STOPPING TIMES FOR RECURRENT MARKOV PROCESSES

BY

J. R. BAXTER AND R. V. CHACON

1. Introduction

Let $\{\xi_n\}$ be a discrete-time Markov process with stationary transition probabilities, and let μ be the distributions of ξ_0. Let τ be a randomized stopping time, and let ν be the distribution of ξ_τ. Then say that μ can be balayaged to ν, and write $\mu \to \nu$, or $\mu \to \nu(\tau)$ to indicate the stopping time that effects the balayage. In this paper we consider the problem of giving an analytical expression for $E[\tau]$ when $\mu \to \nu(\tau)$.

This problem has a well-known solution in the transient case. Let P be the transition operator of the process, and define the potential operator $G \equiv \sum_{k=0}^{\infty} P^k$. If $\mu \to \nu(\tau)$ then $E[\tau] = \int (\mu - \nu)G$. This is the discrete-time analogue of the case of Brownian motion in dimension three or higher. In the Brownian motion case for dimension one or two the potential still exists as an operator on differences of probability measures, and the same formula remains valid. The discrete-time analogue of this situation would be a recurrent process such that the potential exists and such that $\mu P^n \to 0$ as $n \to \infty$ for all probability measures μ. For such a process the above formula is again valid, and we shall not deal with this case further. A general discussion of potential theory for recurrent processes is given in [9] and [11].

In the present paper we wish to consider processes which are strongly recurrent. It is assumed that $\mu P^n \to \lambda$ as $n \to \infty$ for all probability measures μ, where λ is an invariant probability measure, and that the potential operator G exists for differences of probability measures. It is then shown that $\mu \to \nu$ if and only if the negative part of $(\mu - \nu)G$ is of the form $\phi \, d\lambda$, that is, absolutely continuous with respect to λ. Furthermore

$$\text{ess sup } |\phi| = \min \{E[\tau] \mid \mu \to \nu(\tau)\}.$$

Thus a supremum has replaced the integral which occurred in the transient case.

2. Balayage sequences

Let $(S, \mathscr{B})$ be a measurable space and let $\xi = \{\xi_n, n = 0, 1, \ldots\}$ be a discrete-time Markov process with state space S, having stationary transition probabilities $p(x, A)$, where p is a Markov kernel on $S \times \mathscr{B}$. If μ is a measure on $\mathscr{B}$, μP will as usual denote the measure

$$\mu P(A) = \int \mu(dx) p(x, A). \tag{2.1}$$

Received April 14, 1975.

Let η be a random variable independent of ξ and let

$$\mathscr{F}_n = \sigma(\xi_0, \ldots, \xi_n, \eta), \quad n = 0, 1, 2, \ldots.$$

A stopping time relative to $(\mathscr{F}_n)$ will be called a *randomized stopping time.*

Let μ be the distribution of ξ_0. Let

$$(2.2) \quad r_n(A) = P(\xi_n \in A, \tau > n) \quad \text{and} \quad r_n'(A) = P(\xi_n \in A, \tau = n) \quad \text{for all } A \text{ in } \mathscr{B}, 0 \leq n < \infty.$$

It is easy to show that

$$(2.3) \qquad r_{n+1} + r_{n+1}' = r_n P \quad \text{for } 0 \leq n < \infty,$$

$$(2.4) \qquad r_0 + r_0' = \mu.$$

Conversely, if r_n and r_n' are sequences of nonnegative measures satisfying (2.3) and (2.4), it is a straightforward matter to construct a randomized stopping time τ such that (2.2) holds. (Cf. [10].)

We shall call a double sequence $r = (r_n, r_n')$ satisfying (2.3) a *balayage sequence.* For our purposes, we can work entirely with such sequences, rather than with the stopping times that induce them. The *initial measure* of the sequence is defined to be $r_0 + r_0'$. The *terminal measure* of the sequence is $\sum_{k=0}^{\infty} r_k'$. If μ is the initial and ν is the terminal measure of a balayage sequence r, we shall write $\mu \to \nu$ or $\mu \to \nu(r)$, in accordance with the notation of Section 1. We shall write $\sum_{k=0}^{\infty} r_k(S) = E[r]$, since this is $E[\tau]$ for any τ which induces r.

The arguments in the rest of the paper can be readily visualized if one keeps the following facts in mind: r_n is the mass still in motion at time n (i.e., not yet stopped). r_n' is the mass that was in motion prior to time n but which has just been stopped at time n. $\sum_{k=0}^{n} r_k'$ is the total mass stopped at times up to and including time n. $\sum_{k=0}^{\infty} r_k'$ is the total mass stopped. $\sum_{k=0}^{n} r_k' + r_n$ is all the mass at time n (i.e., the distribution of $\xi_{\tau \wedge n}$).

The balayage defined here, which uses an operator on measures (or L_1-functions) has been treated by several authors. See for example [2], [10], [13].

3. The filling scheme

Let μ and ν be two given nonnegative measures. We shall define a special balayage sequence (f) using μ and ν, called the filling scheme. The filling scheme was defined in [5] and has been studied in various contexts. (Cf. [1], [10], [13].)

μ will be the initial measure of f. ν will not necessarily be the terminal measure but will dominate the terminal measure. We think of μ as dirt, and ν as a hole, and try to fill the hole with the initial measure, using P as a "shovel". We stop as much mass as possible at each time. Define $f = (f_n, f_n')$ by

$$(3.1) \qquad f_0' = \mu \wedge \nu,$$

$$(3.2) \qquad f_{n+1}' = f_n P \wedge \left(\nu - \sum_{k=0}^{n} f_n'\right).$$

Clearly

$$f_n \wedge \left(\nu - \sum_{k=0}^{n} f'_k\right) = 0 \quad \text{for all } n. \tag{3.3}$$

We now prove two results which show the special nature of the filling scheme. We shall not use the fact that P is mass-preserving, only that P is a positive contraction.

(3.4) LEMMA. *Let f and r be balayage sequences with $f + f'_0 = r_0 + r'_0 = \mu$, and let $\psi = \sum_{k=0}^{\infty} r'_k$. Suppose that*

$$f_n \wedge \left(\psi - \sum_{k=0}^{n} f'_k\right) \leq 0 \quad \textit{for all } n. \tag{3.5}$$

Then

$$\sum_{k=0}^{n} f_k \leq \sum_{k=0}^{n} r_k \quad \textit{for all } n. \tag{3.6}$$

Proof. $f_0 \wedge (r'_0 - f'_0) \leq 0$, so $\mu \wedge r'_0 \leq f'_0$, or $r'_0 \leq f'_0$. Thus $r_0 \geq f_0$, so (3.6) holds for $n = 0$.

Suppose that (3.6) holds for some n. We shall prove it for $n + 1$.

Choose some reference measure with respect to which all measures involved are absolutely continuous. We shall denote a measure and its density by the same letter. All equations are to hold almost everywhere. (The use of densities seems to make the argument clearer, but could easily be avoided.

Consider a point x in S. There are two possible cases:

Case 1. $\psi(x) - \sum_{k=0}^{n+1} f'_k(x) > 0$. Then $f_k(x) = 0$ for $k = 0, \ldots, n + 1$. Hence (2.10) holds trivially at x.

Case 2. $\psi(x) - \sum_{k=0}^{n+1} f'_k(x) \leq 0$. Then by (2.3) and (2.10),

$$\begin{aligned} \sum_{k=0}^{n+1} f_k(x) &= \mu(x) + \sum_{k=0}^{n} f_k P(x) - \sum_{k=0}^{n+1} f'_k(x) \\ &\leq \mu(x) + \sum_{k=0}^{n} r_k P(x) - \psi(x) \\ &\leq \sum_{k=0}^{n+1} r_k(x). \end{aligned}$$

Thus (3.6) holds for $n + 1$, so the lemma is proved.

Remark. If we exchange primed and unprimed terms in the hypothesis of Lemma (3.4), then (3.6) holds with the inequality reversed. The proof is almost identical to that given above. Results of this sort are stated in [6], apparently with a different method of proof.

(3.7) LEMMA. *Let f and r be balayage sequences with* $f_0 + f_0' = r_0 + r_0' = \mu$, *such that* (3.6) *holds. Let* $a_n = \sum_{k=0}^{n} f_k'(S)$ *and let* $b_n = \sum_{k=0}^{n} r_k'(S)$. *Then*

$$\sum_{k=0}^{n} a_k \geq \sum_{k=0}^{n} b_k \quad \textit{for all } n. \tag{3.8}$$

Proof. We note that $\gamma(S') - \gamma P(S) \geq 0$ for any nonnegative measure γ. Hence if $\gamma_1 \geq \gamma_2$, then $\gamma_1(S) - \gamma_1 P(S) \geq \gamma_2(S) - \gamma_2 P(S)$.

$$\begin{aligned} a_n &= \mu(S) + \sum_{k=0}^{n-1} (f_k P(S) - f_k(S)) - f_n(S) \quad \text{by (2.3)}, \\ &\geq \mu(S) + \sum_{k=0}^{n-1} (r_k P(S) - r_k(S)) - f_n(S) \quad \text{by (3.6)}, \\ &= b_n - f_n(S) + r_n(S). \end{aligned}$$

Summing and applying (3.6) once more, the lemma is proved.

In order to apply these lemmas to the filling scheme, let μ and ν be given, and let f be the filling scheme for μ and ν. Let r be any other balayage sequence such that $\mu \to \nu(r)$. It follows from (3.3) and Lemma (3.4) that (3.6) holds. Hence by Lemma (3.7), (3.8) holds. Since $\mu \to \nu(r)$, $b_n \to \nu(S)$ as $n \to \infty$, in the notation of Lemma (3.7). Since $a_n \leq \nu(S)$ and (3.8) holds, it must also be true that $a_n \to \nu(S)$ as $n \to \infty$. Thus $\mu \to \nu(f)$. This result is stated in [10] as a consequence of the theorems of [13].

Equation (2.10) implies the further fact that $E[f] \leq E[r]$.

To summarize: If μ can be balayaged to ν in any way then μ can be balayaged to ν using the filling scheme. Furthermore the filling scheme stopping time has minimal expectation.

Various authors have studied the problem of determining when one measure can be balayaged to another, especially in the case of Brownian motion. (Cf. [14], [7], [13].)

4. The main result

We now restrict our attention to a class of "highly recurrent" processes. We shall assume that there exists an invariant probability measure λ. Let N denote the space of bounded signed measures γ on $\mathscr{B}$ with $\gamma(S) = 0$. We assume that for each γ in N, $\gamma(I + P + \cdots + P^n)$ converges in total variation norm as $n \to \infty$. We define the operator G on N by

$$\gamma G = \sum_{k=0}^{\infty} \gamma P^k. \tag{4.1}$$

By the uniform boundedness principle G is a bounded linear operator on N, with respect to total variation norm. We extend G to a bounded linear operator on the space of all bounded signed measures γ on $\mathscr{B}$ by setting

$$\gamma G \equiv (\gamma - \gamma(S)\lambda)G. \tag{4.2}$$

We define the "Laplacian", Δ, as usual:

$$\Delta = P - I. \tag{4.3}$$

Let Λ be defined by

$$\gamma\Lambda = \gamma(S)\lambda. \tag{4.4}$$

(4.5) LEMMA. $\Delta G = -I + \Lambda.$

The proof is immediate.

Operators like G, and more general operators, are studied systematically in [8] and [9]. G is also the discrete-time analogue of the Green operator defined in [3] for diffusions on compact manifolds.

A simple example of a situation where the assumptions of this section hold is that of an ergodic Markov chain with a finite number of states. Let P denote the transition matrix. Let Λ denote the matrix with every row equal to the invariant measure λ. The matrix of G is

$$G \equiv \sum_{k=0}^{\infty} (P^k - \Lambda) = (I - (P - \Lambda))^{-1} - \Lambda. \tag{4.6}$$

Another example, this time of a rather trivial nature, is helpful in understanding the theorem that follows. Let $S = [0, 1]$, $\mathscr{B}$ = the Borel sets, and let λ be the ordinary Lebesgue measure on $\mathscr{B}$. Define $P = \Lambda$. Then $G = I - \Lambda$.

We shall prove

(4.7) THEOREM. *Let μ and ν be probability measures on $\mathscr{B}$. Then $\mu \to \nu$ if and only if the negative part of $(\mu - \nu)G$ is absolutely continuous with respect to λ. If $\mu \to \nu$ then*

(4.8) $\min \{E[r] \mid \mu \to \nu(r)\} = \operatorname{ess\,sup} |\phi|$, *where ϕ is the density of the negative part of $(\mu - \nu)G$ with respect to λ.*

The proof of the theorem requires some preliminary lemmas.

(4.9) LEMMA. *For any bounded measure γ on $\mathscr{B}$, $\gamma P^n \to \gamma(S)\lambda$ as $n \to \infty$.*

The proof follows at once from the invariance of λ and the convergence of (4.1).

(4.10) LEMMA. *Let $\{r_n\}$ be a sequence of nonnegative, bounded measures, such that $r_nP \geq r_{n+1}$ for all n. Let $c = \lim_{n\to\infty} r_n(S)$. Then $r_n \to c\lambda$ as $n \to \infty$.*

Proof.

$$\begin{aligned} \|r_nP^m - g_{n+m}\| &\le \sum_{k=1}^{m} \|r_{n+k-1}P^{m-k+1} - r_{n+k}P^{m-k}\| \\ &\le \sum_{k=1}^{m} \|r_{n+k-1}P - r_{n+k}\| \\ &\le \sum_{k=1}^{m} (r_{n+k-1}(S) - r_{n+k}(S)) \\ &= r_n(S) - r_{n+m}(S). \end{aligned}$$

It follows easily from Lemma (4.9) that r_n is a Cauchy sequence, and hence that the lemma holds.

(4.11) LEMMA. *Let μ and v be bounded, mutually singular measures on $\mathscr{B}$ with $\mu(S) = v(S)$. Suppose that*

$$(4.12) \qquad (\mu - v)G + c\lambda \ge 0 \text{ on } \mathscr{B} \text{ for some } c, -\infty < c \le \infty.$$

Then

$$(4.13) \qquad (\mu - v)G - \mu + c\lambda \ge 0 \quad \text{on } \mathscr{B}.$$

Proof. Choose sets A and B in $\mathscr{B}$ with $A \cap B = \emptyset$, $A \cup B = S$, $v = 0$ on $A \cap \mathscr{B}$ and $\mu = 0$ on $B \cap \mathscr{B}$. Let $\gamma = (\mu - v)G - \mu + c\lambda$. Then clearly $\gamma \ge 0$ on $B \cap \mathscr{B}$, by (4.12). Since γ is "superharmonic" on $A \cap \mathscr{B}$, that is, since $\Delta\gamma \le 0$ on $A \cap \mathscr{B}$ we can finish the proof by proving a simple "domination principle" and applying it to γ. However, it is faster to proceed directly.

It also follows from (4.12) that

$$(4.14) \qquad (\mu - v)GP + c\lambda \ge 0 \quad \text{on } \mathscr{B}.$$

This is trivial if c is finite. If $c = \infty$, one shows first that the operator P preserves absolute continuity with respect to λ. Applying this fact to the negative part of $(\mu - v)G$ then gives the inequality.

By Lemma (4.5), we write (4.14) as

$$(\mu - v)G - \mu + v + c\lambda \ge 0 \quad \text{on } \mathscr{B}.$$

But then (4.13) holds on $A \cap \mathscr{B}$ so the lemma is proved.

COROLLARY. *Using Lemma* (4.5) *we see that* (4.13) *can also be stated as*

$$(4.15) \qquad (\mu P - v)G + (c - \mu(S))\lambda \ge 0 \quad \text{on } \mathscr{B}.$$

Proof of Theorem (4.7). Let μ and v be given. Let $f = (f_n, f_n')$ be an arbitrary balayage sequence with $f_0 + f_0' = \mu$ and $\sum_{k=0}^{\infty} f_k' \le v$. For example, f could be the filling scheme for μ and v.

By Lemma (4.9), f_n converges to a limit f_∞ which is a multiple of λ. Define

$$g_n = \nu - \sum_{k=0}^{n} f'_n \quad \text{for all } n. \tag{4.16}$$

The g_n decrease to a limit which is called g_∞. Clearly $\mu - \nu = f_0 - g_0$ and

$$f_{n+1} - g_{n+1} = f_n P - g_n \quad \text{for all } n. \tag{4.17}$$

Applying Lemma (4.5) to (4.17),

$$(f_n - g_n)G = (\mu - \nu)G - \sum_{k=0}^{n-1} f_k + \sum_{k=0}^{n-1} f_k(S)\lambda. \tag{4.18}$$

Hence

$$(f_\infty - g_\infty)G \leq (\mu - \nu)G + \sum_{k=0}^{\infty} f_k(S)\lambda. \tag{4.19}$$

(a) Suppose that $\mu \to \nu(f)$. Then $f_\infty - g_\infty = 0$, so

$$0 \leq (\mu - \nu)G + E[f]\lambda. \tag{4.20}$$

Hence the negative part of $(\mu - \nu)G$ must be of the form $\phi\, d\lambda$, with

$$\text{ess sup } |\phi| \leq E[f]. \tag{4.21}$$

(b) Conversely, let the negative part of $(\mu - \nu)G$ be of the form $\phi\, d\lambda$, and let $c = \text{ess sup } |\phi|$. Then

$$(\mu - \nu)G + c\lambda \geq 0 \quad \text{on } \mathscr{B}. \tag{4.22}$$

Let $f = (f_n, f'_n)$ now denote the filling scheme for μ and ν. By (3.3), $f_n \wedge g_n = 0$ for all n. By (4.17) and (4.15),

$$(f_n - g_n)G + \left(c - \sum_{k=0}^{n-1} f_k(S)\right)\lambda \geq 0 \quad \text{on } \mathscr{B} \tag{4.23}$$

for all n. Hence

$$(f_\infty - g_\infty)G + c\lambda \geq \sum_{k=0}^{\infty} f_k(S)\lambda \quad \text{on } \mathscr{B}. \tag{4.24}$$

Since f_∞ is a multiple of λ, it follows that $g_\infty G$ must be absolutely continuous with respect to λ. Since P preserves absolute continuity with respect to λ, g_∞ must be absolutely continuous with respect to λ, by Lemma (4.5). Hence $f_\infty = 0 = g_\infty$, since f_∞ and g_∞ are mutually singular. Thus $\mu \to \nu(f)$, and by (4.24).

$$c \geq E[f]. \tag{4.25}$$

By combining (4.21) and (4.25), the theorem is proved.

Some sufficient conditions for the assumptions of this section to hold can be found in [11]. It seems very likely that the assumptions of this section can be weakened, particularly as concerns the manner in which $\sum_{k=0}^{\infty} P^k$ converges on N.

In order to apply Theorem (4.7) more efficiently, we note that $(\mu - \nu)G$ should "assume its minimum" on the support of ν. More formally, the following lemma holds:

(4.26) LEMMA. *Let μ and ν be probability measures on $\mathscr{B}$. Let the negative part of $(\mu - \nu)G$ be of the form $\phi\, d\lambda$. Let B be any set in $\mathscr{B}$ with $\nu(B) = 1$ and $\lambda(B) > 0$. Then*

$$\text{ess sup } |\phi| \text{ on } B = \text{ess sup } |\phi| \text{ on } S.$$

Proof. Let $c = \text{ess sup } |\phi|$ on B. If $c = \infty$ the lemma is obvious, so assume $c < \infty$. Let $\gamma = (\mu - \nu)G + c\lambda$. Then $\gamma \geq 0$ on $B \cap \mathscr{B}$ and $\Delta\gamma \leq 0$ on $A \cap \mathscr{B}$, where $A = B^c$. Since γ is "superharmonic" on $A \cap \mathscr{B}$, it follows that $\gamma \geq 0$ on all of $\mathscr{B}$. Indeed, let $Q = \chi_B + P\chi_A$. Then $\gamma P \leq \gamma$ on $A \cap \mathscr{B}$ implies $\gamma Q \leq \gamma$ on $\mathscr{B}$, and hence $\gamma Q^n \leq \gamma$ for all n. Clearly $\gamma Q^n \geq \gamma(P\chi_A)^n$ for all n. Since $\lambda(B) > 0$, Lemma (4.10) implies that $\gamma(P\chi_A)^n \to 0$ as $n \to \infty$. Thus $\gamma \geq 0$ on $\mathscr{B}$, i.e., $|\phi| \leq c$ on $\mathscr{B}$, so the lemma is proved.

As an example, consider again the case of a Markov chain with a finite number of states. Let $\mu = \lambda$ and let $\nu = m_j$, a unit mass on the jth point of S. In this case the minimal stopping time τ such that $\mu \to \nu(\tau)$ is obviously unique. By the theorem,

$$E[\tau] = -\inf\{(\mu - \nu)G(\{i\})/\lambda(\{i\}) \mid i \in S\}.$$

By Lemma (4.26), $E[\tau] = -(\mu - \nu)G(\{j\})/\lambda(\{j\})$, or, since $\lambda G = 0$,

$$E[\tau] = m_j G(\{j\})/\lambda(\{j\}).$$

This agrees with the formula for $E[\tau]$ given in Proposition 9-79(1) of [9].

Finally, we will mention an alternative proof of part of Theorem (4.7), which was suggested by the referee. Let two probability measures μ and ν be given, and suppose we know that $\mu \to \nu$ via the filling scheme f. We wish to derive the formula for $E[f]$. Let $h = \sum_{k=0}^{\infty} f_k$. Then $h(S) = E[f]$. The following theorem is known: h is the *minimal nonnegative* solution of the equation $hP + \mu = h + \nu$.

This equation can be regarded as the Poisson equation for the potential h of the distribution $\mu - \nu$. Another solution of the same equation is of course the potential $(\mu - \nu)G$, which we will call g. One should note that g is by assumption a finite measure, whereas h will have infinite total mass if $E[f] = \infty$.

Since both h and g satisfy the Poisson equation, the difference $h - g$ is invariant under P. This invariance implies that $h - g = c\lambda$, for some c, $0 \leq c \leq \infty$. To see this, we note that the assumed existence of $(\mu - \lambda)G$ implies easily that h, and hence $h - g$, must have at most a finite singular part with respect to λ. Lemma (4.9) then shows that $h - g = c\lambda$.

We have $h = g + c\lambda \geq 0$, and this, together with the *minimality* of h, implies that c is the ess. sup. of the negative part of g. On the other hand $E[f] = h(S) = c$, so the formula is proved.

References

1. M. A. Akcoglu, *An ergodic lemma*, Proc. Amer. Math. Soc., vol. 16 (1965), pp. 388–392.
2. M. A. Akcoglu and R. W. Sharpe, *Ergodic theory and boundaries*, Trans. Amer. Math. Soc., vol. 132 (1968), pp. 447–460.
3. J. R. Baxter and G. A. Brosamler, *Energy and the law of the iterated logarithm*, Math. Scand., to appear.
4. J. R. Baxter and R. V. Chacon, *Potentials of stopped distributions*, Illinois J. Math., vol. 18, (1974), pp. 649–656.
5. R. V. Chacon and D. S. Ornstein, *A general ergodic theorem*, Illinois J. Math., vol. 4 (1960), pp. 153–160.
6. H. Dinges, "Stopping sequences" in *Séminaire de probabilitiés VIII*, Lecture notes in mathematics, no. 381, Springer, New York, 1974.
7. L. E. Dubins, *On a theorem of Skorohod*, Ann. Math. Stat., vol. 39 (1968), pp. 2094–2097.
8. J. G. Kemeny and J. L. Snell, *Finite Markov chains*, Van Nostrand, Princeton, 1960.
9. J. G. Kemeny, J. L. Snell and A. W. Knapp, *Denumerable Markov chains*, Van Nostrand, Princeton, 1966.
10. P. A. Meyer, "Travaux de H. Rost en théorie du balayage" in *Séminaire de probabilitiés V*, Lecture notes in mathematics, no. 191, Springer, New York, 1971.
11. S. Orey, *Potential kernels for recurrent Markov chains*, J. Math. Anal. Appl., vol. 8, (1964), pp. 104–132.
12. ———, *Limit theorems for Markov chain transition probabilities*, Van Nostrand Reinhold, London, 1971.
13. H. Rost, *The stopping distributions of a Markov process*, Inventiones Math., vol. 14, (1971), pp. 1–16.
14. A. B. Skorokhod, *Studies in the theory of random processes*, Addison-Wesley, Reading, Mass., 1965.

University of British Columbia
Vancouver, British Columbia