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

If $\omega \in \Omega$, $\Lambda \in \mathcal{A}$ and $\mu \in \mathcal{P}(\Omega)$, let $\omega^+_\Lambda \in \Omega$ be defined by $\omega^+_\Lambda(t) = \omega(s+t)$,
$\Lambda^+_s$ be the set of all $\omega \in \Omega$ for which $\omega^+_\Lambda \in \Lambda$ and let $\lambda^+_\Lambda \in \mathcal{P}(\Omega)$ be defined by
$\lambda^+_\Lambda(\Lambda) = \lambda(\Lambda^+_s)$. Let $D^\Lambda_0$ 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
$\Sigma$, $\mu(x_{s_1} \in A_1, \cdots, x_{s_n} \in A_n) > 0$ and

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

for each $\Lambda \in \mathcal{A}$. Let $\mathcal{E}^\mu$ be the set of all weak $^*$ compact simplexes
$\mathcal{D}$ in $\mathcal{P}(\Omega)$ which contain $D^\Lambda_0$ and have the property that $\mu \in D$ implies that

(i) $\mu^+_\Lambda \in D$ for each $t \in T$;

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

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

\[
f(\mu) = \int_{\mathcal{Y}^{\mu}} f(\nu) P^{\mu}(d\nu) \quad \text{and} \quad f(\lambda^{+}_{i}) = \int_{\mathcal{Y}^{\mu}} f(\nu) P^{\mu}_{i}(\lambda, d\nu).
\]

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

\[
\mu^{*}(x_{1}^{\mu} \in B_{1}, \ldots, x_{n}^{\mu} \in B_{n}) = \int_{\mathcal{Y}^{\mu}} P(d\nu) \int_{B_{1}} P_{1}(\nu_{0}, d\nu_{1}) \cdots \int_{B_{n}} P_{n}(\nu_{n-1}, d\nu_{n}).
\]

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

\[
\int_{\mathcal{Y}^{\mu}} f(\nu) \left( \int_{\mathcal{Y}^{\mu}} P^{\mu}_{i}(\lambda, d\xi) P^{\mu}_{i}(\xi, d\nu) \right) = \int_{\mathcal{Y}^{\mu}} P^{\mu}_{i}(\lambda, d\xi) \int_{\mathcal{Y}^{\mu}} f(\nu) P^{\mu}_{i}(\xi, d\nu) = \int_{\mathcal{Y}^{\mu}} f(\xi^{+}) P^{\mu}_{i}(\lambda, d\xi) = f(\lambda^{+}_{i}) = \int_{\mathcal{Y}^{\mu}} f(\nu) P^{\mu}_{i+1}(\lambda, d\nu)
\]

and so by the uniqueness of \( P^{\mu}_{i}(\lambda, \cdot) \)

\[
P^{\mu}_{i+1}(\lambda, \cdot) = \int_{\mathcal{Y}^{\mu}} P^{\mu}_{i}(\lambda, d\nu) P^{\mu}_{i}(\nu, \cdot).
\]

Thus not only is \( \mu^{*} \) consistently defined, but \( \mathcal{X}^{\mu*} = (\Omega^{\mu}, \mathcal{A}^{\mu}, x_{1}^{\mu}, \mathcal{Y}^{\mu}, \mu^{*}) \) is a temporally homogeneous Markov process with initial distribution \( P^{\mu} \) and transition probability function \( P^{\mu}_{i} \).

If \( \mu \in \mathcal{G}(\Omega) \) 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 \( \mathcal{D}^{\mu} \) 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 $\delta_*$. For each $t \in T$, we now let $\hat{x}_t = \delta_*^t$ and $\hat{X}_t$ be the stochastic process

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

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 $\Omega$

$$\int g(\omega) \lambda_t^+(d\omega) = \int_y \left( \int_\Lambda g(\omega) \nu(d\omega) \right) P_t(\lambda, dv)$$

$$= \int_\Lambda g(\omega) \int_y \nu(d\omega) P_t(\lambda, dv),$$

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

$$\lambda_t^+(\Lambda) = \int_y \nu(\Lambda) P_t(\lambda, dv).$$

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

$$\lambda_t^+(x_0 \in A, \Lambda) = \int_{\sigma A} \nu(x_0 \in A, \Lambda) P_t(\lambda, dv) = \int_{\sigma A} \nu(\Lambda) P_t(\lambda, dv).$$

Thus

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

When $\Lambda = \Omega$,

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

and so

$$\mu(x_t \in A) = \int_y \lambda(x_t \in A) P(d\lambda) = \int P(d\lambda) P_t(\lambda, \sigma A).$$

Using induction on $n$ we see that if $\lambda \in \mathcal{Y}$, then
\( \lambda(x_1 \in A_1, \ldots, x_t + \ldots + t_n \in A_n) \)
\[= \int_{\sigma A_1} P_t(\lambda, d\nu_1) \int_{\sigma A_2} P_t(\nu_1, d\nu_2) \cdots \int_{\sigma A_n} P_t(\nu_{n-1}, d\nu_n). \]

Indeed if \( n = 1 \) we have already proven it and if it is true for \( n = r - 1 \), then
\[\lambda(x_1 \in A_1, \ldots, x_t + \ldots + t_r \in A_r) \]
\[= \lambda(x_1 \in A_1, (x_t \in A_2, \ldots, x_t + \ldots + t_r \in A_r) \tau_t) \]
\[= \int_{\sigma A_1} \nu_1(x_t \in A_2, \ldots, x_t + \ldots + t_r \in A_r) P_t(\lambda, d\nu_1) \]
\[= \int_{\sigma A_1} P_t(\lambda, d\nu_1) \int_{\sigma A_2} P_t(\nu_1, d\nu_2) \cdots \int_{\sigma A_r} P_t(\nu_{r-1}, d\nu_r). \]

Thus
\[\mu(x_0 \in A_0, x_t \in A_1, \ldots, x_t + \ldots + t_n \in A_n) \]
\[= \int_{\sigma A_0} \nu_0(x_0 \in A_0, x_t \in A_1, \ldots, x_t + \ldots + t_n \in A_n) P(d\nu) \]
\[= \int_{\sigma A_0} \nu(x_t \in A_1, \ldots, x_t + \ldots + t_n \in A_n) P(d\nu) \]
\[= \int_{\sigma A_0} P(d\nu_0) \int_{\sigma A_1} P_t(\nu_0, d\nu_1) \cdots \int_{\sigma A_n} P_t(\nu_{n-1}, d\nu_n). \]
\[= \mu^*(x_0 \in A_0, x_t \in A_1, \ldots, x_t + \ldots + t_n \in A_n) \]

and the proof is complete.

**References**


University of California, Riverside