The Annals of Mathematical Statistics
1968, Vol. 39, No. 1, 164–170

SOME PROPERTIES OF AN ALGEBRAIC REPRESENTATION OF STOCHASTIC PROCESSES[1]

BY PAUL W. HOLLAND

Harvard University

0. Summary. Heller's (1965) concept of a stochastic module is examined with the purpose of seeing what are the algebraic implications of various probabilistic properties of discrete time, finite state processes. kth order Markovity, recurrence, stationarity, and ergodicity are given characterizations in terms of the stochastic module.

1. Introduction. Heller (1965) introduced the concept of a stochastic module associated with a stochastic process in order to obtain a characterization of those processes which are functions of Markov chains. Briefly, he associates with every finite state, discrete time stochastic process (hereafter referred to simply as a process) a certain algebraic object which he calls a stochastic module. He then examines how certain *stochastic* properties of the process are reflected by corresponding *algebraic* properties of the stochastic module. The object of this paper is to pursue this program in a limited way and to examine some additional properties of processes and their corresponding interpretation in terms of associated stochastic modules. Due to the novelty of Heller's approach, in this section the basic ideas and elementary results of his paper will be stated to provide the necessary language requirement for the rest of this paper. The reader is referred to Heller's article for the proofs. I deviate slightly from his notation.

Let $\{X_n, n \geqq 1\}$ be a process as defined above and S be its (finite) set of states. Now suppose we can find (i) a real vector space L, (ii) a linear functional q on L, (iii) an element $e \in L$ and (iv) a set of linear operators on L into L, $\{T_x, x \in S\}$, such that for all $x_1, \cdots, x_n \in S$ and $n = 1, 2, \cdots$

$$P\{X_1 = x_1, \cdots, X_n = x_n\} = q(T_{x_1} \cdot T_{x_2} \cdot \cdots \cdot T_{x_n} e). \tag{1}$$

The triple (L, q, e) is called a *stochastic S-module* associated with the process $\{X_n\}$ and we identify T_x with $x \in S$, i.e. we write xl for $T_x l$ for $l \in L$. Heller shows that every process has a stochastic S-module associated with it. Actually there are many stochastic S-modules associated with a given process so that to obtain a unique one they must be restricted further. Herein we shall call them simply S-modules. Because we shall be viewing the states of the process as linear operators on some vector space and because one may naturally consider linear combinations of products of linear operators we shall want to do the same thing with the elements of S. Hence we let A_s be the free associative real algebra generated

Received 11 April 1967.
[1] This work was facilitated by a grant from the National Science Foundation (GS-341).

by S, i.e. the set of all formal linear combinations of formal products of elements of S. We identify A_s with the algebra of operators generated by $\{T_x, x \varepsilon S\}$ in the natural way. The empty product is denoted by 1 and the special linear combination $\sum_{x \varepsilon S} x$ by σ. The vector space L may be viewed as a module over A_s, consequently the name. We let $U_s \subseteq A_s$ be the set of all products of elements of S.

It is easy to verify that if (L, q, e) is an S-module associated with a process then

(a) $q(e) = 1$,
(b) $q(\pi e) \geqq 0, \qquad \pi \varepsilon U_s$,
(c) $q(\xi(\sigma - 1)e) = 0, \qquad \xi \varepsilon A_s$.

Conversely if (L, q, e) is a vector-space—linear functional—vector triple and L is acted on by a finite set, S, of linear operators and if (a), (b), and (c) are satisfied then via

$$P\{X_1 = x_1, \cdots, X_n = x_n\} = q(x_1 \cdots x_n e) \tag{2}$$

a process $\{X_n\}$ is defined with state space S and (L, q, e) is an S-module associated with it. Any such triple (L, q, e) satisfying (a), (b), and (c) for some set, S, of linear operators will be called an S-module, and the process $\{X_n\}$ defined by (2) its induced process.

We say an S-module is *reduced* if

(d) $L = A_s e$,
(e) $q(A_s l) = 0 \Rightarrow l = 0 \quad$ for all $\quad l \varepsilon L$.

Furthermore two S-modules (L, q, e) and (L', q', e') are homomorphic if there is a linear map $\phi: L \rightarrow L'$ such that $\phi(e) = e'$ and $q = q' \circ \phi$. They are isomorphic if ϕ is an isomorphism. Homomorphic S-modules induce the same process. Heller shows that up to isomorphism there is one and only one *reduced* S-module associated with every process. Sometimes it is convenient to use S-modules which are not reduced but which still induce a particular process. From such an S-module one can always construct a reduced S-module inducing the same process by replacing L by $A_s e$ and then forming the quotient space of this by its largest subspace N for which $A_s N \subseteq N$ and $q(N) = 0$. Then e is replaced by its equivalence class $[e]$ and q is replaced by the factor map of q under the natural map $l \rightarrow [l]$. T_x is defined by the usual maneuver since it does not split these equivalence classes.

Since to every process there can be associated essentially one and only one reduced S-module it follows that every property of the process is reflected by some corresponding property of its reduced S-module. For example a state $x \varepsilon S$ is *prohibited* if it can never be achieved by the process. In the reduced S-module (L, q, e) this is equivalent to

$$xL = 0.$$

Similarly, $\{X_n\}$ is a 1st order Markov chain if there is a map $t: S \times S \rightarrow \mathbf{R}$ such that

$$P\{X_1 = x_1, \cdots, X_n = x_n, X_{n+1} = x_{n+1}\}$$
$$= t(x_n; x_{n+1})P\{X_1 = x_1, \cdots, X_n = x_n\}.$$

Heller shows that in the reduced S-module this is equivalent to

$$xL = \mathbf{R}xe \quad \text{for all} \quad x \varepsilon S,$$

that is, the image of L under each operator in S is at most one dimensional.

We observe that since $q(e) = 1$, $e \neq 0$ and hence L is at least one dimensional. In general L is at most countably infinite dimensional. If in the reduced S-module L is finite dimensional the corresponding process is said to be of finite rank. Finally, from (c) and (e) we see that for a reduced S-module $\sigma e = e$.

2. Independence and Markovity. We begin by asking the simplest question, namely, when is the reduced S-module one dimensional? This means $L = \mathbf{R}e$ since $\mathbf{R}e \subseteq L$. The first theorem answers this question and the method of proof is a good illustration of how (d) and (e) are exploited.

THEOREM 1. *If (L, q, e) is the reduced S-module of $\{X_n\}$ then the process is a sequence of independent and identically distributed random variables if and only if $L = \mathbf{R}e$.*

PROOF. Assume $\{X_n\}$ are independent and identically distributed, then

$$P\{X_1 = x_1, \cdots, X_n = x_n, X_{n+1} = x_{n+1}, \cdots, X_{n+m} = x_{n+m}\}$$
$$= P\{X_1 = x_1, \cdots, X_n = x_n\}P\{X_1 = x_{n+1}, \cdots, X_m = x_{n+m}\}.$$

Then for all π, $\pi' \varepsilon U_s$ we have

$$q(\pi\pi'e) = q(\pi e)q(\pi'e). \tag{3}$$

Let $c = q(\pi'e) \varepsilon \mathbf{R}$, then (3) may be written

$$q(\pi(\pi' - c)e) = 0 \quad \text{for all} \quad \pi \varepsilon U_s$$

so that

$$q(A_s(\pi' - c)e) = 0$$

or

$$\pi'e = ce \varepsilon \mathbf{R}e.$$

Hence $A_s e \subseteq \mathbf{R}e$ or $L \subseteq \mathbf{R}e$. Conversely suppose $L = \mathbf{R}e$. Then if $y \varepsilon S$

$$ye = c_y e \quad \text{for some} \quad c_y \varepsilon \mathbf{R}$$
$$q(ye) = c_y q(e) = c_y = P\{X_1 = y\}.$$

Induction yields

$$q(x_1 \cdots x_n e) = P\{X_1 = x_1\}P\{X_1 = x_2\} \cdots P\{X_1 = x_n\}. \qquad \square$$

The next question we ask is what does kth order Markovity imply about the S-module? More precisely we say $\{X_n\}$ is a kth order Markov chain if there is a map $t: S^k \times S \to \mathbf{R}$ such that

$$\begin{aligned} P\{X_1 &= x_1, \cdots, X_{n+k+1} = x_{n+k+1}\} \\ &= t(x_{n+1}, \cdots, x_{n+k}; x_{n+k+1})P\{X_1 = x_1, \cdots, X_{n+k} = x_{n+k}\}. \end{aligned}$$

In terms of (L, q, e) this may be expressed as

$$q(\pi\pi_k xe) = t(\pi_k; x)q(\pi\pi_k e) \tag{4}$$

for all π, $\pi_k \varepsilon U_s$, $x \varepsilon S$ where π_k is a product of length k. The next theorem gives the generalization of Heller's characterization of 1st order Markov chains to kth order chains. Its proof follows from Theorem 5 in Section 3.

THEOREM 2. *If (L, q, e) is the reduced S-module of $\{X_n\}$ then the process is a kth order Markov chain if and only if for all $x_1, \cdots, x_k \varepsilon S$*

$$x_1 \cdots x_k L = \mathbf{R}x_1 \cdots x_k e.$$

The notion of a prohibited state may be naturally generalized to that of a prohibited sequence of states. A sequence of states $x_1, \cdots, x_n$ is prohibited if it is impossible for the process to ever consecutively occupy $x_1, \cdots, x_n$. In terms of (L, q, e) this is $q(\pi x_1 \cdots x_n e) = 0$ for all $\pi \varepsilon U_s$. Using the type of argument illustrated in proving Theorem 1 the following characterization of prohibited sequences may be proved.

THEOREM 3. *If (L, q, e) is the reduced S-module for $\{X_n\}$ then the sequence of states $x_1, \cdots, x_n$ is prohibited if and only if*

$$x_1 \cdots x_n L = 0.$$

Using this we may give a slightly more detailed version of Theorem 2.

THEOREM 4. *If (L, q, e) is the reduced S-module for $\{X_n\}$ then the process is a kth order Markov chain if and only if for every $x_1, \cdots, x_k \varepsilon S$ which is not prohibited*

$$\dim (x_1 \cdots x_k L) = 1.$$

Theorems 1 and 2 yield the following interesting hierarchy of conditions for Markov chains

$$\begin{array}{rcll} L &=& \mathbf{R}e & \text{0th order chain,} \\ xL &=& \mathbf{R}xe & \text{1st order chain,} \\ \vdots & & \vdots & \vdots \\ x_1 \cdots x_n L &=& \mathbf{R}x_1 \cdots x_n e & n\text{th order chain.} \end{array} \tag{5}$$

Finally we observe that being an nth order Markov chain for some n implies that the reduced S-module is finite dimensional because e, x_1e, x_1x_2e, $\cdots$, $x_1x_2 \cdots x_ne$ for all $x_1, \cdots, x_n \varepsilon S$ will span L although they need not be linearly independent.

3. Recurrent sequence of states. In this section we examine the structure of Markovity a little more closely. We say a sequence of states $x_1, \cdots, x_n$ is *recurrent* if given that the process has just executed the sequence $x_1, \cdots, x_n$ any other past information is irrelevant for the probability of an event occurring in the future. More formally in terms of (L, q, e) this may be written as

$$(6) \quad q(\pi' x_1 \cdots x_n \pi e)/q(\pi' x_1 \cdots x_n e) = q(\pi'' x_1 \cdots x_n \pi e)/q(\pi'' x_1 \cdots x_n e)$$

for all $\pi, \pi', \pi'' \varepsilon U_s$ for which $q(\pi' x_1 \cdots x_n e)$ and $q(\pi'' x_1 \cdots x_n e)$ are positive.

Letting $\rho = x_1 \cdots x_n$ (6) may be rewritten as

$$(7) \qquad q(\pi' \rho\pi e) q(\pi'' \rho e) = q(\pi'' \rho \pi e) q(\pi' \rho e)$$

which is trivially true if either $q(\pi' \rho e) = 0$ or $q(\pi'' \rho e) = 0$. This leads us to define a *recurrent sequence* by:

DEFINITION. $\rho \varepsilon U_s$ is a *recurrent sequence* if for all $\pi, \pi', \pi'' \varepsilon U_s$ equation (7) holds. This definition allows prohibited sequences to be recurrent but this presents no difficulty and from vacuity considerations they should be recurrent. The next result gives the impact of recurrence on the reduced S-module.

THEOREM 5. *If (L, q, e) is the reduced S-module of $\{X_n\}$ then $\rho = x_1 \cdots x_n$ is a recurrent sequence if and only if $\rho L = \mathbf{R}\rho e$.*

PROOF. We first suppose ρ is a recurrent sequence. If ρ is also prohibited then $\rho L = 0 = \mathbf{R}\rho e$, so we suppose there exists $\pi_0 \varepsilon U_s$ for which $q(\pi_0 \rho e) > 0$. By recurrence, for all $\pi, \pi' \varepsilon U_s$,

$$q(\pi' \rho \pi e) q(\pi_0 \rho e) = q(\pi_0 \rho \pi e) q(\pi' \rho e),$$

or

$$q[\pi'(q(\pi_0 \rho e)\rho\pi - q(\pi_0 \rho \pi e)\rho)e] = 0.$$

The usual argument gives $q(\pi_0 \rho e)\rho\pi e = q(\pi_0 \rho \pi e)\rho e$, so that $\rho \pi e \varepsilon \mathbf{R}\rho e$ for all $\pi \varepsilon U_s$ and hence $\rho L \subseteq \mathbf{R}\rho e$. Conversely suppose $\rho L = \mathbf{R}\rho e$. If $q(\pi' \rho e) = 0$, then equation (7) holds trivially for all $\pi, \pi'' \varepsilon U_s$, so suppose $q(\pi' \rho e) > 0$. Now for any $\pi \varepsilon U_s$, $\pi' \rho \pi e = c \pi' \rho e$, where $c \varepsilon \mathbf{R}$ and depends only on π,

$$c = q(\pi' \rho \pi e)/q(\pi' \rho e) \qquad \text{for all} \qquad \pi' \varepsilon U_s.$$

Hence for any $\pi'' \varepsilon U_s$,

$$\pi'' \rho \pi e = [q(\pi' \rho \pi e)/q(\pi' \rho e)]\pi'' \rho e,$$

from which (7) follows. □

In a kth order Markov chain any sequence of length k is recurrent so that Theorem 2 follows from Theorem 5.

4. Stationarity. A process $\{X_n\}$ is stationary if for all $x_1, \cdots, x_n, n$ and t

$$P\{X_{t+1} = x_1, \cdots, X_{t+n} = x_n\} = P\{X_1 = x_1, \cdots, X_n = x_n\}.$$

In terms of (L, q, e) this says that for all $\pi \varepsilon U_s$ and $t = 1, 2, \cdots$

$$q(\sigma^t \pi e) = q(\pi e),$$

which is equivalent to

$$q(\sigma\xi e) = q(\xi e) \quad \text{for all} \quad \xi \varepsilon A_s. \tag{8}$$

Equation (8) may be rewritten as $q((\sigma - 1)L) = 0$. In general $(\sigma - 1)L$ is a subspace of L but it need not be a submodule of L unless

$$A_s(\sigma - 1)L \subseteq (\sigma - 1)L. \tag{9}$$

If (9) does hold then the usual argument implies that for such a stationary process

$$\sigma l = l \quad \text{for all} \quad l \varepsilon L. \tag{10}$$

This is an interesting condition on the S-module, what does it mean for the process? If (10) holds then for all $\pi, \pi' \varepsilon U_s$,

$$q(\pi\sigma\pi' e) = q(\pi\pi' e). \tag{11}$$

This means that not only is $\{X_n\}$ stationary but conditional on $X_1, \cdots, X_t$, $\{X_{t+n}\}$ is also a stationary process for every t. These "hyperstationary" processes may be interesting in their own right. I mention them because the program of asking what types of processes follow from imposing conditions on S-modules may lead to interesting new classes of processes.

5. Ergodic processes. The type of ergodic theorem proved for processes such as chains of infinite order (see for example Suppes and Lamperti (1959)) suggest the following definition of ergodicity. $\{X_n\}$ is an ergodic process if for all $x_1, \cdots, x_n, y_1, \cdots, y_m$

$$\lim_{t\to\infty} P\{X_{n+t+1} = y_1, \cdots, X_{n+t+m} = y_m \mid X_1 = x_1, \cdots, X_n = x_n\} = p(y_1, \cdots, y_m)$$

uniformly in $x_1, \cdots, x_n$ and n. In terms of (L, q, e) this means that for every $\pi \varepsilon U_s$

$$\lim_{n\to\infty} q(\pi'\sigma^n\pi e)/q(\pi' e) = p(\pi)$$

uniformly in π' for which $q(\pi' e) > 0$. As yet S-modules have no topological structure with which to describe a convergence property like ergodicity. However, a norm may be quite naturally defined on a *reduced* S-module which is relevant to the present discussion. It is given by

$$\|l\| = \sup_{\pi \varepsilon U_s, q(\pi e)>0} |q(\pi l)/q(\pi e)|. \tag{12}$$

There are a couple of details that require the reduced nature of the S-module for their veracity. First $\|l\| < \infty$ for every $l \varepsilon L$ since $l = \xi e = \sum_\alpha c_\alpha \pi_\alpha e$, $c_\alpha \varepsilon \mathbf{R}$, $\pi_\alpha \varepsilon U_s$ and $\|l\| \leqq \sum_\alpha |c_\alpha| < \infty$. Secondly if $\|l\| = 0$ then $q(\pi l) = 0$ for all $\pi \varepsilon U_s$ for which $q(\pi e) > 0$. But if $q(\pi e) = 0$ then $q(\pi l) = 0$ so we have $q(A_s l) = 0$ or $l = 0$.

Henceforth all convergence in a reduced S-module will refer to the norm convergence in $\|\cdot\|$ which we call the ergodic- or E-norm. Since an S-module is of at most countable *algebraic* dimension, L will be complete in the E-norm if and only if it is finite dimensional, in which case *all* norms are equivalent. We summarize some obvious facts about the E-norm as follows.

THEOREM 6.

$$\begin{aligned} &(a) \quad \|e\| = 1, \\ &(b) \quad \|\pi l\| \leqq \|l\| \quad \textit{for all} \quad \pi \, \varepsilon \, U_s, \\ &(c) \quad \|\sigma l\| \leqq \|l\|. \end{aligned}$$

By restating the definition of ergodicity in terms of the E-norm we obtain

THEOREM 7. *If (L, q, e) is the reduced S-module of $\{X_n\}$ then the process is ergodic if and only if for every $l \, \varepsilon \, L$ there is $c_l \, \varepsilon \, \mathbf{R}$ such that* $\lim_{n\to\infty} \sigma^n l = c_l e$.

If we let $K_\sigma = \{l: \sigma^n l \to 0\}$ then K_σ is a linear σ-invariant subspace of L such that $K_\sigma \cap \mathbf{R}e = \{0\}$. $\mathbf{R}e$ is also a σ-invariant subspace of L. It is a straightforward matter to prove the following fact.

THEOREM 8. *If (L, q, e) is the reduced S-module of $\{X_n\}$, then the process is ergodic if and only if $L = \mathbf{R}e \oplus K_\sigma$.*

As a corollary for processes of finite rank we obtain:

COROLLARY. *If* $\dim (L) = n$ *then the process is ergodic if and only if* $\dim (K_\sigma) = n - 1$.

The only way a "hyperstationary" process of Section 4 can be ergodic in the present sense is if it is a sequence of independent and identically distributed random variables.

6. Acknowledgments. I would like to thank the referee for his useful suggestions and comments.

REFERENCES

[1] HELLER, A. (1965). On stochastic processes derived from Markov chains. *Ann. Math. Statist.* **36** 1286–1291.

[2] LAMPERTI, J. and SUPPES, P. (1959). Chains of infinite order and their application to learning theory. *Pacific J. Math.* **9** 739–754.