A NEW PROOF OF THE STRONG MARKOV THEOREM OF CHUNG

BY D. G. AUSTIN*

YALE UNIVERSITY, UNIVERSITY OF MIAMI

Communicated by Einar Hille, April 22, 1958

We consider time-homogeneous Markov processes with continuous time parameter and discrete phase space. The recent important result of K. L. Chung1 that every such process has the strong Markov property will be obtained by approximating the optional time by simple random variables (see below for a precise statement).

It should be pointed out that Chung's result includes the critical point \( t = 0 \) in the theorem below with respect to the set where \( y(0) \neq \infty \). In addition to the strong Markov theorem Chung also gives the explicit form of the conditional distribution of \( y(t) \) relative to \( \alpha \) and proves it continuous in \( t \)—a result useful in applications of the theorem (see footnote 6). I should like to thank Professor Chung and Professor Kakutani for many helpful conversations during the development of this proof.

Let \( (P, \mathcal{F}, \Omega) \) be a probability triple, where \( \Omega \) is any abstract space, \( \mathcal{F} \) is a Borel field, of subsets of \( \Omega \), and \( P \) is a probability measure on \( \mathcal{F} \). We consider stochastic processes of the form \( x(t, w) \), where \( w \in \Omega \), \( t \) is on the non-negative real line \( R \), and where it is always assumed that the range is contained in the space of positive integers.2 We denote the time-homogeneous Markov processes3 by \( \mathcal{M} \); let \( \{ p_{ij}(t) \} \) be a transition matrix for \( x(t, w) \in \mathcal{M} \); such a matrix will be called standard if

\[
\lim_{t \to 0} p_{ij}(t) = \delta_{ij}.
\]

When a sequence of random variables \( y_n(w) \) converges in probability to \( y(w) \) we write \( \lim_{n} y_n(w) = y(w) \).

Two conditional probabilities will be described as equal if they are equal almost everywhere \([P]\). We then state a convergence result in a form stronger than will be needed for our applications.

**Lemma.** i) Let \( y(t, w), y_n(t, w) \) \((n = 1, 2, \ldots)\) be stochastic processes: and suppose that \( y_n(t_m, w) \to y(t_m, w) \) a.e. \([P]\) for \( t_m \in R, m = 0, 1, 2, \ldots, k; \) then if \( \Lambda \) is any set in \( \mathcal{F} \) and \( A_m \) \((m = 1, 2 \ldots k)\) are collections of integers, we have

\[
\lim_{n} P \left[ \Lambda; y_n(t_m) \in A_m, m = 1, 2, \ldots, k \mid y_n(t_0) \right] = P \left[ \Lambda; y(t_m) \in A_m, m = 1, 2, \ldots, k \mid y(t_0) \right].
\]  

ii) If the \( y_n(t, w) \) are Markov processes and if for each \( t, \) \( \lim_{n} y_n(t, w) = y(t, w), \)

then \( y(t, w) \) is a Markov process.
iii) If $y_n(t, w) \in \mathfrak{M}$ with transitions $p_{ij}^{(n)}(t)$ then $y(t, w)$ will be in $\mathfrak{M}$ with transitions $p_{ij}(t)$ and $\lim_n p_{ij}^{(n)}(t) = p_{ij}(t)$ for any state $i$ such that $P\{y(\tau) = i\} > 0$

for some $\tau > 0$.

iv) Finally, under the conditions of the last sentence if $p_{ij}^{(n)}(t)$ are identical for all $n$, then $p_{ij}^{(n)}(t)$ will be a transition matrix for $y(t, w)$.

Proof: We prove only (i) the remainder of the lemma will then follow readily. The proof of (i) will be clear if we fix an integer $\{i\}$ such that $P\{y(\ell) = i\} > 0$ and let $k = 1$ where $A_i$ is the single integer $\{j\}$ and prove

$$\lim_n P[A; y_n(t_i) = j]/P[y_n(\ell_o) = i] = P[A; y(t_i) = j]/P[y(\ell_o) = i]. \quad (2)$$

Now let $\epsilon$ be any positive number; then we may pick a number $\eta > 0$ so that for any $\Delta \in \mathfrak{S}$ with $P(\Delta) < \eta$ we have

$$P[A \cap (y(t_i) = j) \cup \Delta]/P[y(\ell_o) = i] \cap \Delta] = P[A \cap (y(t_i) = j) \cap \Delta]/P[y(\ell_o) = i] \cup \Delta] < \epsilon. \quad (3)$$

Now by the Egorov theorem we may pick $\Delta$ and $N$ so that $P\Delta < \eta$ and $|y(\ell_o, w) - y_n(\ell_o, w)| < \epsilon/2$ for $w \in \Delta$, $k = 0,1$ and $n > N$. That is, for $n > N$, $w \not\in \Delta$, $y_n(\ell_o, w) = y(\ell_o, w)$; hence it follows from (3) that, for $n > N$,

$$|P[A \cap (y(t_i) = j)/P[y(\ell_o) = i] - P[A \cap (y(t_i) = j)]/P[y(\ell_o) = i]| < \epsilon,$$

and (2) is proved.

We proceed to the strong Markov theorem. Let $x(t, w)$ be a time-homogeneous Markov process with standard transitions and let $\mathfrak{B}$ denote the Borel sets of $R$; then Chung\(^2\) shows that $x(t, w)$ has a standard modification which is measurable $(\mathfrak{B} \times \mathfrak{S}, R \times \Omega)$ and indeed has almost all of its sample functions right lower semi-continuous in $t$ (the standard modification may take on the value $+ \infty$, but for each $t$ this value will be assumed only in a set of probability 0). We denote by $\mathfrak{M}^*$ the class $(x(t, w); \{p_{ij}(t)\})$, where $x(t, w)$ is in $\mathfrak{M}$ and has right lower semicontinuous sample functions and where $p_{ij}(t)$ is a standard transition matrix for $x(t, w)$. If $y(t), t \in T, T \subset R$ is some collection of random variables, then $\mathfrak{S}(y(t); t \in T)$ denotes the Borel field which they generate. A random variable $\alpha(w)$ with range some subset of $R$ will be called optional if $[w: \alpha(w) \leq t] \in \mathfrak{S}(x(s); s \leq t)$. A set $\Lambda$ will be called admissible if $[\Lambda \cap (\alpha(w) \leq t)] \in \mathfrak{S}(x(s); s \leq t)$ for each $t$.

**Theorem.** If $x(t, w) \in \mathfrak{M}^*$ and $\alpha(w)$ is optional, then $y(t, w) = x(\alpha(w) + t, w)$ is a stochastically continuous process in $\mathfrak{M}(t > 0)$ having the same transition as $x(t, w)$ and satisfying $P\{y(t) = \infty\} = 0$, $t > 0$. Furthermore, if $\Lambda$ is admissible and if $M_1 \in \mathfrak{S}(y(s); s < t^*)$ and $M_2 \in \mathfrak{S}(y(s); s \geq t^*)$ for some $t^*$, then

$$P[\Lambda M_1 M_2 | y(t^*)] = P[\Lambda M_1 | y(t^*)] \cdot P[M_2 | y(t^*)]. \quad (4)$$

**Proof:** Let $\alpha_n(w)$ be a sequence of optional random variables having denumerable range such that $\lim_n \alpha_n(w) = \alpha(w)$. Then the stochastic processes $y_n(t, w) = x(\alpha_n(w) + t, w)$ converge to $y(t, w)$ in measure on $(\mathfrak{B} \times \mathfrak{S}, R \times \Omega)$; this follows from the product space version of Auerbach's theorem.

Hence, by possibly choosing a subsequence, we may suppose that $\lim_n y_n(t, w) = y(t, w)$ almost everywhere on the product space. It is clear however that $y_n(t, w)$
is in $\mathfrak{M}$ with transitions $\{p_{ij}(t)\}$ and that (4) is satisfied with $y(t, w)$ replaced by $y_n(t, w)$; to see this, one need merely note that if $\alpha_n(w) = c$ on $\Delta \times \mathfrak{F}(x(t); t \leq c)$ where $P(\Delta) > 0$, then $y(t, w) = x(t + c, w)$ is in $\mathfrak{M}$ with transitions $\{p_{ij}(t)\}$ on the probability triple $(P(\cdot | \Delta), \Delta, \mathfrak{F}, \Delta)$. We now show that $y_n(t, w)$ form a sequence of Markov processes which are equi-uniformly right continuous in probability on any finite interval $[t_1, t_2]$ where $t_1 > 0$. We pick a $t^*$ with $0 < t^* < t_1$ such that $P[y(t^*) = \infty] = 0$ and $\lim_{n} y_n(t^*) = y(t^*)$ a.e. $[P]$; such a choice is possible by Fubini's theorem. Let $\epsilon > 0$ be assigned; then, using Egorov's theorem, we see that there exist positive integers $M$ and $N$ and a set $\Delta$ with $P(\Delta) < \epsilon$ such that for $n > N$, $w \in \Delta$ we have $y_n(t^*) = y(t^*) < M$. Let $p_t = P[y(t^*) = i]$; then, for $n > N$ and $t > t^*$,

$$P[y_n(t) \neq y_n(t + \tau)] \leq \sum_{i=1}^{M} \sum_{j=1}^{\infty} p_{ij} p_{ij}(t - t^*) \left[1 - p_{ij}(\tau)\right] + P(\Delta). \tag{5}$$

But the elements of a regular transition matrix are continuous functions; so, in view of Dini's theorem, the double sum in (5) converges to 0 as $\tau \to 0^+$ uniformly for $t$ in any finite interval. The uniform continuity in probability assertion is proved, and from this it follows that for each $t$, $y_n(t, w)$ is a Cauchy sequence in measure and the function $\bar{y}(t, w) = \lim_{n} y_n(t, w)$ is stochastically continuous, satisfying

$$P(\bar{y}(t, w) = \infty) = 0 \text{ for each } t > 0.$$  

By (iii) and (iv) of the lemma, $y(t, w)$ is in $\mathfrak{M}$ with transition matrix $\{p_{ij}(t)\}$, and by (i) of the lemma, (4) is satisfied with $y(t, w)$ replaced by $\bar{y}(t, w)$. We have left only to show that $\bar{y}(t, w)$ is a standard modification of $y(t, w)$. It follows from a theorem of Doob$^4$ that $y(t, w)$ has at most one finite right-hand limit point; hence if $\Omega_t = \{w; \bar{y}(t, w) \neq y(t, w)\}$; then by the right lower semicontinuity of $y(t, w)$ and the stochastic continuity of $y(t, w)$, we have that $y(t, w) = \infty$ a.e. $[P]$ in $\Omega_t$. But this implies that $P(\Omega_t) = 0$ and hence $\bar{y}(t, w)$ is indeed a standard modification of $y(t, w)$. 

The argument used here seems applicable to more general phase spaces. The arguments become more delicate however and our work in this direction is still incomplete. Professor S. Kakutani has shown us by example that the lemma is not true in its full strength even with such restrictions as continuity of the sample functions or denumerability of the phase space. Also D. Ray$^5$ has shown that right continuity of the sample functions is not sufficient to conclude the strong Markov property in the case where the phase space is the real line.$^6$

* This work was in part supported by the U.S. Air Force under contract No. AF 49(638)-184, monitored by the Office of Scientific Research.


$^2$ Definitions and results not explicitly referred to here may be found in J. L. Doob, *Stochastic Processes* (New York, 1953).

$^3$ It should be noted that a given Markov process $x(t, w)$ may have many transitions and that the property of being time-homogeneous is not invariant under the choice of transitions. When we say that a process is time-homogeneous we mean that it has a set of stationary transitions, or more specifically that for any $\tau_j$ with $P[x(\tau_k) = i] > 0 \ (k = 1, 2)$ we have $P[x(t + \tau_j) = j | x(\tau_i) = i] = P[x(t + \tau_j) = j | x(\tau_i) = i]$ for $t > 0, j = 1, 2, \ldots$. 

---

**Vol. 44, 1958**

**MATHEMATICS: AUSTIN**

577


A. Added in proof. It appears that Chung’s result that \( y(t, w) \in R^* \) on \( t \geq 0 \) with respect to the triple \( (P(\cdot \Delta), \Delta F, \Delta) \) where \( \Delta = \cup \Delta_j = \cup \Delta_j \{ w; x(\alpha(w), w) = j \} \) will follow using our proof if one chooses the approximating \( \alpha_n(w) \) more delicately. Let \( R_p = \{ 2^{-m}; m = 1, 2, \ldots \} \) for \( p = 1, 2, \ldots \) and let \( R = \cup R_p \) be a separability set for \( x(t, w) \). Since \( x(t, w) \in R^* \) the set \( \{ w; w \in \Delta_j, x(w) < t \} \) for \( \Delta_j \leq \Delta \}. \) Let \( \pi_p[0 = \alpha_1, \ldots, \alpha_p < \ldots \} \) be a sequence of partitions of \( [0, \infty) \) with norms tending to 0 and define the optional random variables as \( \alpha_p(w) = a_k \) on \( B_k = \{ w; \alpha_{k-1} < \alpha \leq \alpha_k \} \) with left inclusion at \( k = 1 \). Fix integers \( j, k \) and \( p \) and let \( A_i = A_i(k, j) = \{ w; w \in \Delta_j \cap B_k, x(s, w) = j \} \) where \( s_1 < s_2 < \ldots < s_q \) is an ordering of \( R_p \cap \{ a_{k-1}, a_k \} \).

We define \( \alpha_p(w) = s_i \) on \( A_i = U A_i, l = 1, 2, \ldots q; \) the r.v.’s are defined similarly for each \( k, j \) and we let \( \alpha_p(w) = \alpha_p(w) \) elsewhere. The r.v.’s \( \alpha_p(w) \) are optional and satisfy \( \alpha_p(w) \downarrow \alpha(w) \) and \( x(\alpha_p, w) \to x(\alpha, w) \) on \( \Delta; \) hence we may let \( t_i = t^* = 0 \) on the triple \( (P(\cdot \Delta), \Delta F, \Delta) \) in the proof of our theorem and obtain the stronger result.

Finally we mention that Jushkevich has quite recently (Russian J. of Prob.) obtained results on the strong Markov property in a different form from that presented here. Also Chung has employed the continuity of the conditional distribution of \( y(t) \) relative to \( \alpha \) to show that \( y(t) \) is separable without modification.

\[ DYNAMIC\ PROGRAMMING,\ SUCCESSIVE\ APPROXIMATIONS,\ AND\ MONOTONE\ CONVERGENCE \]

\[ \text{BY Richard Bellman} \]

\[ \text{RAND CORPORATION, SANTA MONICA, CALIFORNIA} \]

\[ \text{Communicated by Philip M. Morse, April 21, 1958} \]

1. Introduction.—Our object in this paper is to show that a blend of dynamic programming, successive approximations, and digital computers enables us to approach various classes of variational problems formerly far beyond our reach.

To illustrate the application of these methods, we shall consider two problems.

The first is that of minimizing the functional

\[ J(v) = \int_0^T F(x_1, x_2, \ldots, x_N)dt + G(x_1(T), x_2(T), \ldots, x_N(T)) \quad (1.1) \]

over all forcing functions \( v_i(t) \) related to the \( x_i(t) \) by means of relations of the type

\[ \frac{dx_i}{dt} = H_i(x_1, x_2, \ldots, x_N) + v_i(t), \quad x_i(0) = e_i, \quad i = 1, 2, \ldots, N, \quad (1.2) \]

and subject to the constraints of the form

\[ a) \int_0^T K_j(v_1, v_2, \ldots, v_N)dt \leq b_j, \quad j = 1, 2, \ldots, L, \]

\[ b) p_i(t) \leq v_i(t) \leq q_i(t), \quad 0 \leq t \leq T, i = 1, 2, \ldots, N. \quad (1.3) \]

The second problem is a generalized Hitchcock-Koopmans transportation problem. It involves the minimization of the function \( C(x) = \sum g_i(x_i) \) subject to the constraints