A LIMIT THEOREM FOR PASSAGE TIMES IN ERGODIC REGENERATIVE PROCESSES

By Julian Keilson[1]

Sylvania Electronics Systems

1. Introduction. In a previous note, "A Technique for Discussing the Passage Time Distribution for Stable Systems" (Keilson, 1965) it was pointed out that for an ergodic Markov diffusion process or birth-death process on a state space having an inaccessible boundary, the distribution of passage times to states of low probability was approximately exponential, and was asymptotically exponential for any sequence of states approaching the inaccessible boundary. It was stated that this well known limiting exponential behavior was to be expected for a much broader class of ergodic processes with inaccessible states, but no explicit results were given. In this note, a theorem will be presented demonstrating such behavior for any ergodic regenerative process in continuous time.

The reader is reminded that a regenerative process in continuous time (which may or may not be Markov) is a temporally homogeneous process $\mathbf{X}(t)$ on an abstract state space of elements $\mathfrak{x}$ characterized by an imbedded sequence of regenerating events having a positive probability of recurrence. (The term event is used here to mean an occurrence in time, with zero duration in time, such as an arrival to or departure from a given set of states. For a given regenerative process, there may be many classes of regenerating events available and attention will focus on some single specified class.) Each such regeneration destroys the "memory" of a process sample, i.e., the statistical behavior of a sample subsequent to such a regeneration is independent of the history of the sample before the regeneration. The time intervals separating successive regenerations then constitute a sequence of independent identically distributed positive random variables. When these have a finite expectation, a renewal process may be associated with the regenerations, and the regenerative process $\mathbf{X}(t)$ is then ergodic, i.e., for any subset A of $\mathfrak{X}$ and any initial state $\mathbf{z}$, $\lim_{t\to\infty} P(\mathbf{X}(t) \varepsilon A \mid \mathbf{X}(0) = \mathbf{z})$ is a non-degenerate measure $P_\infty(A)$ independent of $\mathbf{z}$. For an extensive discussion of such regenerative processes see W. L. Smith (1955), and J. F. C. Kingman (1964).

It will be convenient and will entail no loss of generality to regard the process $\mathbf{X}(t)$ as being a multivariate Markov process, with the number of random variables finite or infinite, and to focus attention on the regeneration events associated with arrival to, departure from, or passage through some particular state $\mathbf{x}_0$ in the state space $\mathfrak{X}$. We will be interested in a sequence of passage time distributions $F_N(x)$ defined in the following way. For each value N there is given a decomposition $\mathfrak{x} = \mathfrak{x}_1^{(N)} + \mathfrak{x}_2^{(N)}$ of the state space into two disjoint subspaces

Received 10 January 1966; revised 22 March 1966.

[1] Now at University of Rochester.

with $P_\infty(\mathfrak{X}_1^{(N)}) > 0$ and $P_\infty(\mathfrak{X}_2^{(N)}) > 0$, and we will suppose that the regeneration state $\mathbf{x}_0$ is always in $\mathfrak{X}_1^{(N)}$. The distribution $F_N(x)$ is the pdf for the first passage time from a regeneration event to the set $\mathfrak{X}_2^{(N)}$. For the decomposition $\mathfrak{D}_N$ we define the conditional regeneration probability p_N to be the probability that a system which has just regenerated will have its next regeneration without having visited $\mathfrak{X}_2^{(N)}$.

Let T_N denote the time interval between the $(n - 1)$st and nth regeneration. Let its measure be denoted by

$$\mu(A) = P(T_n \,\varepsilon\, A), \tag{1}$$

and its finite expectation be denoted by

$$m = E(T_n) = \int_0^\infty t\mu\,(dt). \tag{2}$$

Our basic theorem may now be stated.

THEOREM. *Let $\mathbf{X}(t)$ be an ergodic regenerative process in continuous time on the state space $\mathfrak{X}$. Suppose there is available a sequence of decompositions $\mathfrak{D}_N$ of $\mathfrak{X}$, for which the sequence of conditional regeneration probabilities p_N defined above has the limit 1 when $N \to \infty$. Let ζ_N be the first passage time from the zeroth regeneration to $\mathfrak{X}_2^{(N)}$. Then, as $N \to \infty$,*

$$[(1 - p_N)/m]E(\zeta_N) \to 1. \tag{3}$$

Moreover, for each $x \geqq 0$,

$$\lim_{N\to\infty} P\{[(1 - p_N)/m]\zeta_N \leqq x\} = 1 - e^{-x}. \tag{4}$$

Equation (4) states that the distribution $F_N(x)$ of the passage time ζ_N is asymptotically exponential. The class of ergodic regenerative processes to which the limit theorem is applicable includes univariate Markov diffusion processes, birth-death processes on the lattice of positive integers, more general lattice processes in continuous time (since any skip-free character of the motion is irrelevant), the classical fiber spinning process (cf Daniels, 1945) and related congestion process describing the number of calls in a trunk of infinite capacity, the queue length process and virtual waiting time process for the G/G/1 queue, and a great variety of other processes arising in queueing theory, reliability theory, operations research, etc. A few of these applications are discussed in Section 3.

2. Proof of the theorem. Let the process $\mathbf{X}(t)$ commence at $t = 0$ with the zeroth regeneration. Let $A_n^{(N)}$ denote the event that $\mathbf{X}(t) \,\varepsilon\, \mathfrak{X}_2^{(N)}$ for some time t between the $(n - 1)$st and nth regeneration. Then one must have

$$P(A_n^{(N)}) = 1 - p_N = q_N > 0. \tag{5}$$

The events $A_n^{(N)}$ $(n = 1, 2, \cdots)$ are independent. Let n_N denote the smallest value of n for which $A_n^{(N)}$ occurs. Then for the first passage time ζ_N, one must have

$$0 \leqq \zeta_N \leqq T_1; \qquad n_N = 1 \tag{6}$$
$$\textstyle\sum_1^{n_N-1} T_n \leqq \zeta_N \leqq \sum_1^{n_N} T_n; \qquad n_N > 1.$$

Moreover,

$$P(n_N = n) = p_N^{\,n-1} q_N. \tag{7}$$

Let $\mu_{1N}(A)$ be the regeneration time distribution for paths in which the set $\mathfrak{X}_2^{(N)}$ is not reached, and let $\mu_{2N}(A)$ be the regeneration time distribution for paths in which $\mathfrak{X}_2^{(N)}$ is reached, i.e., let

$$\mu_{1N}(A) = P(T_1 \varepsilon A \mid \overline{A_1^{(N)}})$$

and

$$\mu_{2N}(A) = P(T_1 \varepsilon A \mid A_1^{(N)}),$$

so that

$$\mu(A) = p_N \mu_{1N}(A) + q_N \mu_{2N}(A). \tag{8}$$

For the subset of paths for which $n_N = n$, the regeneration times $T_1, T_2, \cdots, T_n$ are independent random variables with $T_1, T_2, \cdots, T_{n-1}$ having the distribution $\mu_{1N}(A)$ and T_n having the distribution $\mu_{2N}(A)$. For this subset of paths, $T_1 + T_2 + \cdots + T_{n-1}$ has the moment generating function $\phi_{1N}^{n-1}(s)$ and $T_1 + T_2 + \cdots + T_n$ has the m.g.f. $\phi_{1N}^{n-1}(s)\phi_{2N}(s)$, where

$$\phi_{iN}(s) = \textstyle\int_0^\infty e^{-st} \mu_{iN}(dt), \qquad i = 1, 2, .$$

It then follows from (6) and (7) by summation over n that, for $s > 0$,

$$q_N/[1 - p_N\phi_{1N}(s)] \geqq E(e^{-s\zeta_N}) \geqq q_N\phi_{2N}(s)/[1 - p_N\phi_{1N}(s)]; \tag{9}$$

and that

$$(p_N/q_N)m_{1N} \leqq E(\zeta_N) \leqq (p_N/q_N)m_{1N} + m_{2N}, \tag{10}$$

where $m_{iN} = \int_0^\infty t\mu_{iN}(dt)$. We note from (8) that $p_N m_{1N} + q_N m_{2N} = m$. (Note from (10) that for any $\mathfrak{D}_N$, $q_N E(\zeta_N) \leqq m$. Consequently, if it is known that $E(\zeta_N) \to \infty$ for a sequence $\mathfrak{D}_N$, it follows that $q_N \to 0$, and $p_N \to 1$, and the theorem is available.) Hence (10) implies (3) provided that

$$q_N m_{2N} \to 0. \tag{11}$$

Moreover, from (8), $p_N\phi_{1N}(s) + q_N\phi_{2N}(s) = \phi(s)$, where $\phi(s)$ is the m.g.f. of $\mu(A)$. Hence

$$[1 - p_N\phi_{1N}(q_N s)]/q_N = [1 - \phi(q_N s)]/q_N + \phi_{2N}(q_N s). \tag{12}$$

When $q_N \to 0$, the first term on the right-hand side goes to ms when $N \to \infty$, by virtue of (2). If it were further true that

$$\phi_{2N}(q_N s) \to 1, \qquad s > 0, \tag{13}$$

it would follow from (9) and (12) that

$$\lim_{N\to\infty} E(\exp(-sq_N\zeta_N)) = 1/(1 + ms) \tag{14}$$

for all $s > 0$, and the theorem would follow at once from a variant of the continuity theorem for characteristic functions, as given for example in Feller (1966), p. 408. It only remains to prove (11) and (13). We note that (13) follows from (11), since

$$0 \leqq [1 - \phi_{2N}(q_N s)]/q_N = \int_0^\infty \{(1 - \exp(-q_N st))/(q_N st)\}\, st\mu_{2N}(dt)$$
$$\leqq \int st\mu_{2N}(dt) = sm_{2N}\,.$$

To prove (11), let $\epsilon > 0$ be given, and choose $t_0 > 0$ such that $\int_{t_0}^\infty t\mu(dt) < \epsilon$. Since $\mu(A) - q_N\mu_{2N}(A) \geqq 0$, by (8), we have $\int_{t_0}^\infty tq_N\mu_{2N}(dt) < \epsilon$. Moreover, $\int_0^{t_0} tq_N\mu_{2N}(dt) \leqq t_0 q_N$. Consequently,

$$q_N m_{2N} = \int_0^\infty tq_N\mu_{2N}(dt) < \epsilon + t_0 q_N < 2\epsilon$$

as soon as $q_N < \epsilon/t_0$, proving (11). QED.

3. Applications. The classical fiber spinning process alluded to in the introduction was discussed in Keilson (1965). This is a process $K(t)$ on the nonnegative lattice of integers, samples of which may be represented as

$$K(t) = \sum_1^\infty H(t - \tau_i\,, T_i)$$

where $H(t, T) = 1$ in the closed interval $[0, T]$, and vanishes outside this interval. The sequence of epochs $\tau_1, \tau_2 \cdots$ are associated with a Poisson process of given intensity, and the sequence of durations $T_1, T_2 \cdots$, constitute a sequence of independent identically distributed random variables having a prescribed distribution. This process is ergodic if the durations T_i have a finite positive expectation. Imbedded in the process will be the regenerative epochs at which $K(t)$ returns to the value zero. The conditional regeneration probabilities p_N for return to zero without having reached the level N go to one as $N \to \infty$. Hence by our basic theorem, the passage time distributions to level N from the state zero are asymptotically exponential, as conjectured in Keilson (1965).

As a second example, consider the queue-length process $K(t)$ for a $G/G/1$ type queue. Such a queue is characterized by a single server who services customers individually, and a customer stream for which the sequence of successive interarrival times and the sequence of successive service time requirements are independent identically distributed random variables each having a given distribution function. When the mean interarrival time exceeds the mean service time, the queue length process is stable. The server will thereupon always return to a state of idleness and there will be a subsequent epoch at which the next customer appears. Such epochs constitute a sequence of regenerative events for the queue length process $K(t)$. We then may infer via our basic theorem that the passage time distribution from the regeneration epochs to a queue length N become asymptotically exponential.

The mean passage time to a given level will often be required for practical application of the theorem. This parameter will in many cases be available directly from the ergodic distribution for the process, as described in Keilson (1965).

4. Acknowledgment. I would like to thank H. Young and P. Belmont for helpful discussion. I would also like to thank J. H. B. Kemperman for many useful suggestions and for his generous aid in refining the proof of the theorem.

REFERENCES

DANIELS, H. E. (1945). The statistical theory of the strength of bundles of threads. *Proc. Roy. Soc. London Ser. A* **183** 405–435.

FELLER, W. (1966). *Probability Theory and its Applications*, Vol. II. Wiley, New York.

KEILSON, J. (1965). A technique for discussing the passage time distribution for stable systems. *RN 596, Applied Research Laboratory, Sylvania Electronic Systems.* (To appear in the *J. Roy. Statist. Soc. Ser. B.*)

KINGMAN, J. F. C. (1964). The stochastic theory of regenerative events. *Z. Wahrscheinlichkeitstheorie* **2** 180–224.

SMITH, W. L. (1955). Regenerative stochastic processes. *Proc. Roy. Soc. London Ser. A* **232** 6–31.