STATIONARY MEASURES FOR CERTAIN STOCHASTIC PROCESSES

JOHN LAMPERTI

Introduction. In a recent paper [1], T.E. Harris has studied stationary processes $\{Z_n\}$ with a finite number of states, taken to be the integers $0, 1, \cdots, D\text{-}1$. His technique is to map the half-infinite sample sequences $Z_n, Z_{n-1}, \cdots$ onto the unit interval by means of the correspondence

$$(1) \qquad X_{n+1} = Z_n/D + Z_{n-1}/D^2 + \cdots .$$

The X_n then form a stationary Markov process. In § 5 of [1] Harris shows (Theorem 7) that if the process $\{Z_n\}$ is of mixing type, then either the stationary distribution $G(x) = \Pr(X_n \leq x)$ has a unit step, or is the uniform distribution, or $G(x)$ is continuous and totally singular.

The purpose of this paper is to investigate correspondences such as (1) in general, using two simple lemmas in ergodic theory which are given in the next section. If $g(\{i_0, i_1, \cdots\})$ is any essentially one-to-one and measurable mapping of the space of sequences $\{i_0, i_1, \cdots\}$ onto another measurable space X, then a correspondence similar to (1) may be defined between stochastic processes with states i and processes on X:

$$(2) \qquad X_{n+1} = g(\{Z_n, Z_{n-1}, \cdots\}) .$$

Theorem 1 describes the resulting distributions on X; Theorem 2 is a specialization to the case of (1). Finally an additional application (Theorem 3) is made to certain of the processes studied by Karlin in [3]. Theorem 2 contains Theorem 7 of [1], and Theorem 3 overlaps with § 7 of [3]. In addition to a unified approach, some extension of the previous results is obtained in both cases.

2. Ergodic theory lemmas.

LEMMA 1. *Let (Ω, W) be a measurable space and T a measurable transformation of Ω onto itself. Let μ_1 and μ_2 be two sigma-finite measures on (Ω, W) such that for each, T is a measure preserving, metrically-transitive transformation. Then if μ_1 and μ_2 are not proportional, they are orthogonal (i.e., have their positive mass on disjoint sets).*

Proof. Suppose μ_1 and μ_2 are both finite measures, and assume they have been normalized. Let A be a set such that $\mu_1(A) \neq \mu_2(A)$. Define

$$B_i = \left\{ \omega \in \Omega \,\middle|\, \lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} \phi_A(T^j \omega) = \mu_i(A) \right\}, \qquad i = 1, 2,$$

where $\phi_A(\cdot)$ is the characteristic function of the set A. By the individual ergodic theorem, $\mu_i(B_i)=1$. Hence B_1 and $B_1^c \supset B_2$ are a decomposition of Ω demonstrating the orthogonality of μ_1 and μ_2. In the case where one or both of the measures are infinite, the same idea may be carried out using Hopf's ratio ergodic theorem (see, for instance, [2]).

Now let Ω be the $\overset{\infty}{\underset{i=-\infty}{\Pi}} Y_i$ of sequences $\{\omega_i\}$ where $\omega_i \in Y_i$ with each $Y_i = Y$, Y a measurable space. Let W be the Borel field generated by the " cylinder sets " of Ω^2. Denote by S the " shift " transformation

$$(3) \qquad\qquad S\{\omega_i\} = \{\nu_i\} \quad \text{where} \quad \nu_i = \omega_{i+1} .$$

LEMMA 2. *Let μ be a measure on (Ω, W) such that S is measure preserving and metrically-transitive and $\mu(\Omega)=1$. Then one of the following is the case*:

(a) *there is a finite sequence $a_1, a_2, \cdots, a_m$ of points of Y such that μ has mass $\dfrac{1}{m}$ on each of the m points of Ω given by*

$$\omega_i = a_j \quad \text{for} \quad i \equiv j+k \mod m ; \quad k = 0, 1, \cdots, m-1.$$

(b) $\mu\{\omega \in \Omega | \omega_0 = a_0, \omega_{-1} = a_1, \cdots\} = 0$ *for any sequence $\{a_n\}$ of points of Y.*

Proof. Suppose that case (a) does not hold. Then we shall show that $\mu(A)=0$, where

$$A = \overset{\infty}{\underset{n=-\infty}{\cup}} A_n, A_n = \{\omega \in \Omega | \omega_n = a_0 \omega_{n-1} = a_1, \cdots\} .$$

Now A is invariant under S, and so $\mu(A)=0$ or 1. Assume $\mu(A)=1$. It is not hard to see that if $\{a_i\}$ were a periodic sequence, case (a) would hold. But if $\{a_i\}$ is not periodic, a value of n such that $\omega_n = a_0$, $\omega_{n-1} = a_1, \cdots$ must be unique, and so the A_n are disjoint. Since S is measure preserving, $\mu(A_n)$ are all equal. This contradicts the assumption that $\mu(A)=1$.

Finally we remark that, speaking somewhat less precisely, Lemma 2, may be re-expressed as : A stationary ergodic stochastic process either executes deterministically a certain periodic motion, or else each path function has probability 0. We shall refer to these alternatives as case a and b.

3. **Induced Markov processes.** In this section we continue to use the notation $\Omega = \overset{\infty}{\underset{i=-\infty}{\Pi}} Y_i$, but Y is restricted to be a fixed (not necessarily

---

[2] In the applications we shall make of this lemma, Y will be restricted (in fact, will be denumerable), so that the Kolmogorov extension theorem will hold.

finite) set of integers.   Let Ω_0 denote the space of half-infinite sequences $\{i_0, i_1, \cdots\}$ where $i_n \in Y$; the measurable sets of both Ω and Ω_0 are again those belonging to the Borel fields generated by the cylinder sets. If $\omega^{(0)} = \{i_0, i_1, \cdots\}$ is an element of Ω_0, the notation $\{i, \omega^{(0)}\}$ will mean the sequence $\{i, i_0, i_1, \cdots\}$.

Let (X, F) be a measurable space, and suppose that there exists a mapping $g(\omega^{(0)})$ of Ω_0 onto X which is one to one if sets at most denumerable are deleted from Ω_0 and from X; suppose also that both g and g^{-1} are measurable.   Let μ be any probability measure on the space Ω, and let $\{Z_n\}$ mean the stochastic process consisting of the random variables

$$(4) \qquad\qquad Z_n(\omega) = i_n .$$

A new process $\{X_n\}$, with state-space X, may be defined by (2) and (4).

LEMMA 3.   *Assume that for each particular sequence* $\omega^{(0)}$ *in* Ω_0, *and for each* n,

$$(5) \qquad\qquad \Pr(Z_n = i_0, Z_{n-1} = i_1, \cdots) = 0 .$$

Then $\{X_n\}$ *is a Markov process whose (not necessarily stationary) transition probabilities are given with probability one by*

$$(6) \qquad \begin{aligned} &X_{n+1} = g[\{i, g^{-1}(X_n)\}] \text{ with probability} \\ &f_i(X_n) = \Pr[Z_n = i \,|\, \{Z_{n-1}, Z_{n-2}, \cdots\} = g^{-1}(X_n)] . \end{aligned}$$

Proof.   Let $E \subset \Omega$ denote the set of all " path functions " for the $\{Z_n\}$ process such that some segment $\{Z_n, Z_{n-1}, \cdots\}$ belongs to the (denumerable) set which must be deleted from Ω_0 in order to secure a one-to-one map onto X; it follows from (5) that $\mu(E) = 0$.   Therefore with probability one, knowledge of X_n determines the sequence $\{Z_{n-1}, Z_{n-2}, \cdots\} = g^{-1}(X_n)$ uniquely.   Then X_n also determines $X_{n-1}, X_{n-2}, \cdots$ and so the process is Markovian.   That (6) gives the transition law is clear. There are, of course, many cases where the Markov property and (6) hold even though (5) does not.

Consider now measures μ such that the shift operation (3) s measure preserving and metrically-transitive ; in other words, measures such that $\{Z_n\}$ is a stationary, ergodic stochastic process.   In this case, $\{X_n\}$ will also be stationary ; let Q_μ denote the stationary probability measure of (each) X_n.

THEOREM 1.   *If* $\{Z_n\}$ *executes deterministically a cycle of period* m *(case a), then* Q_μ *concentrates its mass upon at most* m *points of* X. *Otherwise (case b)* $\{X_n\}$ *is a Markov process, the measure* Q_μ *is non-atomic,*

and any two measures of this type resulting from different μ's are orthogonal.

Proof. In case a, the measure μ concentrates on m points, and so Q_μ will concentrate on the images of these points, which may or may not all be distinct. (If they are distinct, then $\{X_n\}$ must be a Markov process.) Otherwise, (case b), it follows from Lemma 2 that (5) holds, and hence that Q_μ is non-atomic. Lemma 3 then implies that $\{X_n\}$ is a Markov process. Under the mapping g the relation of orthogonality of non-atomic measures is preserved, and so Lemma 1 yields the remaining assertion of the theorem.

A remark about infinite measures will conclude this section. Suppose the shift operation is measure-preserving and metrically-transitive for sigma-finite measures μ on Ω which have the property (b) of Lemma 2. Let Q_μ denote the perhaps infinite measures which are then induced by g on (X, F). Lemma 1 is still available, and so the conclusion of orthogonality of distinct Q_μ's remains valid.

4. Applications.

EXAMPLE 1. We now consider the particular case studied in [1]. The set Y will consist of the intergers $0, 1, \cdots, D\text{-}1$, and (X, F) will be the unit interval and the field of Borel sets. Let

$$(7) \qquad\qquad g(\{i_0, i_1, \cdots\}) = i_0/D + i_1/D^2 + \cdots .$$

Then the correspondence between a process $\{Z_n\}$ with states Y and a process $\{X_n\}$ is given by (1). In this situation we have

THEOREM 2. *Let $\{Z_n\}$ be a metrically-transitive, stationary process with state-space Y; let $G(x) = \Pr(X_n \leq x)$ be the stationary distribution function of X_n. Then $\{X_n\}$ is always a Markov process, and one of the following holds:*

(i) *$\{Z_n\}$ executes deterministically a cycle of period m; in this case, $G(x)$ has m discontinuities each of leap $1/m$.*

(ii) *Each Z_n is independently uniformly distributed $\{0, 1, \cdots, D\text{-}1\}$. In this case $G(x) = x$, $0 \leq x \leq 1$.*

(iii) *$G(x)$ is continuous and singular with respect to Lebesgue measure. Finally, any two continuous distributions $G(x)$ are orthogonal.*

Proof. The fact that $\{X_n\}$ is always a Markov process, and the statement (i), follow since the mapping (7) cannot map two sequences of Ω_0 having positive measure into the same point. Statement (ii) is easily verified, and then the remainder of the theorem follows from Theorem 1.

EXAMPLE 2. In this application, the $\{X_n\}$ process is the primary object of interest; it is a type of learning model [3]. Let Y consist of the integers 0 and 1, and again take (X, F) as the unit interval and Borel field. Let σ and α be two numbers between 0 and 1 such that $\sigma + \alpha \leqq 1$. (The present approach does not seem to apply when $\sigma + \alpha > 1$.) Define inductively a family of subintervals of the unit interval as follows:

$$A(\{0\}) = [0, \sigma], \quad A(\{1\}) = [1 - \alpha, 1],$$

and if $\omega_m^0 = \{i_0, i_1, \cdots, i_{m-1}\}$, then

$$A(\{0, \omega_m^0\}) = \sigma A(\omega_m^0) \quad \text{and} \quad A(\{1, \omega_m^0\}) = 1 - a + a A(\omega_m^0).$$

Now since both $\sigma < 1$ and $\alpha < 1$, for any sequence $\omega^{(i)}$ the $A(\omega_m^0)$ are a sequence of nested intervals of length approaching zero. Therefore the following definition is meaningful:

$$(8) \qquad g(\{i_0, i_1, \cdots\}) = \bigcap_{m=1}^{\infty} A(\omega_m^0).$$

Let $\{Z_n\}$ be a stationary stochastic process with states Y, and define $\{X_n\}$ by (2) and (8). Let $G(x)$ again denote the stationary distribution of the X_n.

THEOREM 3. $\{X_n\}$ is a Markov process with transition law

$$(9) \qquad X_{n+1} = \begin{cases} \sigma X_n \\ 1 - \alpha + \alpha X_n \end{cases} \text{with probability} \begin{array}{l} f_0(X_n) \\ f_1(X_n) = 1 - f_0(X_n). \end{array}$$

Any stationary Markov process of this form is induced by some process $\{Z_n\}$: If $\{Z_n\}$ is in addition metrically transitive, one of the following cases must hold:

(i) *$\{Z_n\}$ executes deterministically a cycle of period m; $G(x)$ has m discontinuities each of leap $1/m$.*

(ii) *$G(x) = x$; this occurs if and only if $\sigma + \alpha = 1$ and $f_0(x) \equiv \sigma$, $f_1(x) \equiv 1 - \sigma$.*

(iii) *$G(x)$ is continuous and singular with respect to Lebesque measure.*

Any two continuous distributions $G(x)$ arising from different metrically-transitive processes $\{Z_n\}$ but the same mapping g (that is, the same σ and α) are orthogonal.

Proof. If $\sigma + a < 1$, then the mapping g is not onto the whole unit interval, but onto a cantor-like subset of measure zero; it is precisely one-to-one onto this set. Therefore X_n must be a Markov process, and the transition law (9) is obtained from (6) and (8). The continuity of $G(x)$ if (i) does not hold follows from Lemma 2. Since $G(x)$ is a distribution on a set of measure zero, it must automatically be a singular

distribution for any process $\{Z_n\}$.

Now suppose $\sigma + \alpha = 1$. In this case there are some points x corresponding under g^{-1} to two points of Ω_0; however, just as in Theorem 2 the ambiguities of the mapping do not affect either the Markov property or the m distinct discontinuities of statement (i). Statement (ii) is readily verified, and then (iii) follows from Theorem 1. Whether $\sigma + \alpha = 1$, or < 1, the last statement of the theorem also follows from Theorem 1.

If $\{X_n\}$ is a stationary Markov process of the form (9), the distribution $G(x) = \Pr(X_n \leq x)$ concentrates positive mass only on the domain of g^{-1}; hence a stationary measure is induced by g^{-1} on Ω_0, which extends to a measure μ on Ω. The process $\{Z_n\}$ inducing $\{X_n\}$ is then defined by (4). This completes the proof.

REFERENCES

1. T. E. Harris, *On chains of infinite order, Pacific J. of Math*, **5** (1955), 707–724.
2. E. Hopf, *Ergodentheorie*, Chelsea, 1948.
3. S. Karlin, *Some random walks arising in learning models I, Pacific J. Math.* **3** (1953), 725–756.

CALIFORNIA INSTITUTE OF TECHNOLOGY