1 + \lambda_3) & \lambda_1 + c(\lambda_3 - \lambda_2) \end{pmatrix},$$

with the stationary initial distribution $(.5, c \times .5, -c \times .5, .5)$. Let us number the states as 01, 02, 03 and 11. Let f be the function defined by $f(0i) = 0$, $(i = 1, 2, 3)$ and $f(11) = 1$. Then we will get a 2-state function process for which we can compute the pseudo-probabilities using (1.5). These will satisfy all the consistency relations like $p(00) + p(01) = p(0)$, because the rows of M add to 1, as also does the stationary distribution. We will therefore get a proper process as soon as we ensure that all the pseudo-probabilities are nonnegative. Let ϵ^n denote the sequence of n ϵ's in succession and let $\epsilon^0 = \varnothing$, by convention. Since only one state of the underlying process goes into the state 1 of the function process, it is enough to ensure that pseudo-probabilities of the form $p(0^n)$, $p(0^n1)$, $p(1^n)$ and $p(10^n1)$ are nonnegative. We have

$$\begin{aligned} p(0^n) &= (.5)[\lambda_1^{n-1} + c(-1)^{n-1}(\lambda_2^{n-1} - \lambda_3^{n-1})], \\ p(0^n1) &= (.5)[\lambda_1^{n-1}(1 - \lambda_1) + c(-1)^{n-1}\{\lambda_2^{n-1}(1 + \lambda_2) - \lambda_3^{n-1}(1 + \lambda_3)\}], \\ p(1^n) &= (.5)[\lambda_1 + c(\lambda_3 - \lambda_2)]^{n-1}, \\ p(10^n1) &= p(0^n1) - p(0^{n+1}1) \\ &= (.5)[\lambda_1^{n-1}(1 - \lambda_1)^2 + c(-1)^{n-1}\{\lambda_2^{n-1}(1 + \lambda_2)^2 - \lambda_3^{n-1}(1 + \lambda_3)^2\}]. \end{aligned}$$

These expressions show that the pseudo-probabilities will come out nonnegative if we choose the λ's and c in such a way that

$$\begin{aligned} &0 < \lambda_i < 1, (i = 1, 2, 3); \quad \lambda_1 > \lambda_i, (i = 2, 3); \quad 0 < c < 1; \\ &\lambda_1 + c(\lambda_3 - \lambda_2) > 0; \quad (1 - \lambda_1)^k > c(1 + \lambda_i)^k, (i = 2, 3; k = 1, 2). \end{aligned} \tag{3.4}$$

It is clear that Conditions (3.4) can be satisfied. For example, let $\lambda_1 = .5$, $\lambda_2 = .4$, $\lambda_3 = .3$ and $c = .06$. Then $\lambda_1 + c(\lambda_3 - \lambda_2) = .494$, $(1 - \lambda_1) = .5$, $c(1 + \lambda_2) = .084$, $c(1 + \lambda_3) = .078$, $(1 - \lambda_1)^2 = .25$, $c(1 + \lambda_2)^2 = .1176$, $c(1 + \lambda_3)^2 = .1014$. The Conditions (3.4) are thus fulfilled. The values of $p(0^n)$ for $n = 1, \cdots, 5$ are, respectively, .5, .247, .1271, .06139 and .031775. Numerical computations give the determinant of $P_0(\varnothing, 0, 00; \varnothing, 0, 00)$ as 2.3328×10^{-6}, which is nonzero. Hence $n(0) = 3$ and $n(1) = 1$. Now M can be partitioned as

$$\begin{pmatrix} M_{00} & M_{01} \\ M_{10} & M_{11} \end{pmatrix},$$

where M_{00} has order 3×3. From part II of Lemma 3.1 we know that for the 2-state function process obtained above, A_{00} must be similar to M_{00}. But the trace of M_{00} is $-.2$, which is negative. Hence A_{00} cannot be similar to a nonnegative matrix. What we have done so far shows that *the condition* $\sum n(\epsilon) < \infty$ *does not characterize regular functions of Markov chains.*

In order to complete the proof that Theorem 3.1 is the best possible in the sense described before, we only need to show that the process constructed above is mixing. This will follow from the following lemma whose proof is straightforward.

LEMMA 3.5. *Let B be a pseudo-Markov matrix of order n. Let 1, λ_i, $(i = 2, \cdots, n)$ be the characteristic roots of B. Let $|\lambda_i| < 1$, $(i = 2, \cdots, n)$. Then $B^n \to W$, where each row of W equals the same vector w.*

For the numerical example constructed above computations show that all the characteristic roots of M, other than 1, lie strictly within the unit circle. The conclusion of Lemma 3.5, together with Lemmas 3.3 and 3.4, then shows that the process is mixing. This shows that all the assumptions of Theorem 3.1 are satisfied and A_{00} is still not similar to a nonnegative matrix. Under the conditions imposed, the m^* given by Theorem 3.1 cannot therefore be brought down to 1 in general. In this sense the theorem is therefore the best possible.

4. Acknowledgments. The author is grateful to Professor Edward W. Barankin who suggested the problem and under whose guidance this work was carried out. Thanks are due to Professor David Blackwell for his interest in this work, to Dr. Shishirkumar S. Jogdeo for checking certain sections of the manuscript and to Professor Martin Fox for keeping the author posted with his results on the problem.

REFERENCES

[1] Blackwell, D. and Koopmans, L. (1957). On the identifiability problem for functions of finite Markov chains. *Ann. Math. Statist.* **28** 1011–1015.

[2] Dharmadhikari, S. W. (1963). Sufficient conditions for a stationary process to be a function of a finite Markov chain. *Ann. Math. Statist.* **34** 1033–1041.

[3] Fox, Martin. (1962). Conditions under which a given process is a function of a Markov chain (abstract). *Ann. Math. Statist.* **33** 1206.

[4] Gilbert, Edgar J. (1959). On the identifiability problem for functions of finite Markov chains. *Ann. Math. Statist.* **30** 688–697.