ON TRANSIENT MARKOV CHAINS WITH APPLICATION TO THE UNIQUENESS PROBLEM FOR MARKOV PROCESSES[1]

By Leo Breiman

University of California, Berkeley

1. Summary. We focus our attention herein on a Markov chain $x_0, x_1, \cdots$ with a countable number of states indexed by a subset I of the integers and with stationary transition probabilities p_{ij}, and explore the sets of states defined by:

A transient set of states C is said to be *denumerably atomic* if $P(x_n \varepsilon C \text{ i.o.}) > 0$ and if for every infinite set $A \subset C$ we have $x_n \varepsilon C$ i.o. implies $x_n \varepsilon A$ i.o. with probability one (a.s.).

Following Blackwell's basic paper [1] which introduced the systematic use of martingales into the study of Markov chains, we use the semi-martingale convergence theorem [2] to characterize denumerably atomic sets in terms of the bounded solutions of the inequality

$$\phi(i) \leqq \sum_{j \varepsilon I} p_{ij}\phi(j), \qquad i \varepsilon I.$$

For chains whose state space contains a denumerably atomic set a convergence criterion for certain sums $\sum_{n=0}^{\infty} f(x_n)$ is then developed. The application of this criterion to a restricted class of continuous parameter Markov processes gives simple necessary and sufficient conditions for the existence of a unique process satisfying given infinitesimal conditions. This last result illuminates the connection between the necessary and sufficient conditions given by Feller [3] for uniqueness and the simpler conditions for birth and death processes given recently by Dobrusin [4], more recently by Karlin and McGregor [5], and by Reuter and Lederman [6] (see also [7]).

2. Characterization theorem.

Theorem 1. *The necessary and sufficient condition for a transient set of states C such that $P(x_n \varepsilon C$ i.o.$) > 0$ to be denumerably atomic is that any bounded solution $\phi(i)$ of*

$$\text{(A)} \qquad \phi(i) \leqq \sum_{j \varepsilon I} p_{ij}\phi(j)$$

satisfy $\liminf_{i \varepsilon C} \phi(i) = \limsup_{i \varepsilon C} \phi(i)$.

Proof. Let C be denumerably atomic and $\phi(i)$ any bounded solution of (A). Then $E(\phi(x_n) \mid x_{n-1}, \cdots, x_0) \geqq \phi(x_{n-1})$ so that by the semi-martingale convergence theorem $\phi(x_n)$ converges a.s. to a function $f(\omega)$. Let A_1, A_2 be infinite subsets of C such that for $i \varepsilon A_1$, $\phi(i) < \alpha$ and for $i \varepsilon A_2$, $\phi(i) > \beta > \alpha$. Then, since almost every sample path which is in A_1 i.o. is in A_2 i.o. the limit of $\phi(x_n)$ cannot exist.

Received July 2, 1956.

[1] This paper was prepared with the support of the Office of Ordnance Research, U. S. Army under Contract DA-04-200-ORD-171.

Conversely, let A be any infinite subject of C and take $\mathcal{E}$ to be the event that x_n is never in A. Let $\phi(i) = P(\mathcal{E} \mid x_0 = i)$, then:

$$\phi(i) = E(P(\mathcal{E} \mid x_0, x_1) \mid x_0 = i) = \textstyle\sum_{j \varepsilon I-A} p_{ij}\phi(j) \leqq \sum_{j \varepsilon I} p_{ij}\phi(j).$$

Since $\phi(i)$ is zero on A, we have $\liminf_{i \varepsilon C} \phi(i) = 0$ whence $\limsup_{i \varepsilon C} \phi(i) = 0$. This implies that for almost every sample path $x_0, \cdots$ which is in C i.o., $\phi(x_n)$ converges to zero. But by the martingale convergence theorem, since $P(\mathcal{E} \mid x_n, \cdots, x_0) \leqq \phi(x_n)$, the indicator $I_{\mathcal{E}}(\omega)$ of the set $\mathcal{E}$ is zero for almost every $\omega \varepsilon [x_n \varepsilon C \text{ i.o.}]$ and therefore almost every such ω is in A i.o.

We note, for future use, that if C is denumerably atomic and if $\phi(i)$ is a solution of (A) satisfying $E \mid \phi(x_n) \mid < K$, the conclusion $\liminf_{i \varepsilon C} \phi(i) = \limsup_{i \varepsilon C} \phi(i)$ remains unaltered.

It is interesting, as well as necessary, to know that any denumerably atomic set C can be embedded in a set $\bar{C}$ which is a maximal denumerably atomic set. That this is so follows from Blackwell's work [1] in the following sense: *there is a set* $\bar{C} \supset C$ *such that* $\bar{C}$ is denumerably atomic and $x_n \varepsilon \bar{C}$ i.o. implies $x_n \varepsilon \bar{C}$ for all sufficiently large n a.s.

3. Convergence criterion. The above characterization leads to a convergence criterion reminiscent of the Three-Series theorem.

THEOREM 2. *Let C be denumerably atomic and $f(i)$ a finite nonnegative function on I such that f is zero outside of C. Then the sum $\sum_{n=0}^{\infty} f(x_n)$ converges* a.s. *if and only if $\sum_{n=0}^{\infty} Ef(x_n)$ converges and otherwise diverges with probability equal to* $P(x_n \varepsilon C \text{ i.o.})$.

PROOF. Let $S = \sum_{n=0}^{\infty} f(x_n)$ and $\phi(i) = P(S < d \mid x_0 = i)$. Then $\phi(i)$ satisfies inequality (A). Suppose $\liminf_{i \varepsilon C} \phi(i) = 0$ for every value of d, then $\limsup_{i \varepsilon C} \phi(i) = 0$, and $\phi(x_n) \to 0$ a.s. on the set $[x_n \varepsilon C \text{ i.o.}]$. Since

$$P(S < d \mid x_n, \cdots, x_0) \leqq \phi(x_n)$$

it follows that $I_{[S<d]}(\omega) = 0$ a.s. for $\omega \varepsilon [x_n \varepsilon C \text{ i.o.}]$ and hence that S diverges a.s. on this set.

From the above it follows that if S converges on some subset of $[x_n \varepsilon C \text{ i.o.}]$ of positive measure, there is a $\delta > 0$ and $d_1 > 0$ such that $P(S < d_1 \mid x_0 = i) \geqq \delta$ for $i \varepsilon C$. Let $R_m = \sum_{n=m}^{\infty} f(x_n)$ and define S_m as the set $[R_m < d_1]$. Writing

$$\int_{S_m} R_m - \int_{S_{m+1}} R_{m+1} = \int_{S_m} [R_m - R_{m+1}] - \int_{S_{m+1}-S_m} R_{m+1},$$

using the definition of S_m

$$\int_{S_m} f(x_m) \leqq d_1 P(S_{m+1} - S_m) + \int_{S_m} R_m - \int_{S_{m+1}} R_{m+1},$$

and summing over all m results in

$$\sum_{m=0}^{\infty} \int_{S_m} f(x_m) < 2\, d_1.$$

But,

$$\int_{S_m} f(x_m) = \sum_{i \varepsilon C} f(i)P(R_m < d_1 \mid x_m = i)\, P(x_m = i)$$

$$= \sum_{i \varepsilon C} f(i)P(S < d_1 \mid x_0 = i)P(x_m = i) \geqq \delta Ef(x_m)$$

which proves the theorem.

COROLLARY 1. *Under the conditions of the above theorem, a necessary and sufficient condition for the* a.s. *convergence of* $\sum_{n=0}^{\infty} f(x_n)$ *is that the equation*

$$\text{(B)} \qquad a(i) = f(i) + \sum_{j \varepsilon I} p_{ij} a(j)$$

have a bounded solution.

PROOF. Let $\sum_{n=0}^{\infty} f(x_n)$ converges a.s. to $S(\omega) < \infty$. Then $a(i) = E(S \mid x_0 = i)$ is a solution of (B) and $-a(i)$ is a solution of (A) with $Ea(x_n) \leqq ES$. If $a(i)$ is unbounded, then $\lim \sup_{i \varepsilon C} a(i) = \lim \inf_{i \varepsilon C} a(i) = \infty$, which implies that $a(x_n) \to \infty$ on a set of positive measure and contradicts the boundedness of $Ea(x_n)$. Conversely, if (B) has a bounded solution $a(i)$, then the iteration of (B) gives $| a(i) - E(\sum_{n=0}^{\infty} f(x_n) \mid x_0 = i) | \leqq \sup_{j \varepsilon I} a(j)$ which implies the convergence of $\sum_{n=0}^{\infty} Ef(x_n)$.

We relate Theorem 2 to the uniqueness problem which involves global structure, by confining ourselves to chains with a fairly simple decomposition. The following theorem is appropriate. Its proof follows immediately from the various definitions.

THEOREM 3. *Let the state space I of a Markov chain be completely decomposable into the set C_0 of recurrent states, a set M of transient states such that $P(x_n \varepsilon M$ i.o.$) = 0$, and a finite number of maximal denumerably atomic sets $C_1, \cdots, C_N$. If f is a function on I and if f_k is that function which equals f on C_k and is zero elsewhere, then $\sum_{n=0}^{\infty} f(x_n)$ diverges almost surely if and only if each $\sum_{n=0}^{\infty} f_k(x_n)$ diverges* a.s. *on the set* $[x_n \varepsilon C_k$ i.o.$]$.

4. The uniqueness problem. The synthetic uniqueness problem for continuous parameter Markov processes having states indexed by a subset I of the integers begins with a set of nonnegative constants q_i, p_{ij}, defined for $i, j \varepsilon I$, and asks concerning the existence of a unique process $X(t)$, $0 \leqq t < \infty$, having a given initial distribution and satisfying

(C)
i. $P(X(t)$ is constant in interval $[s, s + \tau] \mid X(s) = i) = 1 - q_i\tau + o(\tau)$
ii. P (First discontinuity of $X(t)$, $t \geqq s$, is a jump to $j \mid X(s) = i) = p_{ij}$.

Our remarks are restricted to the simple and common situation $q_i, p_{ij} < \infty$, $\sum_{j \varepsilon I} p_{ij} = 1$.

There is a general answer, [3], [6], and [8]: if no "explosions" are possible, if an infinite number of jumps cannot occur in any finite time interval, then there is a unique process satisfying (C) and having, in addition, all of the properties that could reasonably be desired. This is the "minimal" solution. In the contrary case, there is in general no unique solution and the solutions that do exist are analytically or probabilistically pathological.

To be more exact; the traversal time of each path of infinite length $(i_1, i_2, \cdots)$ is a sum $Q_{i_1} + Q_{i_2} + \cdots$ of independent random variables with distributions $P(Q_{i_n} > t) = \exp(-q_{i_n}t)$. There is a Markov measure $\bar{P}$ on the space of all paths induced by the p_{ij} and the given initial distribution. The minimal solution exists if and only if the transversal time is a.s. infinite for each path in a set of $\bar{P}$ measure one. Using the Three-Series criterion for the divergence of a sum of independent random variables and writing $x_0, x_1, \cdots$ for the chain associated with the measure $\bar{P}$ leads to an equivalent formulation.

Uniqueness criterion. A minimal solution exists if and only if $\sum_0^\infty 1/q_{x_n}$ diverges a.s.

The applicability of Theorem 2 is now apparent. For instance, in the birth and death process, the given constants are

if $i \geqq 1$, $q_i = \lambda_i + \mu_i$, $p_{i,i+1} = \lambda_i/(\lambda_i + \mu_i)$, $p_{i,i-1} = \mu_i/(\lambda_i + \mu_i)$, $p_{ij} = 0$ otherwise;

if $i = 0$, $q_i = 0$, $p_{00} = 1$, $p_{ij} = 0$ otherwise.

If return to the origin is uncertain, the positive integers form a maximal denumerably atomic set. The condition for the existence of a minimal solution, as given by Corollary 1, is that the equation

$$(\lambda_i + \mu_i)a(i) = 1 + \lambda_i a(i + 1) + \mu_i a(i - 1), \qquad i \geqq 1$$

have no bounded solution. A little formal computation yields the condition as stated in [4], [5], and [7].

Another interesting application is to the case

$$p_{ij} = p_{j-i}, \qquad 0 < \textstyle\sum_{j \varepsilon I} j p_j < \infty,$$

where we restrict the state space I to those states with a positive probability of being entered. As pointed out to me by D. Blackwell, the basic theorem of renewal theory, Chung and Wolfowitz [9], provides the simplest proof that the nonnegative integers I^+ in I form a maximal denumerably atomic set. By this theorem, the expression

$$E\,(\text{number of entrances into } j \mid x_0 = i) = \sum_{n=0}^{\infty} p_{ij}^{(n)}$$

approaches a positive limit as $j \to +\infty$ through I, which implies that for any infinite set A of positive integers in I, $P(x_n \varepsilon A \text{ i.o.}) = 1$. As the negative integers I^- in I have the property $P(x_n \varepsilon I^- \text{ i.o.}) = 0$, the necessary and sufficient condition for the existence of the minimal solution is

$$\sum_{n=0}^{\infty} \sum_{j \varepsilon I^+} \frac{1}{q_j} p_{ij}^{(n)} = \infty.$$

Interchanging the order of summation, and applying the renewal theorem once more gives the equivalent condition

$$\sum_{j\varepsilon I^+} \frac{1}{q_j} = \infty.$$

A slight alteration of this discussion is sufficient to establish the same condition when the negative integers are absorbing states, that is, if

$$p_{ij} = p_{j-i} \quad \text{if} \quad i \geqq 0; \qquad q_i = 0 \quad \text{if} \quad i < 0.$$

If the state space of the chain $x_0, x_1, \cdots$ cannot be decomposed as indicated in Theorem 3, complications set in and Feller's criterion, which necessitates a close examination of every set of states A such that $P(x_n \varepsilon A \text{ all } n \mid x_0 \varepsilon A) > 0$ must be referred to. Simple necessary and sufficient conditions for uniqueness are possible only when some uniformity, such as denumerable atomicity, is present.

REFERENCES

[1] D. Blackwell, "On transient Markov processes, with a countable number of states and stationary transition probabilities," *Ann. Math. Stat.*, Vol. 26 (1955), pp. 654–658.

[2] J. L. Doob, *Stochastic Processes*, Chapter 7, John Wiley and Sons, New York, 1953.

[3] W. Feller, "On the integro-differential equations of purely discontinuous Markov processes," *Trans. Amer. Math. Soc.*, Vol. 48 (1940), pp. 488–515.

[4] R. L. Dobrusin, "On the conditions of regularity of stationary Markov processes with a denumerable set of possible states," *Progr. Math. Sci. Moscow* (N.S.), Vol. 7 (1952), pp. 185–191.

[5] S. Karlin and J. McGregor, "Representation of a class of stochastic processes," *Proc. Nat. Acad. Sci.*, Vol. 41 (1955), pp. 387–391.

[6] W. Lederman and G. E. H. Reuter, "On the differential equations for the transition probabilities of Markov processes with enumerably many states," *Proc. Cambridge Philos. Soc.*, Vol. 49 (1953), pp. 247–262.

[7] M. S. Bartlett, *An Introduction to Stochastic Processes*, Cambridge, 1955.

[8] J. L. Doob, "Markov chains—denumerable case," *Trans. Amer. Math. Soc.*, Vol. 63 (1945), pp. 455–473.

[9] K. L. Chung and J. Wolfowitz, "On a limit theorem in renewal theory," *Ann. Mathematics*, Vol. 55 (1952), pp. 1–6.