ON STOCHASTIC PROCESSES DERIVED FROM MARKOV CHAINS[1]

By Alex Heller

University of Illinois

0. Introduction. If $X_0, X_1, \cdots$ is a stochastic process with finite state space S, T is a finite set and $f: S \rightarrow T$ then $fX_0, fX_1, \cdots$ is a stochastic process with state space T, often referred to as a "functional" of $X_0, X_1, \cdots$. Our object here is to characterize processes which are functionals of Markov chains.

This problem was considered by E. J. Gilbert [4] who adduced a necessary condition. S. W. Dharmadikari [1], [2], [3] showed that Gilbert's condition was not sufficient and provided a sufficient condition but, though he has given an excellent analysis of the problem, did not complete the characterization. We do this here (Theorem 5.1).

Dharmadikari has indicated the essentially geometrical nature of the problem. With a stochastic process he associates vector spaces on which the states operate linearly. His conditions (as well as ours) have to do with invariant convex cones in these spaces. We have chosen here to regard such spaces as modules over the free associative algebra generated by the state space S (i.e. the algebra of polynomials in the noncommuting variables $x \varepsilon S$). This point of view, still unconventional perhaps in probability theory, seems indicated by the fact that the vector spaces in question lack preferred bases, so that linear transformations are not naturally represented by matrices. We believe the argument is made simpler and more conceptual by its adoption.

To each stochastic process with finite state space we associate canonically a module; the process is then discussed in terms of this module. If the module is finite dimensional (this is Gilbert's necessary condition) it is reasonable to say that the process is characterized by finitely many parameters: this condition obviously defines an interesting class of processes. We suggest that the proper apparatus for the discussion of such processes is the one introduced here.

The author would like to thank Professor Joseph Jaffe of the College of Physicians and Surgeons, Columbia University, for bringing the problem to his attention.

1. Stochastic S-modules. We shall be concerned here with stochastic processes with discrete time and finite state space; "stochastic process" is used below in this sense only. Such a process, with state space S, is completely specified by the probabilities $p(x_1, \cdots, x_n)$ of the finite sequences $x_1, \cdots, x_n$ in S.

It is more convenient for our purposes to describe these processes in the following way. Let A_S be the free associative $\mathbf{R}$-algebra generated by S. We then

Received 2 October 1964; revised 3 February 1965.

[1] This work was in part supported by a contract with the National Science Foundation.

write $p(x_1 \cdots x_n)$ instead of $p(x_1, \cdots, x_n)$, set $p(1) = 1$ and, observing that p is now defined on an **R**-basis of A_S, extend p linearly to $p: A_S \to \mathbf{R}$.

We denote by P_S the *coordinate cone* of A_S, consisting of polynomials with nonnegative coefficients, and we let $\sigma = \sum_{x \varepsilon S} x$. Then an **R**-linear $p: A_S \to \mathbf{R}$ is a stochastic process if and only if the conditions

$$\text{(P0)} \qquad p(1) = 1$$

$$\text{(P1)} \qquad p(P_S) \subset [0, \infty)$$

$$\text{(P2)} \qquad \text{for all } \xi \,\varepsilon\, A_S,\ p(\xi\sigma) = p\xi$$

hold.

Now suppose L is a left A_S-module, $l_0 \,\varepsilon\, L$ and $q: L \to \mathbf{R}$ (linear). Then $p\xi = q(\xi l_0)$ defines a linear $p: A_S \to \mathbf{R}$.

LEMMA 1.1. *p is a stochastic process if and only if*

(i) $ql_0 = 1$

(ii) $q(P_S l_0) \subset [0, \infty)$

(iii) *for all* $\xi \,\varepsilon\, A_S$, $q(\xi(\sigma - 1)l_0) = 0$.

If these conditions are satisfied we shall say that the triple (L, q, l_0) is a *stochastic S-module* (sS-module), and the stochastic process p just defined is *associated with* (L, q, l_0).

A *morphism* of stochastic S-modules (L, q, l_0), (L', q', l_0') is a homomorphism $\varphi: L \to L'$ of left A_S-modules such that $q'\varphi = q$, $\varphi l_0 = l_0'$; *isomorphisms* are defined accordingly as invertible morphisms. The existence of a morphism of sS-modules clearly implies that the associated stochastic processes are the same.

We shall say that an sS-module (L, q, l_0) is *reduced* if (i) L is cyclic with generator l_0, i.e., $A_S l_0 = L$, (ii) L has no nonzero submodules L' with $q(L') = 0$.

PROPOSITION 1.2. *Any stochastic process $p: A_s \to \mathbf{R}$ is associated with a reduced sS-module; any two such modules are isomorphic.*

For let $N \subset A_S$ be the left ideal $N = \{\xi \mid p(A_S\xi) = 0\}$, i.e., the largest left ideal in the kernel of p. Let $L = A_S/N$ and let $\lambda: A_S \to L$ be the canonical map. Since p vanishes on N there is a unique linear $q: L \to \mathbf{R}$ with $p = q\lambda$. It remains only to set $l_0 = \lambda(1)$. So defined, L is certainly cyclic. To see that (L, q, l_0) is reduced it is only necessary to observe that if $L' \subset L$ is a submodule then $\lambda^{-1}(L')$ is a left ideal; if further $q(L') = 0$ then $p\lambda^{-1}(L') = 0$ so that $\lambda^{-1}(L') \subset N$, $L' = 0$.

On the other hand all reduced sS-modules must arise in just this way, with $\xi \to \xi l_0$ playing the role of the map λ. This observation leads immediately to the proof of the second statement.

In view of this result we may permit ourselves to speak of the reduced sS-module of a stochastic process.

Finally, suppose (L, q, l_0) is any sS-module. The reduced sS-module of the associated stochastic process is clearly obtained in the following way. First replace L by the cyclic submodule $A_S l_0$; then divide by the largest submodule on which q vanishes.

LEMMA 1.3. *If (L, q, l_0) is a reduced sS-module then $\sigma l_0 = l_0$.*

For $A_S(\sigma - 1)l_0$ is a submodule in the kernel of q.

A state x of a stochastic process $p: A_S \to \mathbf{R}$ is *prohibited* if for any $x_1, \cdots, x_n, x_1', \cdots, x_m'$, $p(x_1 \cdots x_n x x_1' \cdots x_m') = 0$, i.e., if any sequence of states containing x has probability zero.

LEMMA 1.4. *Let (L, q, l_0) be a reduced sS-module, and suppose $x \varepsilon S$. Then the following are equivalent:*

(i) $xl_0 = 0$

(ii) *x is a prohibited state of the associated stochastic process p*

(iii) $xL = 0$.

For the implication (i) $\Rightarrow$ (ii), we have

$$0 \leqq p(x_1 \cdots x_n x x_1' \cdots x_m') \leqq p(x_1 \cdots x_n x \sigma^m) = q(x_1 \cdots x_n x l_0) = 0.$$

2. Induced stochastic processes. If $p: A_S \to \mathbf{R}$ is a stochastic process and $f: S \to S'$ (S' being finite) then a stochastic process $p': A_{S'} \to \mathbf{R}$ is defined by the formulae

$$p'(y_1 \cdots y_n) = \textstyle\sum_{fx_1 = y_1} \cdots \sum_{fx_n = y_n} p(x_1 \cdots x_n). \tag{2.1}$$

The process p' is sometimes referred to as a "functional" or "function" of p; this seems a bit misleading, and we shall call it the process *induced from p by f.*

To see how this is reflected in the theory of stochastic modules we introduce the following notation: if $\varphi: A \to B$ is a homomorphism of rings and L is a left B-module then ${}_{[\varphi]}L$ is the left A-module whose underlying abelian group is that of L, the operation being given by $\alpha l = (\varphi\alpha)l$. Further, for $f: S \to S'$, as above, we define $f^*: A_{S'} \to A_S$ to be the homomorphism given on the free generators $y \varepsilon S'$ of $A_{S'}$ by $f^* y = \sum_{fx=y} x$.

PROPOSITION 2.2. *Let (L, q, l_0) be an sS-module and suppose $f: S \to S'$. Then $({}_{[f^*]}L, q, l_0)$ is an sS'-module and the stochastic process associated with $({}_{[f^*]}L, q, l_0)$ is induced by f from that associated with (L, q, l_0).*

This is really no more than a restatement of (2.1), which may be written $p'(y_1 \cdots y_n) = pf^*(y_1 \cdots y_n)$.

We may at this point define a *finitary* stochastic process as one whose reduced sS-module has finite dimension (over $\mathbf{R}$). Clearly any process associated with a finite-dimensional module is finitary.

PROPOSITION 2.3. *A stochastic process induced from a finitary process is itself finitary.*

3. Markov chains. We shall discuss here only Markov chains with stationary transition probabilities and shall accordingly omit the phrase "with stationary transition probabilities" henceforth.

A stochastic process $p: A_S \to \mathbf{R}$ is a *Markov chain* if there is a map $t: S \times S \to \mathbf{R}$, the *transition matrix*, such that for any sequence $x_1, \cdots, x_n$ in S

$$p(x_1 \cdots x_n) = (px_1)t(x_1, x_2)t(x_2, x_3) \cdots t(x_{n-1}, x_n). \tag{3.1}$$

We may also state this in the following equivalent form: if $\xi \varepsilon A_S$, $x, y \varepsilon S$ then

$$p(\xi(xy - t(x, y)x)) = 0. \tag{3.1$'$}$$

The transition matrix may always be taken nonnegative and stochastic, i.e. $\sum_y t(x, y) = 1$.

PROPOSITION 3.2. *If (L, q, l_0) is a reduced sS-module then the associated stochastic process p is a Markov chain if and only if for each $x \varepsilon S$, xL has dimension* $\leqq 1$.

If p is a Markov chain then (3.1$'$) asserts that q vanishes on the submodule $A_S(xy - t(x, y)x)l_0$ for any $x, y \varepsilon S$. Since (L, q, l_0) is reduced we have $xyl_0 = t(x, y)xl_0$. But $A_S l_0 = L$, and we see immediately that $xL \subset \mathbf{R}xl_0$.

Conversely, in view of Lemma 1.4, we may write for $x, y \varepsilon S$ $xyl_0 = t(x, y)xl_0$, for some $t(x, y) \varepsilon \mathbf{R}$. We then have, for $\xi \varepsilon A_S$, $q(\xi(xy - t(x, y)x)l_0) = 0$, which of course is just (3.1$'$).

COROLLARY 3.3. *If (L, q, l_0) is an sS-module and for each $x \varepsilon S$,* $\dim xL \leqq 1$ *then the associated stochastic process is a Markov chain.*

For the reduced module obviously shares the property. We shall call such modules *Markovian.*

COROLLARY 3.4. *A process induced from a Markov chain is finitary.*

We may indeed sharpen the last result as follows. Suppose $f: S \to S'$ and denote by $\mu(y)$, for $y \varepsilon S'$, the number of x such that $fx = y$.

PROPOSITION 3.5. *If (L', q', l_0') is the reduced S'-module of a stochastic process induced by f from a Markov chain on S then for each $y \varepsilon S'$,* $\dim y'L' \leqq \mu(y)$.

This is, essentially, the result of [4]. We leave the proof for the reader.

For application below we define the *regular module* of a Markov chain $p: A_S \to \mathbf{R}$ with transition matrix t as follows. Let L be a vector space with basis $\{[x] \mid x \varepsilon S\}$ in bijective correspondence with S, and define an operation of A_S on L by $x[y] = t(x, y)[x]$ for $x, y \varepsilon S$. Let $l_0 = \sum_x [x]$ and define $q: L \to \mathbf{R}$ by $q[x] = p(x)$. Then (L, q, l_0) is the regular module of p.

PROPOSITION 3.6. *The stochastic process associated with (L, q, l_0) is p.*

If $x \varepsilon S$, then $xl_0 = \sum_y t(x, y)[x] = [x]$. If $x_1, \cdots, x_n$ is a sequence in S then, inductively, $x_1 \cdots x_n l_0 = t(x_1, x_2) \cdots t(x_{n-1}, x_n)[x_1]$ and thus $q(x_1 \cdots x_n l_0) = p(x_1)t(x_1, x_2) \cdots t(x_{n-1}, x_n)$ as required.

4. Cones. By a *cone* in a real vector space V we mean a union of rays from the origin. A convex cone C is *strongly convex* if it contains no line through the origin, i.e., if $x, -x \varepsilon C$ imply $x = 0$. A convex cone C is *polyhedral* if it is the convex hull of the union of finitely many rays, or, equivalently, the intersection of finitely many half-spaces.

A subspace $W \subset V$ intersects a cone $C \subset V$ *extremally* if $x, y \varepsilon C$, $x + y \varepsilon W$ imply $x, y \varepsilon W$.

The reader will readily supply the proof of the following observations.

(4.1) If $W \subset V$ is a subspace and $C \subset V$ a strongly convex cone then $W \cap C$ is a strongly convex cone.

(4.2) if $W \subset V$ is a subspace and $C \subset V$ a polyhedral cone then so is $W \cap C$.

(4.3) if $W, W' \subset V$ are subspaces and W' intersects the cone $C \subset V$ extremally then also $W \cap W'$ intersects $W \cap C$ extremally.

(4.4) if $W \subset V$ is a subspace, $\eta: V \to V/W$ the canonical map and $C \subset V$ is a polyhedral cone then so is $\eta C \subset V/W$.

(4.5) if $W \subset V$ is a subspace, $\eta: V \to V/W$ the canonical map, $C \subset V$ a strongly convex cone and W intersects C extremally then ηC is strongly convex.

5. Processes induced from Markov chains. Our principal result is the following:

THEOREM 5.1. *Let* (L, q, l_0) *be a reduced sS-module. The associated stochastic process is induced from a Markov chain if and only if there is a cone* $C \subset L$ *such that* (i) $l_0 \varepsilon C$, (ii) $q(C) \subset [0, \infty)$, (iii) C *is invariant under* P_S, i.e., $P_S C \subset C$, (iv) C *is strongly convex and polyhedral.*

Suppose our process is induced by a map $f: S' \to S$ from a Markov chain with regular sS'-module (L', q', l_0'). Then the nonnegative orthant $C' \subset L'$ certainly has, with respect to S', the Properties (i)–(iv). If we set $L_1 = A_S l_0'$ in the A_S-module ${}_{[f^*]}L$ then, since $f^*(P_S) \subset P_{S'}$, the cone $L_1 \cap C'$ has by (4.1), (4.2), the Properties (i)–(iv) with respect to S.

Now if $N_1 \subset L_1$ is the largest submodule on which q' vanishes L may be identified with L_1/N_1, with q', l_0' going into q, l_0. But N_1 intersects $L_1 \cap C'$ extremally. For suppose $u, v \varepsilon L_1 \cap C'$, $u + v \varepsilon N_1$. For any $\xi \varepsilon P_S$ we have $\xi u \varepsilon L_1 \cap C'$ and $\xi(u + v) \varepsilon N_1$ so that $0 \leqq q'(\xi u) \leqq q'(\xi u + \xi v) = 0$. Thus u, and similarly v, are in N_1.

Thus the image of $L_1 \cap C'$ in $L = L_1/N_1$, by (4.4), (4.5), satisfies Conditions (i)–(iv).

Conversely, let $C \subset L$ be a cone satisfying (i)–(iv) and let $U \subset C$ be a finite subset such that C is the convex hull of the rays through elements of U. Then we may write

$$l_0 = \sum\nolimits_{u \varepsilon U} \lambda_u u, \qquad \lambda_u \geqq 0$$

$$xu = \sum\nolimits_{v \varepsilon U} \alpha_{xuv} v, \qquad x \varepsilon S,\ u \varepsilon U,\ \alpha_{xuv} \geqq 0.$$

We define $S' = S \times U$ and let L' be a vector space with base $\{[x, u]\}$ in bijective correspondence with S', and we make L' a left $A_{S'}$-module by setting

$$(x, u)[y, v] = \alpha_{xvu}[x, u] \qquad x, y \varepsilon S,\ u, v \varepsilon U.$$

If n is the number of elements in S we define

$$l_0' = n^{-1} \sum\nolimits_{x \varepsilon S, u \varepsilon U} \lambda_u [x, u]$$

and determine $q': L' \to \mathbf{R}$ by $q'[x, u] = qu$.

In view of 3.3, (L', q', l_0') is an sS'-module whose associated stochastic process is a Markov chain.

Finally let $f: S' \to S$ be the projection. We claim that the process induced by f from the one just defined is associated with (L, q, l_0). To see this define $g: L' \to L$ as the linear extension of $g[x, u] = u$. Then $gl_0' = l_0$ and $qg = q'$. Further $g: {}_{[f^*]}L' \to L$ is a homomorphism of A_S-modules, for if $x \varepsilon S$ then

$$\begin{aligned} g\{(f^{*}x)[y, v]\} &= g\{\textstyle\sum_u (x, u)[y, v]\} \\ &= g\{\textstyle\sum_u \alpha_{xvu}[x, u]\} \\ &= \textstyle\sum_u \alpha_{xvu} u = xv = xg[y, v]. \end{aligned}$$

Thus g is a morphism of sS-modules, and the theorem is proved.

6. An example. We apply here the criterion of Theorem 5.1 to construct a finitary stochastic process which is not induced from a Markov chain (cf. also [1]).

Let L be Euclidean 3-space with orthonormal basis $\{e_0, e_1, e_2\}$. Let $l_0 = e_0 + \epsilon e_1$ with ϵ small positive, and define q to be the inner product with $|l_0|^{-2} l_0$, so that $q_0 l_0 = 1$.

For S we take the set of two elements x, y and make L into an A_S-module as follows: x acts on L as $\frac{1}{2}\theta$ where θ is a rotation about e_0 through a small angle which is an irrational multiple of π, while y acts on L as $\alpha\rho$, where ρ is the orthogonal projection on $l_0 - xl_0$, and $\alpha \varepsilon \mathbf{R}$ is chosen so that $\sigma l_0 = (x + y)l_0 = l_0$.

The closure of $P_S l_0$ is clearly the right circular cone with axis e_0 and element $l_0 - xl_0$; thus if ϵ is small enough we have $q(P_S l_0) \subset [0, \infty)$. Thus (L, q, l_0) is a stochastic S-module which is irreducible, even as an A_S-module.

On the other hand, the only P_S-invariant cones in L are the right circular cones with axis e_0. If these are to be strongly convex, i.e., distinct from the half space, it is clear they cannot be polyhedral.

PROPOSITION 6.1. *The stochastic process associated with (L, q, l_0) is not induced from a Markov chain.*

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] DHARMADHIKARI, S. W. (1965). A characterization of a class of functions of finite Markov chains. *Ann. Math. Statist.* **36** 524–528.

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