A CHARACTERISATION OF A CLASS OF FUNCTIONS OF FINITE MARKOV CHAINS

BY S. W. DHARMADHIKARI

Indian Statistical Institute

0. Summary. In a previous paper ([2], Theorem 3.1) this author has given some sufficient conditions for a stationary process to be a function of a finite Markov chain. Suppose $\mathfrak{F}$ denotes the class of functions which satisfy these conditions. In this paper we give a characterisation of $\mathfrak{F}$. Using this characterisation $\mathfrak{F}$ is shown to be wider than the class of regular functions of finite Markov chains.

1. Preliminaries. Suppose $\{Y_n, n \geqq 1\}$ is a stationary process with a finite state-space $J = \{0, 1, \cdots, D - 1\}$. We will use the notation of [1] and [2] and assume that $\sum_\epsilon n(\epsilon) < \infty$.

The following are the conditions of Theorem 3.1 of [2].

DEFINITON. *We say that* $\{Y_n\}$ *satisfies the Conditions* (c) *if, for every* ϵ, *there exists a convex polyhedral cone* $\mathcal{C}_\epsilon$ *such that*

(c 1): $\mathcal{C}(\alpha_\epsilon) \subset \mathcal{C}_\epsilon \subset [\mathcal{C}(\pi_\epsilon)]^+$, *for every* ϵ; *and*

(c 2): $\beta_\epsilon A_{\epsilon\mu}$ *belongs to* $\mathcal{C}_\mu$ *for every* ϵ, μ *and for every* β_ϵ *in* $\mathcal{C}_\epsilon$.

If $\{Y_n\}$ satisfies (c), let $\beta_{\epsilon j}$, $j = 1, \cdots, N(\epsilon)$, be the generators of $\mathcal{C}_\epsilon$. Let B_ϵ denote the $N(\epsilon) \times n(\epsilon)$ matrix whose jth row is $\beta_{\epsilon j}$. The Condition (c 1) and Lemma 1.1 of [2] show that B_ϵ has rank $n(\epsilon)$.

For future use we need the following lemma.

LEMMA 1. *The vector* $\pi_\epsilon(\phi)$ *is in the interior of* $\mathcal{C}(\pi_\epsilon)$.

PROOF. Let us recall ([1], p. 1025) that $s_{\epsilon 1}$ and $t_{\epsilon 1}$ can be taken to be ϕ. If $n(\epsilon) = 1$, the lemma thus asserts that the point 1 is in the interior of the non-negative real line. So, let $n(\epsilon) \geqq 2$. Observe that $\pi_\epsilon(\phi) = \sum^{(n)} \pi_\epsilon(\mu_1, \cdots, \mu_n)$, where $\sum^{(n)}$ denotes summation over all possible sequences $(\mu_1, \cdots, \mu_n)$ of length n. Thus, for every t, the vector $\pi_\epsilon(\phi) - \pi_\epsilon(t)$ belngs to $\mathcal{C}(\pi_\epsilon)$.

Let $\xi_i = \pi_\epsilon(t_{\epsilon i})$ and $\eta_i = \pi_\epsilon(\phi) - \pi_\epsilon(t_{\epsilon i})$, $i = 2, \cdots, n(\epsilon)$. Then the η_i's belong to $\mathcal{C}(\pi_\epsilon)$; ξ_i and η_i are linearly independent; and $\xi_i + \eta_i = \pi_\epsilon(\phi)$. If $\mathcal{C}$ is the convex cone generated by the ξ's and η's, $\pi_\epsilon(\phi)$ is thus in the interior of $\mathcal{C}$. But it is easy to see that $\mathcal{C}$ has dimension $n(\epsilon)$. Since $\mathcal{C} \subset \mathcal{C}(\pi_\epsilon)$, the lemma is proved.

Suppose $\{X_n, n\ 1\} \geqq$ is a stationary Markov chain with a finite state-space $I = \{(\epsilon, j) \mid j = 1, \cdots, N(\epsilon); \epsilon \text{ in } J\}$, and suppose f is the function on I to J defined by $f[(\epsilon, j)] = \epsilon$. Let $\mathbf{m}$ be the initial distribution and M the transition matrix of $\{X_n\}$. The function f can be used to partition $\mathbf{m}$ into sub-vectors m_ϵ, $(\epsilon = 0, 1, \cdots, D - 1)$ and to partition M into submatrices $M_{\epsilon\mu}$, $(\epsilon, \mu = 0, 1, \cdots, D - 1)$ in the natural way. If s has length m and t has length n, we define

$$q_{\epsilon j}(s) = P[(f(X_1), \cdots, f(X_m)) = s, \quad X_{m+1} = (\epsilon, j)],$$

Received 26 May 1964; revised 9 September 1964.

and

$$r_{\epsilon j}(t) = P[(f(X_2), \cdots, f(X_{n+1})) = t \mid X_1 = (\epsilon, j)].$$

Let $q_\epsilon(s)$ be the row vector whose jth element is $q_{\epsilon j}(s)$ and let $r_\epsilon(t)$ be the column vector whose jth element is $r_{\epsilon j}(t)$. By convention $q_\epsilon(\phi) = m_\epsilon$ and $r_\epsilon(\phi) = e_\epsilon$, the column vector all of whose $N(\epsilon)$ elements are equal to unity. The Markov character of $\{X_n\}$ shows that

$$(1.1) \qquad q_\mu(s\epsilon) = q_\epsilon(s)M_{\epsilon\mu} \quad \text{and} \quad r_\epsilon(\mu t) = M_{\epsilon\mu}r_\mu(t)$$

Let K_ϵ be the linear space generated by $\{q_\epsilon(s), \text{ all } s\}$ and let L_ϵ be the linear space generated by $\{r_\epsilon(t), \text{ all } t\}$. Suppose $k(\epsilon)$ and $l(\epsilon)$ respectively denote the ranks of K_ϵ and L_ϵ. Since $q_\epsilon(s)$ and $r_\epsilon(t)$ are $N(\epsilon)$-dimensional vectors, we have, for every ϵ,

$$(1.2) \qquad k(\epsilon) \leqq N(\epsilon) \quad \text{and} \quad l(\epsilon) \leqq N(\epsilon).$$

Suppose now that $\{Y_n\}$ is the function f of the Markov chain $\{X_n\}$ above. Then, for all ϵ, s and t,

$$(1.3) \qquad p(s\epsilon t) = q_\epsilon(s)r_\epsilon(t).$$

From (1.3) and from the definitions of $n(\epsilon)$, $k(\epsilon)$ and $l(\epsilon)$ it follows that

$$(1.4) \qquad n(\epsilon) \leqq k(\epsilon) \quad \text{and} \quad n(\epsilon) \leqq l(\epsilon).$$

Combining (1.2) and (1.4), we get

$$(1.5) \qquad n(\epsilon) \leqq k(\epsilon) \leqq N(\epsilon) \quad \text{and} \quad n(\epsilon) \leqq l(\epsilon) \leqq N(\epsilon).$$

2. The characterisation. This section continues the work begun in [2]. The following theorem will be established.

THEOREM. *The following three statements are equivalent:*

I. *The stationary process* $\{Y_n\}$ *has* $\sum_\epsilon n(\epsilon) < \infty$ *and satisfies the Conditions* (c)

II. $\{Y_n\}$ *can be expressed as a function of a stationary finite Markov chain in such a way that* $k(\epsilon) = n(\epsilon)$, *for all* ϵ.

III. $\{Y_n\}$ *can be expressed as a function of a stationary finite Markov chain in such a way that* $l(\epsilon) = n(\epsilon)$ *for all* ϵ.

PROOF. (a) Suppose I holds. Then it was shown in [2] that $\{Y_n\}$ can be expressed as a function of a finite Markov chain in such a way that $r_{\epsilon j}(t) = (\beta_{\epsilon j}, \pi_\epsilon(t))$ for all ϵ, j and t. This means that $r_\epsilon(t) = B_\epsilon \pi_\epsilon'(t)$. Since B_ϵ has rank $n(\epsilon)$ it follows that $l(\epsilon) \leqq n(\epsilon)$. But now (1.5) shows that $l(\epsilon) = n(\epsilon)$. Thus I $\Rightarrow$ III.

(b) Suppose III holds. Then, as shown by Gilbert [3], $\sum_\epsilon n(\epsilon) < \infty$. Let P_ϵ be the $n(\epsilon) \times n(\epsilon)$ matrix whose (i, j)th element is $p(s_{\epsilon i}\epsilon t_{\epsilon j})$. Let Q_ϵ be the $n(\epsilon) \times N(\epsilon)$ matrix whose ith row is $q_\epsilon(s_{\epsilon i})$. Finally, let R_ϵ be the $N(\epsilon) \times n(\epsilon)$ matrix whose jth column is $r_\epsilon(t_{\epsilon j})$. Then (1.3) shows that $P_\epsilon = Q_\epsilon R_\epsilon$. This means that both Q_ϵ and R_ϵ have rank $n(\epsilon)$.

Since $l(\epsilon) = n(\epsilon)$, the columns of R_ϵ must span L_ϵ. Therefore for each t, there is a column vector $\alpha_\epsilon^*(t)$ such that

$$r_\epsilon(t) = R_\epsilon \alpha_\epsilon^*(t). \tag{2.1}$$

Premultiplication by Q_ϵ yields

$$\pi_\epsilon'(t) = P_\epsilon \alpha_\epsilon^*(t). \tag{2.2}$$

Let $B_\epsilon = R_\epsilon P_\epsilon^{-1}$ and let $\mathfrak{C}_\epsilon$ be the convex polyhedral cone generated by the rows of B_ϵ. Then from (2.1) and (2.2), we get

$$B_\epsilon \pi_\epsilon'(t) = R_\epsilon P_\epsilon^{-1} P_\epsilon \alpha_\epsilon^*(t) = R_\epsilon \alpha_\epsilon^*(t) = r_\epsilon(t).$$

This shows that $\mathfrak{C}_\epsilon \subset [\mathfrak{C}(\pi_\epsilon)]^+$.

Now $\alpha_\epsilon(s)\pi_\epsilon'(t) = p(s\epsilon t) = q_\epsilon(s) r_\epsilon(t)$. That is, $\alpha_\epsilon(s) P_\epsilon = q_\epsilon(s) R_\epsilon$. Thus $\alpha_\epsilon(s) = q_\epsilon(s) B_\epsilon$. This proves that $\mathfrak{C}(\alpha_\epsilon) \subset \mathfrak{C}_\epsilon$. Thus (c 1) holds.

Finally from (1.1) and from the relation (1.4) of [2], we get, for all t,

$$B_\epsilon A_{\epsilon\mu} \pi_\mu'(t) = B_\epsilon \pi_\epsilon'(\mu t) = r_\epsilon(\mu t) = M_{\epsilon\mu} r_\mu(t) = M_{\epsilon\mu} B_\mu \pi_\mu'(t).$$

Since $\mathfrak{C}(\pi_\mu)$ has dimension $n(\mu)$ we see that $B_\epsilon A_{\epsilon\mu} = M_{\epsilon\mu} B_\mu$. Thus (c 2) holds. We have proved that III $\Rightarrow$ I.

(c) By taking the duals of all the cones involved, we see that $\{Y_n\}$ satisfies the Condition (c) if, and only if, for every ϵ, there is a convex polyhedral cone $\mathfrak{D}_\epsilon$ such that

(c 1)$'$: $\mathfrak{C}(\pi_\epsilon) \subset \mathfrak{D}_\epsilon \subset [\mathfrak{C}(\alpha_\epsilon)]^+$, for every ϵ; and

(c 2)$'$: $\gamma_\mu (A_{\epsilon\mu})'$ belongs to $\mathfrak{D}_\epsilon$ for every ϵ, μ and for every γ_μ in $\mathfrak{D}_\mu$.

(d) Suppose I holds. We will use the cones $\mathfrak{D}_\epsilon$ introduced above. Let $\mathfrak{D}_\epsilon$ be generated by the non-zero vectors $\gamma_{\epsilon j}$, $(j = 1, \cdots, N(\epsilon))$ and let C_ϵ be the $N(\epsilon) \times n(\epsilon)$ matrix whose jth row is $\gamma_{\epsilon j}$. Condition (c 1)$'$ and Lemma 1.1 of [2] show that C_ϵ has rank $n(\epsilon)$.

Lemma 1 and (c 1)$'$ imply that the vector $\pi_\epsilon(\phi)$ is in the interior of $\mathfrak{D}_\epsilon$. Hence there are positive constants $\lambda_{\epsilon j}$, $(j = 1, \cdots, N(\epsilon))$, such that $\pi_\epsilon(\phi) = \sum_j \lambda_{\epsilon j} \gamma_{\epsilon j}$. Since the γ's are unique only up to positive multiplicative constants we can replace them by $\lambda\gamma$'s and have

$$\pi_\epsilon(\phi) = \sum_{j=1}^{N(\epsilon)} \gamma_{\epsilon j}. \tag{2.3}$$

Define $q_{\epsilon j}(s) = (\gamma_{\epsilon j}, \alpha_\epsilon(s))$, for all s. Denote $q_{\epsilon j}(\phi)$ by $m_{\epsilon j}$. Following the same lines as the proof of Lemma (1.2) of [2] we can show that each $m_{\epsilon j}$ is positive. Taking inner product of (2.3) with $\alpha_\epsilon(s)$, we get

$$p(s\epsilon) = \sum_{j=1}^{N(\epsilon)} q_{\epsilon j}(s). \tag{2.4}$$

The substitution $s = \phi$ in (2.4) shows that $p(\epsilon) = \sum_j m_{\epsilon j}$. Thus the vector $\mathbf{m}$ of $N = \sum_\epsilon N(\epsilon)$ elements formed from the m's defines a probability distribution.

Condition (c 2)$'$ shows that $\gamma_{\mu k}(A_{\epsilon\mu})'$ belongs to $\mathfrak{D}_\epsilon$. Therefore there are non-negative constants $m_{\epsilon j, \mu k}$ such that

$$\gamma_{\mu k}(A_{\epsilon\mu})' = \sum_{j=1}^{N(\epsilon)} \gamma_{\epsilon j} m_{\epsilon j, \mu k}.$$

Post-multiplying by $\alpha_\epsilon'(s)$ and using Lemma (1.3) of [2] we get

$$q_{\mu k}(s\epsilon) = \sum_{j=1}^{N(\epsilon)} q_{\epsilon j}(s) m_{\epsilon j,\mu k} . \tag{2.5}$$

Putting $s = \phi$ and summing over ϵ, we have

$$m_{\mu k} = \sum_{\epsilon=0}^{D-1} \sum_{j=1}^{N(\epsilon)} m_{\epsilon j} m_{\epsilon j,\mu k} . \tag{2.6}$$

Define $\hat{M}$ to be the $N \times N$ matrix for which the (k, j)th element in the (μ, ϵ)th submatrix $\hat{M}_{\mu\epsilon}$ is $\hat{m}_{\mu k,\epsilon j} = m_{\epsilon j} m_{\epsilon j,\mu k}/m_{\mu k}$. Then (2.6) shows that $\hat{M}$ is a transition matrix. Suppose F denotes the $N \times N$ diagonal matrix for which the (j, j)th element in $F_{\epsilon\epsilon}$ is $m_{\epsilon j}$. Then, writing $m_{\epsilon j,\mu k}$ in terms of $\hat{m}_{\mu k,\epsilon j}$, we get from (2.5)

$$q_\mu'(s\epsilon) = F_{\mu\mu}\hat{M}_{\mu\epsilon}F_{\epsilon\epsilon}^{-1}q_\epsilon'(s), \tag{2.7}$$

where $q_\epsilon(s)$ denotes the row vector whose jth element is $q_{\epsilon j}(s)$. Since $F_{\epsilon\epsilon}^{-1}q_\epsilon'(\phi) = e_\epsilon = (1, \cdots, 1)'$, it follows from (2.7) that

$$q_\mu'(\epsilon_n \cdots \epsilon_1) = F_{\mu\mu}\hat{M}_{\mu\epsilon_1} \cdots \hat{M}_{\epsilon_{n-1}\epsilon_n} e_{\epsilon_n} . \tag{2.8}$$

It is now convenient to assume that $\{Y_n\}$ is defined for $-\infty < n < \infty$ rather than just for $n \geqq 1$. This involves no loss of generality because we are interested only in distribution problems. Let $\hat{Y}_n = Y_{-n}$. From (2.4) and (2.8) it is clear that $\{\hat{Y}_n\}$ is a function of a Markov chain $\{Z_n\}$ with transition matrix $\hat{M}$ and with $\mathbf{m}$ as the distribution of Z_0 . This Markov chain need not be stationary. Define $\mathbf{m}^*$ by

$$\mathbf{m}^* = \lim_{N\to\infty}(1/N)\sum_{k=1}^{N} \mathbf{m}\hat{M}^k.$$

Then $\mathbf{m}^*$ is a stationary initial distribution for $\hat{M}$ and the stationarity of $\{Y_n\}$ shows that $\{\hat{Y}_n\}$ is a function of a stationary Markov chain $\{\hat{X}_n\}$ with transition matrix $\hat{M}$ and with $\mathbf{m}^*$ as the common distribution of each $\hat{X}_n$. If $X_n = \hat{X}_{-n}$, then it follows that $\{Y_n\}$ is a function of the finite stationary Markov chain $\{X_n\}$. We will use asterisks to denote quantities connected with this last functional relationship. Using (2.8), we get

$$\begin{aligned} q_\mu^{*\prime}(\epsilon_n \cdots \epsilon_1) &= F_{\mu\mu}^*\hat{M}_{\mu\epsilon_1} \cdots \hat{M}_{\epsilon_{n-1}\epsilon_n}e_{\epsilon_n} \\ &= F_{\mu\mu}^*F_{\mu\mu}^{-1}q_\mu'(\epsilon_n \cdots \epsilon_1). \end{aligned} \tag{2.9}$$

But the definitions of $q_{\epsilon j}(s)$ and C_ϵ show that $q_\mu'(s) = C_\mu\alpha_\mu'(s)$. Further C_μ has rank $n(\mu)$. Therefore the linear span of $\{q_\mu(s), \text{all } s\}$ has rank $n(\mu)$ at the most. Now (2.9) shows that $k^*(\mu) \leqq n(\mu)$. From (1.5) we get $k^*(\mu) = n(\mu)$. This proves that I $\Rightarrow$ II.

(e) Suppose II holds. We will use the matrices P_ϵ , Q_ϵ and R_ϵ introduced in part (*b*) of this proof. Let $\mathfrak{D}_\epsilon$ be the convex polyhedral cone generated by the rows of Q_ϵ'. The relation $\pi_\epsilon'(t) = Q_\epsilon r_\epsilon(t)$ or, equivalently, $\pi_\epsilon(t) = r_\epsilon'(t)Q_\epsilon'$ shows that $\mathfrak{C}(\pi_\epsilon) \subset \mathfrak{D}_\epsilon$.

Since Q_ϵ has rank $n(\epsilon)$ and since $k(\epsilon) = n(\epsilon)$, the rows of Q_ϵ span K_ϵ . Therefore, for each s, there is a unique vector $\alpha_\epsilon(s)$ of $n(\epsilon)$ elements such that

$$q_\epsilon(s) = \alpha_\epsilon(s)Q_\epsilon . \tag{2.10}$$

This vector $\alpha_\epsilon(s)$ must be the same as the vector with the same notation used before, because post-multiplying (2.10) by $r_\epsilon(t)$, we get, for all t, $p(s\epsilon t) = \alpha_\epsilon(s)\pi_\epsilon'(t)$. Since $q_\epsilon(s) \geqq 0$, (2.10) shows that $\mathfrak{D}_\epsilon \subset [\mathfrak{C}(\alpha_\epsilon)]^+$. Thus (c 1)′ is satisfied.

Now from (1.1) and (2.10)

$$q_\epsilon(s_{\epsilon i})M_{\epsilon\mu} = q_\mu(s_{\epsilon i}\epsilon) = \alpha_\mu(s_{\epsilon i}\epsilon)Q_\mu .$$

Therefore $Q_\epsilon M_{\epsilon\mu} = A_{\epsilon\mu}Q_\mu$ or $Q_\mu'(A_{\epsilon\mu})' = (M_{\epsilon\mu})'Q_\epsilon'$. Thus (c 2)′ also holds. This shows that II $\Rightarrow$ I and completes the proof of the theorem.

A stationary process $\{Y_n\}$ which can be expressed as a function of a finite Markov chain in such a way that $n(\epsilon) = N(\epsilon)$ for all ϵ was termed a regular function of a Markov chain by Gilbert [3]. It was shown in [2] that such regular functions are in $\mathfrak{F}$ (i.e. have $\sum_\epsilon n(\epsilon) < \infty$ and satisfy the Conditions (c)). This also follows from the preceding theorem, because, in view of (1.5), the condition $n(\epsilon) = N(\epsilon)$ implies that $k(\epsilon) = l(\epsilon) = n(\epsilon)$. A question arises whether $\mathfrak{F}$ includes some non-regular functions also. We will show that the answer is in the affirmative. We need a simple lemma.

LEMMA 2. *If $\{Y_n\}$ can be expressed as a function of a finite Markov chain in such a way that $k(\epsilon) = l(\epsilon) = N(\epsilon)$, then $n(\epsilon) = N(\epsilon)$.*

PROOF. If $k(\epsilon) = l(\epsilon) = N(\epsilon)$, then we can find s_i, t_i, $(i = 1, \cdots, N(\epsilon))$, such that the $q_\epsilon(s_i)$'s and the $r_\epsilon(t_j)$'s are linearly independent. It follows from (1.3) that the $N(\epsilon) \times N(\epsilon)$ matrix whose (i, j)th element is $p(s_i\epsilon t_j)$ is non-singular. Hence $n(\epsilon) = N(\epsilon)$. This proves the lemma.

We have constructed before ([1], Section 3 and [2], Section 4) a 2-state stationary process $\{Y_n\}$ such that (a) $n(0) = 3$ and $n(1) = 1$; (b) $\{Y_n\}$ is not a regular function of a Markov chain; and (c) $\{Y_n\}$ is a function of a Markov chain $\{X_n\}$ with 5 states in such a way that $N(0) = 4$ and $N(1) = 1$. For this functional relationship suppose $k(0) = l(0) = 4$. Then Lemma 2 shows that $n(0) = 4$, which is false. Thus either $k(0) = 3$ or $l(0) = 3$. In any event $\{Y_n\}$ belongs to $\mathfrak{F}$. This proves that $\mathfrak{F}$ is wider than the class of regular functions.

The results of this paper pose the following question. Does $\mathfrak{F}$ exhaust all functions of finite Markov chains? The answer is not known.

REFERENCES

[1] DHARMADHIKARI, S. W. (1963). Functions of finite Markov chains. *Ann. Math. Statist.* **34** 1022–1032.

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