MARKOV PROCESS REPRESENTATIONS OF GENERAL STOCHASTIC PROCESSES

DUDLEY PAUL JOHNSON

ABSTRACT. In this paper we show that any separable stochastic process on a compact metric space can be derived from a temporally homogeneous Markov process on the extreme points of a compact convex set of measures.

Let $\mathfrak{X}$ be a compact metric space with Borel field Σ. Let T be either the nonnegative integers or the nonnegative rationals and let Ω be the set of all functions mapping T into $\mathfrak{X}$. Ω with its product topology is a compact metric space and so $C(\Omega)$, the Banach space of continuous functions on Ω, is separable [**1**, p. 340], and the weak * topology of the closed unit sphere of the Banach space $\mathrm{rca}(\Omega)$ of regular countably additive set functions on Ω is a metric topology [**1**, p. 426]. If for each $\omega \in \Omega$ and $t \in T$ we define $x_t(\omega) = \omega(t)$ and if we let $\mathfrak{A}$ be the σ-field of Borel subsets of Ω, then for each $\mu \in \mathcal{P}(\Omega)$, the set of all probability measures in $\mathrm{rca}(\Omega)$, we get a stochastic process $X_\mu = (\Omega, \mathfrak{A}, x_t, \mathfrak{X}, \mu)$.

If $\omega \in \Omega$, $\Lambda \in \mathfrak{A}$ and $\mu \in \mathcal{P}(\Omega)$, let $\omega_s^+ \in \Omega$ be defined by $\omega_s^+(t) = \omega(s+t)$, Λ_s^+ be the set of all $\omega \in \Omega$ for which $\omega_s^+ \in \Lambda$ and let $\lambda_s^+ \in \mathcal{P}(\Omega)$ be defined by $\lambda_s^+(\Lambda) = \lambda(\Lambda_s^+)$. Let $\mathfrak{D}_0^\mu$ be the set of all $\lambda \in \mathcal{P}(\Omega)$ which have the property that for some $0 < s_1 < \cdots < s_n$ in T and $A_1, \cdots, A_n$ in Σ, $\mu(x_{s_1} \in A_1, \cdots, x_{s_n} \in A_n) > 0$ and

$$\lambda(\Lambda) = \mu(x_{s_1} \in A_1, \cdots, x_{s_n} \in A_n, \overset{+}{\Lambda}_{s_n}) / \mu(x_{s_1} \in A_1, \cdots, x_{s_n} \in A_n)$$

for each $\Lambda \in \mathfrak{A}$. Let $\mathfrak{S}^\mu$ be the set of all weak * compact simplexes $\mathfrak{D}$ in $\mathcal{P}(\Omega)$ which contain $\mathfrak{D}_0^\mu$ and have the property that $\mu \in \mathfrak{D}$ implies that

(i) $\mu_t^+ \in \mathfrak{D}$ for each $t \in T$;

(ii) $\mu(\cdot \mid x_0 \in A) \in \mathfrak{D}$ for each $A \in \Sigma$. Ordering $\mathfrak{S}^\mu$ by inclusion and applying Zorn's Lemma, we find that $\mathfrak{S}^\mu$ contains minimal elements. Let $\mathfrak{D}^\mu$ be one of these minimal subsets of $\mathcal{P}(\Omega)$. Let $\mathfrak{Y}^\mu$ be the set of extreme points of $\mathfrak{D}^\mu$, Ω^μ the set of all functions mapping T into $\mathfrak{Y}^\mu$, and $\{x_t^\mu, t \in T\}$ the family of functions mapping Ω^μ into $\mathfrak{Y}^\mu$ defined by $x_t^\mu(\omega^\mu) = \omega^\mu(t)$. Finally, let $\mathfrak{A}^\mu$ be the σ-field generated by x_t^μ, $t \in T$.

Received by the editors July 15, 1969.

AMS Subject Classifications. Primary 6040, 6060.

Key Words and Phrases. Stochastic process, temporally homogeneous Markov process, extreme points, Choquet's Theorem.

If $\mu \in \mathcal{P}(\Omega)$ and $\lambda \in \mathcal{Y}^\mu$, then $\lambda_t^+ \in \mathfrak{D}^\mu$. Thus by Choquet's Theorem there exist unique measures $P^\mu(\cdot)$ and $P_t^\mu(\lambda, \cdot)$ on the weak * Borel subsets of $\mathcal{Y}^\mu$ such that for any weak * continuous linear functional f on $\mathcal{P}(\Omega)$,

$$f(\mu) = \int_{\mathcal{Y}^\mu} f(\nu) P^\mu(d\nu) \quad \text{and} \quad f(\lambda_t^+) = \int_{\mathcal{Y}^\mu} f(\nu) P_t^\mu(\lambda, d\nu).$$

Let $\mu^* \in \mathcal{P}(\Omega^\mu, \mathcal{A}^\mu)$ be defined by

$$\mu^*(x_{t_1}^\mu \in B_1, \cdots, x_{t_1+\cdots+t_n}^\mu \in B_n)$$
$$= \int_{\mathcal{Y}^\mu} P(d\nu_0) \int_{B_1} P_{t_1}(\nu_0, d\nu_1) \cdots \int_{B_n} P_{t_n}(\nu_{n-1}, d\nu_n).$$

μ^* is consistently defined since for any continuous linear functional f on $\mathcal{P}(\Omega)$,

$$\begin{aligned}\int_{\mathcal{Y}^\mu} f(\nu) \left(\int_{\mathcal{Y}^\mu} P_s^\mu(\lambda, d\xi) P_t^\mu(\xi, d\nu) \right) &= \int_{\mathcal{Y}^\mu} P_s^\mu(\lambda, d\xi) \int_{\mathcal{Y}^\mu} f(\nu) P_t^\mu(\xi, d\nu) \\ &= \int_{\mathcal{Y}^\mu} f(\xi_t^+) P_s^\mu(\lambda, d\xi) \\ &= f(\lambda_{s+t}^+) \\ &= \int_{\mathcal{Y}^\mu} f(\nu) P_{s+t}^\mu(\lambda, d\nu)\end{aligned}$$

and so by the uniqueness of $P_t^\mu(\lambda, \cdot)$

$$P_{s+t}^\mu(\lambda, \cdot) = \int_{\mathcal{Y}^\mu} P_s^\mu(\lambda, d\nu) P_t^\mu(\nu, \cdot).$$

Thus not only is μ^* consistently defined, but $X_\mu^* = (\Omega^\mu, \mathcal{A}^\mu, x_t^\mu, \mathcal{Y}^\mu, \mu^*)$ is a temporally homogeneous Markov process with initial distribution P^μ and transition probability function P_t^μ.

If $\mu \in \mathcal{P}(\Omega)$ and $\nu \in \mathcal{Y}^\mu$, then for each set $A \in \Sigma$, either $\nu(x_0 \in A)$ or $\nu(x_0 \in A^c)$ is zero. Indeed, suppose that $\nu(x_0 \in A) > 0$ and $\nu(x_1 \in A^c) > 0$. Then

$$\nu(\cdot) = \nu(\cdot \mid x_0 \in A)\nu(x_0 \in A) + \nu(\cdot \mid x_0 \in A^c)\nu(x_0 \in A^c).$$

Since $\nu \in \mathcal{Y}^\mu$ and since $\nu(\cdot \mid x_0 \in A)$ and $\nu(\cdot \mid x_0 \in A^c)$ are both in $\mathfrak{D}^\mu$ we must have

$$\nu(\cdot) = \nu(\cdot \mid x_0 \in A) = \nu(\cdot \mid x_0 \in A^c).$$

Thus $\nu(x_0 \in A) = \nu(x_0 \in A \mid x_0 \in A^c) = 0$ which is a contradiction.

Let $\mathcal{C}(\nu)$ be the class of all sets $A \in \Sigma$ for which $\nu(x_0 \in A) > 0$. Ordering $\mathcal{C}$ by inclusion and applying Zorn's Lemma, we see that $\mathcal{C}$ has a unique minimal element which is a set consisting of a single point δ_ν. For each $t \in T$, we now let $\hat{x}_t = \delta_{x_t^\mu}$ and $\hat{X}_\mu$ be the stochastic process

$$\hat{X}_\mu = (\Omega^\mu, \mathcal{A}^\mu, \hat{x}_t, \mathfrak{X}, \mu^*).$$

We then have the

THEOREM. *If $\mu \in \mathcal{P}(\Omega)$, then $X_\mu = \hat{X}_\mu$ in distribution.*

PROOF. Since for any continuous function g on Ω

$$\int g(\omega) \overset{+}{\lambda}_t(d\omega) = \int_{\mathcal{Y}} \left(\int_\Omega g(\omega)\nu(d\omega) \right) P_t(\lambda, d\nu)$$
$$= \int_\Omega g(\omega) \int_{\mathcal{Y}} \nu(d\omega) P_t(\lambda, d\nu),$$

we have for each $\Lambda \in \mathcal{A}^\mu$

$$\overset{+}{\lambda}_t(\Lambda) = \int_{\mathcal{Y}} \nu(\Lambda) P_t(\lambda, d\nu).$$

Letting $\sigma^\mu A = \{\nu : \delta_\nu \in A\}$ and dropping the superscript μ from $\mathcal{Y}^\mu$, P_t^μ and σ^μ, we have for any $A \in \Sigma$ and $\Lambda \in \mathcal{A}$,

$$\overset{+}{\lambda}_t(x_0 \in A, \Lambda) = \int_{\sigma A} \nu(x_0 \in A, \Lambda) P_t(\lambda, d\nu) = \int_{\sigma A} \nu(\Lambda) P_t(\lambda, d\nu).$$

Thus

$$\lambda(x_t \in A, \overset{+}{\Lambda_t}) = \int_{\sigma A} \nu(\Lambda) P_t(\lambda, d\nu).$$

When $\Lambda = \Omega$,

$$\lambda(x_t \in A) = \int_{\sigma A} P_t(\lambda, d\nu)$$

and so

$$\mu(x_t \in A) = \int_{\mathcal{Y}} \lambda(x_t \in A) P(d\lambda) = \int P(d\lambda) P_t(\lambda, \sigma A).$$

Using induction on n we see that if $\lambda \in \mathcal{Y}$, then

$$\lambda(x_{t_1} \in A_1, \cdots, x_{t_1+\cdots+t_n} \in A_n) = \int_{\sigma A_1} P_{t_1}(\lambda, d\nu_1) \int_{\sigma A_2} P_{t_2}(\nu_1, d\nu_2) \cdots \int_{\sigma A_n} P_{t_n}(\nu_{n-1}, d\nu_n).$$

Indeed if $n=1$ we have already proven it and if it is true for $n=r-1$, then

$$\begin{aligned}
\lambda(x_{t_1} \in A_1, \cdots, x_{t_1+\cdots+t_r} \in A_r) &= \lambda(x_{t_1} \in A_1, (x_{t_2} \in A_2, \cdots, x_{t_2+\cdots+t_r} \in A_r)_{t_1}) \\
&= \int_{\sigma A_1} \nu_1(x_{t_2} \in A_2, \cdots, x_{t_2+\cdots+t_r} \in A_r) P_{t_1}(\lambda, d\nu_1) \\
&= \int_{\sigma A_1} P_{t_1}(\lambda, d\nu_1) \int_{\sigma A_2} P_{t_2}(\nu_1, d\nu_2) \cdots \int_{\sigma A_r} P_{t_r}(\nu_{r-1}, d\nu_r).
\end{aligned}$$

Thus

$$\begin{aligned}
\mu(x_0 \in A_0, x_{t_1} \in A_1, \cdots, x_{t_1+\cdots+t_n} \in A_n) &= \int_{\mathcal{Y}} \nu(x_0 \in A_0, x_{t_1} \in A_1, \cdots, x_{t_1+\cdots+t_n} \in A_n) P(d\nu) \\
&= \int_{\sigma A_0} \nu(x_{t_1} \in A_1, \cdots, x_{t_1+\cdots+t_n} \in A_n) P(d\nu) \\
&= \int_{\sigma A_0} P(d\nu_0) \int_{\sigma A_1} P_{t_1}(\nu_0, d\nu_1) \cdots \int_{\sigma A_n} P_{t_n}(\nu_{n-1}, d\nu_n). \\
&= \mu^*(\hat{x}_0 \in A_0, \hat{x}_{t_1} \in A_1, \cdots, \hat{x}_{t_1+\cdots+t_n} \in A_n)
\end{aligned}$$

and the proof is complete.

References

1. N. Dunford and J. Schwartz, *Linear operators.* I: *General theory,* Pure and Appl. Math., vol. 7, Interscience, New York, 1958. MR 22 #8302.

2. P. A. Meyer, *Probability and potentials,* Blaisdell, Waltham, Mass., 1966. MR 34 #5119.

University of California, Riverside