The Annals of Probability
1984, Vol. 12, No. 4, 1113–1148

CONDITIONAL MARKOV RENEWAL THEORY I. FINITE AND DENUMERABLE STATE SPACE

BY S. P. LALLEY

M.S.R.I./Berkeley and Columbia University

A renewal theory is developed for sums of independent random variables whose distributions are determined by the current state of a Markov chain (also known as "Markov additive" processes, or "semi-Markov" processes). This theory departs from existing theories in that its conclusions are required to be valid *conditionally* for a given realization of the Markov Chain. It rests on a peculiar coupling construction which differs markedly from existing coupling arguments.

1. Introduction. The story begins with a Markov chain $\{Y_n\}_{n\geq 1}$ taking values in a denumerable state space $\mathscr{Y}$. To each state $y \in \mathscr{Y}$ is assigned a probability distribution F_y on the real line, with (finite) mean μ_y. A new sequence $\{X_n\}_{n\geq 1}$ is generated by drawing X_n from F_{Y_n}, i.e.,

$$\mathscr{L}(X_{n+1} \mid \{Y_m\}_{m\geq 1}; X_1, X_2, \cdots, X_n) = F_{Y_{n+1}} \quad \forall n \geq 0; \tag{1.1}$$

the cumulative sum process $S_n = X_1 + \cdots + X_n$ ($S_0 = 0$) will be called a "Markov random walk" relative to the "driving process" $\{Y_m\}$.

When the variables X_n are strictly positive they may be interpreted as random sojourn times which may be pieced together to provide a random time change for the Markov chain $\{Y_n\}$. The new process

$$Z(t) = Y_{N(t)}, \quad \text{where} \quad N(t) = \max\{n: S_{n-1} < t\},$$

is called a "semi-Markov" process (cf. Smith, 1955, and Lévy, 1954). Notice that in this context the natural candidate for a "state-variable" is the current value of the Y-process, while in a very natural sense the X-variables are "subordinate": one is led to ask about the limiting behavior of $Z(t)$ for large t.

One may, however, just as well regard S_n as the "state variable," and ask about the behavior of S_n, or perhaps the pair (S_n, Y_n), for large values of n. In this context the Y-variables are, in a sense, "subordinate": they serve only as a string of instructions for the generation of the increments in the state-variable S_n. If one adopts this perspective one will, perhaps, find it natural to ask whether limit theorems which describe the behavior of (S_n, Y_n) for large n retain their validity for fixed individual strings $y_1, y_2, \cdots$ of instructions. It is to questions of this sort that this paper is devoted.

Markov renewal theory, as developed by Smith (1955), Orey (1959, 1961), Pyke (1961), Cinlar (1969), Jacod (1971), Kesten (1974), and many others, has

Received March 1983; revised December 1983.
AMS 1980 *subject classifications.* Primary 60K15; secondary 60K05
Key words and phrases. Markov renewal theory, coupling, conditional limit theorem.

traditionally been concerned with the study of the renewal measure

$$U_y(A, dz) = \sum_{n=1}^{\infty} P_y\{S_n \in dz,\ Y_n \in A\}$$

and the residual lifetime vector $(S_{\tau(a)} - a;\ Y_{\tau(a)})$, where

$$\tau(a) = \inf\{n: S_n > a\}.$$

The primary concerns of this paper are the analogous conditional quantities

$$U_y(A, dz \mid \mathscr{F}) =_{\Delta} \sum_{n=1}^{\infty} P_y(S_n \in dz,\ Y_n \in A \mid \mathscr{F})$$

and

$$\mathscr{L}_y((S_{\tau(a)} - a;\ Y_{\tau(a)}) \mid \mathscr{F}),$$

where

$$\mathscr{F} =_{\Delta} \sigma(Y_1, Y_2, \cdots).$$

Naturally it cannot be hoped that the conditional renewal measure is in general as well-behaved as the unconditional renewal measure: if all of the distributions F_y are degenerate, then the evolution of the process $\{S_n\}$ is completely determined by $\mathscr{F}$. We shall find, however, that if at least one of the distributions F_y is nondegenerate, then the conditional renewal measure exhibits some of the same limiting behavior as the unconditional renewal measure (cf. Theorem 1).

Periodicity phenomena may be considerably more complicated in Markov renewal theory than in the renewal theory for sums of i.i.d. random variables (cf. Címlar, 1974, for an extended discussion). Such phenomena are not our concern here, however, and we will sidestep them by imposing certain restrictions on the Markov random walks to be studied. Specifically, let

$$N = N(y) = \inf\{n \geq 1: Y_n = y\};$$

our assumption is that either

(1.1NA) $\quad \mathscr{L}(\sum_{n=1}^{N-1} X_n \mid Y_1 = y)$ *is nonarithmetic for all* $y \in \mathscr{Y}$,

or

(1.1A) $\quad \mathscr{L}(\sum_{n=1}^{N-1} X_n \mid Y_1 = y')$ *is supported by* $\mathbb{Z}$, *but by no smaller subgroup of* $\mathbb{R}$, *for all* $y, y' \in \mathscr{Y}$.

The leading result of Markov renewal theory is the

MARKOV RENEWAL THEOREM. *Suppose condition* (1.1) *is satisfied, and that the driving process* $\{Y_n\}$ *is aperiodic, irreducible, and recurrent, with invariant measure* $\pi(\cdot)$. *Suppose also that*

$$\sum_{y \in \mathscr{Y}} \pi(y) \int_{\mathbb{R}} |x| F_y(dx) < \infty \tag{1.2}$$

and

$$0 < \mu =_{\Delta} \sum_{y \in \mathscr{Y}} \pi(y)\mu_y < \infty. \tag{1.3}$$

Then for every $y \in \mathscr{Y}$, *every* $A \subset \mathscr{Y}$ *and every* $h > 0$ ($h \in \mathbb{Z}^+$ *under* (1.1A)),

$$\lim_{a\to\infty} U_y(A, [a, a + h)) = h\pi(A)\mu^{-1}. \tag{1.4}$$

Furthermore, there exists a probability distribution G *on* $\mathbb{R}^+ \times \mathscr{Y}$ (*on* $\mathbb{Z}^+ \times \mathscr{Y}$ *under* (1.1A)) *such that for every* $y \in \mathscr{Y}$, *every* $A \subset \mathscr{Y}$, *and every* $h > 0$ ($h \in \mathbb{Z}^+$ *under* (1.1A)),

$$\lim_{a\to\infty} P_y\{S_{\tau(a)} - a \geq h;\ Y_{\tau(a)} \in A\} = G([h, \infty) \times A). \tag{1.5}$$

NOTE. It should be understood that in the arithmetic case the limits in (1.4) and (1.5) are taken as $a \to \infty$ *through* $\mathbb{Z}$. When the increments X_n are always nonnegative, the limit distribution G is given by

$$G(dh \times A) = \sum_{y\in A} \pi(y) \int_{\{x>h\}} F_y(dx) \cdot dh \quad (h > 0)$$

in the nonarithmetic case, and by

$$G(\{x\} \times A) = \sum_{y\in A} \pi(y) \sum_{x'\geq x} F_y(\{x'\}) \quad (x \in \mathbb{Z}^+)$$

in the arithmetic case (cf. Proposition 2.3). In the general case the limit distribution G is not so easily described.

In the case of nonnegative increments X_n (the "semi-Markov" case) the Markov renewal theorem was first proved by Smith (1955) (Smith's assumptions and conclusions differ somewhat from those given above, however). Smith noticed that the "excursions" between successive visits to a fixed state are i.i.d., and thus reduced the problem to an application of Blackwell's renewal theorem for i.i.d. variates. Much of the subsequent work in Markov renewal theory has been concerned with Markov random walks for which the driving process $\{Y_n\}$ takes its values in an uncountable state space: in such situations there are no obvious sequences of regeneration points, hence Smith's approach fails (cf., however, Athreya, McDonald, and Ney, 1978b). The most general results available at present seem to be those of Kesten (1974). The Markov renewal theorem as stated above is a (very) special case of Kesten's Theorem 3 in the nonarithmetic case, when the invariant measure $\pi(\cdot)$ is finite.

Our first result is a "weak" conditional version of the Markov renewal theorem.

THEOREM 1. *If the hypotheses of the Markov renewal theorem are satisfied, then for every* $y \in \mathscr{Y}$, $A \subset \mathscr{Y}$, *and* $h > 0$ ($h \in \mathbb{Z}^+$ *in the arithmetic case*)

$$U_y(A, [a, a + h) \mid \mathscr{F}) \to_{P_y} h\pi(A)\mu^{-1} \tag{1.6}$$

and

$$P_y(S_{\tau(a)} - a \geq h;\ Y_{\tau(a)} \in A \mid \mathscr{F}) \to_{P_y} G([h, \infty) \times A), \tag{1.7}$$

(*as* $a \to \infty$ *through* $\mathbb{Z}$ *in the arithmetic case*), *provided at least one of the distributions* $\{F_y\colon y \in \mathscr{Y}\}$ *is nondegenerate.*

This theorem is evidently nontrivial even in the case of a finite state space.

Unlike the Markov Renewal Theorem, it cannot be proved by isolating an embedded renewal process and appealing to Blackwell's theorem. Nor does it seem to be amenable to attack by any of the traditional approaches to the renewal theorem. The proof given here (Section 3) is unlike any published proof of the renewal theorem known to the author; it ties together the three most well-known "first-moment" theorems of random walk theory (to wit, the renewal theorem, the law of large numbers, and the Chung–Fuchs–Ornstein theorem) in an unusual way.

Theorem 1 immediately prompts one to ask whether the convergence in (1.6) and (1.7) holds almost surely. The answer: sometimes. Counterexamples are given in Section 4; although these are simple, they illustrate that almost sure convergence depends on a rather delicate balance between the "regularity" (or lack of it) in the assignment $y \to F_y$ and the rate of mixing in the driving process $\{Y_n\}$. We have no definitive solution to the problem of determining the "right" conditions which guarantee almost sure convergence in (1.6)–(1.7). However, we have discovered some "reasonable" sets of sufficient conditions.

THEOREM 2. *If the state space* $\mathscr{Y}$ *is finite, and if each of the distributions* F_y *is supported by* $(0, \infty)$ (*thus* $X_n > 0$ *w.p.*1), *then under the hypotheses of Theorem* 1 *the convergence indicated in* (1.6) *and* (1.7) *holds almost surely.*

The hypothesis that $F_y((0, \infty)) = 1$ for each $y \in \mathscr{Y}$ is extraneous (the weaker hypothesis that $\mu_y > 0$ for each $y \in \mathscr{Y}$ is sufficient, and probably even this is unnecessary). However, it simplifies the proof considerably.

When the state space is infinite, the situation is complicated enormously. Some degree of mixing in the driving chain $\{Y_n\}$ seems to be crucial, but how much is necessary seems to depend on the amount of variability in the tail behavior in the collection $\{F_y\colon y \in \mathscr{Y}\}$. When all of the distributions F_y have finite second moments, relatively weak mixing conditions are sufficient. Let

$$\sigma_y^2 = \operatorname{var}(F_y) = \int_{\mathbb{R}} (x - \mu_y)^2 F_y(dx).$$

THEOREM 3. *Suppose that each of the distributions* F_y *is supported by* $(0, \infty)$. *If for all* $y \in \mathscr{Y}$

$$E_y N(y)^2 < \infty, \tag{1.8}$$

$$E_y\left(\textstyle\sum_{j=1}^{N(y)-1} \mu_{Y_j}\right)^2 < \infty, \tag{1.9}$$

and

$$E_y \textstyle\sum_{j=1}^{N(y)-1} \sigma_{Y_j}^2 < \infty, \tag{1.10}$$

then under the hypotheses of Theorem 1 *the convergence in* (1.6) *and* (1.7) *holds almost surely.*

The astute reader will notice that the conditions (1.8)–(1.10) are appropriate sufficient conditions for

$$(S_n - n\mu)n^{-1/2} \Rightarrow \text{normal};$$

this plays no role in the proof, however. The important "second moment" tool is the Hsu–Robbins (1948) theorem, which provides a rate of convergence in the law of large numbers. It is somewhat curious that such second moment conditions would arise in connection with a renewal theorem, and the reader may feel that they are unnecessarily strong. However, it is possible to construct (i) a driving chain $\{Y_n\}$ and associated distributions $\{F_y\}$ such that the variances σ_y^2 are uniformly bounded, $E_y\exp\{\theta N(y)\} < \infty$ for some $\theta > 0$, $E_y(\sum_{j=1}^{N(y)-1} \mu_{Y_j})^{2-\varepsilon} < \infty$ for any $\varepsilon > 0$, for which almost sure convergence in (1.6)–(1.7) fails (cf. Example 1, Section 4); and (ii) a driving chain $\{Y_n\}$ and associated distributions F_y all supported by the interval [1, 3], such that $E_yN(y)^{2-\varepsilon} < \infty$ for all $\varepsilon > 0$, and such that a.s. convergence in (1.6) and (1.7) fails (cf. Example 2, Section 4). Thus when the variances are bounded, conditions (1.8) and (1.9) are apparently needed.

The assumption (1.10) is by no means necessary for almost sure convergence to obtain in Theorem 1. However, certain restrictions on the tail behavior of the distributions in the family $\{F_y\}$ seem to be necessary; the author's best efforts to articulate appropriate restrictions have led to various unwieldy L^1-quasicompactness hypotheses, which will not be described in this paper. There is a special case which deserves mention, though, to wit, that where all of the distributions F_y are members of the same translation family.

THEOREM 4. *Suppose there exist constants $\theta_y \geq 0$ and a probability distribution $F(dx)$ on $(0, \infty)$ such that for each $y \in \mathscr{Y}$,*

$$F_y(dx) = F(d(x - \theta_y)). \tag{1.11}$$

Suppose also that for all $y \in \mathscr{Y}$,

$$E_yN(y)^2 < \infty \tag{1.12}$$

and

$$E_y[\textstyle\sum_{j=1}^{N(y)-1} \theta_{Y_j}]^2 < \infty. \tag{1.13}$$

Then under the hypotheses of Theorem 1, *the convergence* (1.6) *and* (1.7) *holds almost surely.*

Once again Example 1 of Section 4 demonstrates that, within the context of assumption (1.11), (1.13) is the "right" moment condition.

The problems considered in Theorems 1–4 are all subsumed by the more general problem of developing a renewal theory for sums of independent but nonidentically distributed random variables. The best renewal theoretic results for such processes seem to be those of Smith (1961), obtained by Fourier-analytic methods, and McDonald (1978), who used ergodic theory for Markov chains. McDonalds's results rely on a mixing condition which is somewhat obscure and evidently stronger than the hypotheses of Theorems 2–4. Smith's theorems do not overlap with those of this paper: his hypotheses (cf. Theorem 6, especially) require a certain uniformity in the sequence of distributions which will *never* be present in the sequence $F_{Y_1}, F_{Y_2}, \cdots$ if at least two of the means μ_y $(y \in \mathscr{Y})$ are distinct. Despite the lack of overlap, Smith's results and techniques do shed a certain amount of light on those of this paper (at the very least they should impress on the reader the difficulties of the Fourier-analytic approach, and the complexities of the periodicity problem).

Many of the results of this paper have extensions and analogues for Markov random walks whose driving processes $\{Y_n\}$ take values in nondenumerable spaces. These are considerably more difficult to establish, however, and so we defer all discussion of the uncountable case to a subsequent paper.

The main architectural features of the paper are as follows. Section 2 is a collection of ergodic theorems which will be used in the proof of Theorem 1: these are obtained by familiar arguments which are of no great interest in their own right. The reader may wish to merely note the statements of the results, and move on to Section 3, which contains the proof of Theorem 1. This proof is based on a coupling construction which is the essential "technical" feature of the paper; the reader may find it more interesting than the theorems for which it was designed. The proofs of Theorems 2 and 3 are given in Sections 5–7; they are increasingly subtle variations on the main coupling argument presented in Section 3. The counterexamples alluded to earlier are collected in Section 4.

In many of the proofs the differences between the arithmetic and nonarithmetic cases are small enough that we have omitted the details of the nonarithmetic case. In proving Theorems 2 and 3 we only consider the convergence (1.6), since the arguments needed to establish (1.7) are virtually identical. We assume that the reader is comfortable with Markov chains on denumerable state space, especially in Section 2.

2. Ergodic theorems for Markov random walks. The approach to renewal theory developed in Section 3 relies on certain variants of the ergodic theorem: the purpose of this section is to catalogue these results. Throughout this section the notation and terminology established in Section 1 will be retained, and *the assumptions of the Markov renewal theorem will be in force.*

Let y denote the initial state of the Markov chain $\{Y_n\}_{n\geq 1}$ (i.e., $P_y\{Y_1 = y\} = 1$), and let $N_0 = 1, N_1, N_2, \cdots$ be the instants of successive visits to y:

$$(2.1) \qquad N_0 = 1, \quad N_{j+1} = \min\{n > N_j \colon Y_n = y\}, \quad j \geq 0.$$

Since $\{Y_n\}$ is recurrent, each of the random variables N_j is finite and well-defined. Furthermore, the successive "excursions" from y are i.i.d., by the Markov property: in particular, if $f\colon \mathscr{Y} \times \mathbb{R} \to \mathbb{R}$ is any function, then

$$(2.2) \qquad \textstyle\sum_{j=1+N_i}^{N_{i+1}} f(Y_j, X_j), \quad i = 0, 1, 2, \cdots \quad \text{are i.i.d.}$$

under P_y. Important special cases are

$$f(y', x) = x, \quad f(y', x) = \mu_{y'}, \quad \text{and} \quad f(y', x) = \sigma_{y'}^2.$$

Since the Markov chain $\{Y_n\}_{n\geq 1}$ is aperiodic, irreducible, and recurrent, there is an invariant measure $\pi(\cdot)$ which is unique (up to constant multiples). It is well-known (cf. Derman, 1954) that an invariant measure $\lambda(\cdot)$ is specified by

$$\lambda(A) = E_y \sum_{n=2}^{N_1} 1\{Y_n \in A\}, \quad A \subset \mathscr{Y}; \tag{2.3}$$

by the essential uniqueness, it follows that for any invariant measure $\pi(\cdot)$,

$$\lambda(A) = \pi(A)/\pi(y), \quad A \subset \mathscr{Y}. \tag{2.4}$$

Consequently, under the hypotheses of Theorem 1

$$E_y \sum_{n=2}^{N_1} \mu_{Y_n} = \mu/\pi(y), \tag{2.5}$$

$$E_y \sum_{n=2}^{N_1} |X_n| = E_y \sum_{n=2}^{N_1} \int_{\mathbb{R}} |x| F_{Y_n}(dx) < \infty, \tag{2.6}$$

and

$$E_y \sum_{n=2}^{N_1} X_n = E_y \sum_{n=2}^{N_1} \mu_{Y_n} = \mu/\pi(y). \tag{2.7}$$

ERGODIC THEOREM. *If f: $\mathscr{Y} \times \mathbb{R} \to \mathbb{R}$ is any function for which*

$$\sum_{y' \in \mathscr{Y}} \pi(y') \int_{x \in \mathbb{R}} |f(y', x)| F_{y'}(dx) < \infty, \tag{2.8}$$

then for every $A \subset \mathscr{Y}$ with $0 < \pi(A) < \infty$,

$$\frac{\sum_{j=1}^{n} f(Y_j, X_j)}{\sum_{j=1}^{n} 1\{Y_j \in A\}} \to \frac{\sum_{y' \in \mathscr{Y}} \pi(y') \int_{x \in \mathbb{R}} f(y', x) F_{y'}(dx)}{\pi(A)} \tag{2.9}$$

with P_y-probability one, for every $y \in \mathscr{Y}$. If $\pi(\mathscr{Y}) = 1$, then

$$n^{-1} \sum_{j=1}^{n} f(Y_j, X_j) \to \sum_{y' \in \mathscr{Y}} \pi(y') \int_{x \in \mathbb{R}} f(y', x) F_{y'}(dx) \tag{2.10}$$

with P_y-probability one, for every $y \in \mathscr{Y}$.

The result is well known: in the special case where f depends only on y', it may be found in Section I.15 of Chung (1967). The proof given by Chung may easily be adapted to the general case.

PROPOSITION 2.1. *For each initial state $y \in \mathscr{Y}$, each subset $A \subset \mathscr{Y}$ with $0 < \pi(A) < \infty$, and each $\varepsilon > 0$, there exists an integer $M = M(y, A, \varepsilon)$ so large that whenever $m \geq M$ and $a \geq M$,*

$$P_y\{|m^{-1} \sum_{n \geq 1} 1\{a - m \leq S_n \leq a;\ Y_n \in A\} - \pi(A)\mu^{-1}| > \varepsilon\} < \varepsilon. \tag{2.11}$$

PROOF. Consider first the special case where $\pi(\mathscr{Y}) = 1$ and $A = \mathscr{Y}$. In this case the Ergodic Theorem implies that $S_n/n \to \mu$ almost surely P_y, for every $y \in \mathscr{Y}$. Consequently for each $y \in \mathscr{Y}$ there exists $M_1 = M_1(y)$ large enough that whenever $m \geq M_1(y)$,

$$P_y\{|m^{-1} \sum_{n \geq 1} 1\{0 \leq S_n \leq m\} - \mu^{-1}| > \varepsilon/2\} < \varepsilon/2. \tag{2.12}$$

Now we appeal to the Markov renewal theorem. This asserts that there is a probability distribution G on $(0, \infty) \times \mathscr{Y}$ such that the pair $(S_{\tau(a)} - a, Y_{\tau(a)})$ converges in law to G as $a \to \infty$ ($a \to \infty$ through $\mathbb{Z}$ in the arithmetic case) under each P_y, $y \in \mathscr{Y}$. Therefore for each $y \in \mathscr{Y}$ there exist constants $K(y) < \infty$, $M_2(y) < \infty$ such that for all $a > 0$,

$$P_y\{S_{\tau(a)} - a \geq K(y) \text{ or } M_1(Y_{\tau(a)}) \geq M_2(y)\} < \varepsilon/2. \tag{2.13}$$

If $M(y) \geq M_2(y) + K(y)$ is chosen large enough that

$$\left| \left[\frac{M(y)}{M(y) - K}\right]\left(\mu^{-1} + \frac{\varepsilon}{2}\right) - \mu^{-1} \right| < \varepsilon$$

and

$$\left| \left[\frac{M(y)}{M(y) - K}\right]\left(\mu^{-1} - \frac{\varepsilon}{2}\right) - \mu^{-1} \right| < \varepsilon,$$

then (2.12) and (2.13), together with the Markov property of (Y_n, S_n), imply

$$P_y\{| m^{-1} \textstyle\sum_{n\geq 1} 1\{a - m \leq S_n \leq m\} - \mu^{-1} | > \varepsilon\} < \varepsilon \tag{2.14}$$

for all $a \geq m \geq M(y)$. This proves (2.11) in the special case $A = \mathscr{Y}$, $\pi(\mathscr{Y}) = 1$.

The general case of (2.11) will be deduced from the special case by the device of isolating an appropriate embedded Markov random walk. Without loss of generality assume that $\pi(A) = 1$ (the invariant measure can always be renormalized) and let $T(0), T(1), \cdots$ be the times of successive visits to A by $\{Y_n\}$, i.e.,

$$T(0) = \min\{n \geq 1\colon Y_n \in A\}$$

$$T(j + 1) = \min\{n > T(j)\colon Y_n \in A\}.$$

Set

$$\tilde{Y}_{k+1} = (Y_{T(k)+1}, Y_{T(k)+2}, \cdots, Y_{T(k+1)}), \quad k \geq 0$$

$$\tilde{S}_k = S_{T(k)}, \qquad k \geq 1$$

$$\tilde{S}_0 = 0 \tag{2.15}$$

and

$$\tilde{X}_k = \tilde{S}_k - \tilde{S}_{k-1}.$$

It is easily verfied that $\{\tilde{Y}_k\}_{k\geq 1}$ is an aperiodic, irreducible, recurrent Markov chain, and it follows from (1.1) that

$$\mathscr{L}(\tilde{X}_{k+1} | \tilde{X}_1, \cdots, \tilde{X}_k; \{\tilde{Y}_j\}_{j\geq 1}) = F_{Y_{T(k)+1}} * F_{Y_{T(k)+2}} * \cdots * F_{Y_{T(k+1)}} =_\Delta \tilde{F}_{\tilde{Y}_{k+1}}$$

so $\{\tilde{S}_k\}$ is a Markov random walk with respect to the driving process $\{\tilde{Y}_k\}$. Moreover, it follows from the ergodic theorem (applied to the original process (S_n, Y_n)) that $\tilde{S}_k/k \to \mu$ almost surely. Finally, $\{\tilde{Y}_k\}$ is positive recurrent: its stationary distribution

$$\tilde{\pi}(y_1, \cdots, y_r) = \textstyle\sum_{y_0 \in A} \pi(y_0) \prod_{i=0}^{r-1} p(y_i, y_{i+1})$$

(here $(y_1, y_2, \cdots, y_r)$ is a finite sequence from $\mathscr{Y}$ with $y_r \in A$) has total mass 1.

Consequently the result (2.14) may be applied to the derived process $(\tilde{S}_k, \tilde{Y}_k)$: in particular, for each possible initial state $\tilde{y}$ of $\{\tilde{Y}_k\}$, there exists a constant $M(\tilde{y})$ such that $a \geq m \geq M(\tilde{y})$ implies

$$P_{\tilde{y}}\{|m^{-1}\textstyle\sum_{k\geq 1} 1\{a - m \leq \tilde{S}_k \leq a\} - \mu^{-1}| > \varepsilon\} < \varepsilon. \tag{2.16}$$

Since

$$\textstyle\sum_{k\geq 1} 1\{a - m \leq \tilde{S}_k \leq a\} = \sum_{n\geq 1} 1\{a - m \leq S_n \leq a;\ Y_n \in A\},$$

(2.11) follows from (2.16) by an easy argument. □

COROLLARY 2.2. *For each initial state $y \in \mathscr{Y}$, each subset $A \subset \mathscr{Y}$ with $0 < \pi(A) < \infty$, and each $\varepsilon > 0$, there exists an integer $M = M(y, A, \varepsilon)$ large enough that whenever $a \geq m \geq M$,*

$$P_y\{|E_y(m^{-1}\textstyle\sum_{n\geq 1} 1\{a - m \leq S_n \leq a;\ Y_n \in A\} \mid \mathscr{F}) - \mu^{-1}\pi(A)| > \varepsilon\} < \varepsilon. \tag{2.17}$$

PROOF. Suppose Z is a random variable on a probability space $(\Omega, \mathscr{B}, P)$ and $c, \delta \in \mathbb{R}$ are constants for which

$$P(|Z - c| \geq \delta) < \delta \quad \text{and} \quad |EZ - c| < \delta.$$

Then an easy argument based on the triangle inequality shows that

$$E|Z - c| \leq 3\delta.$$

Consequently, if $\mathscr{A}$ is any sub-σ-albegra of $\mathscr{B}$,

$$P(|E(Z \mid \mathscr{A}) - c| > \varepsilon) \leq 3\delta/\varepsilon, \tag{2.18}$$

by Markov's inequality, for all $\varepsilon > 0$.

Returning to (2.17), recall from Proposition 2.1 that $\exists M_1 = M_1(y, A, \varepsilon)$ large enough that $a \geq m \geq M_1$ implies

$$P_y\{|m^{-1}\textstyle\sum_{n\geq 1} 1\{a - m \leq S_n \leq a;\ Y_n \in A\} - \mu^{-1}\pi(A)| > \delta\} < \delta.$$

Moreover, the Markov renewal theorem implies that $\exists M_2 = M_2(y, A)$ large enough that $a \geq m \geq M_2$ implies

$$|E_y(m^{-1}\textstyle\sum_{n\geq 1} 1\{a - m \leq S_n \leq a;\ Y_n \in A\}) - \mu^{-1}\pi(A)| < \delta.$$

Thus (2.18) implies that if $a \geq m \geq \max(M_1, M_2)$, then LHS (2.17) $\leq 3\delta/\varepsilon$. By choosing $\delta < \varepsilon^2/3$ we obtain (2.17). □

PROPOSITION 2.3. *Suppose* (1.1A) *holds* (*i.e., the Markov random walk is arithmetic*). *For each initial state $y \in \mathscr{Y}$, each positive integer x, each subset $A \subset \mathscr{Y}$, and each $\varepsilon > 0$, there exists a constant $M = M(y, A, x, \varepsilon)$ sufficiently large that whenever $a \geq m \geq M$*

$$P_y\{|m^{-1}\textstyle\sum_{b=a=-m}^{m} 1\{S_{\tau(b)} - b = x;\ Y_{\tau(b)} \in A\} - G(\{x\}; A)| > \varepsilon\} < \varepsilon \tag{2.19}$$

and

$$P_y\{|E_y(m^{-1}\sum_{b=a-m}^{m}1\{S_{\tau(b)}-b=x;\ Y_{\tau(b)}\in A\}\mid\mathscr{F}) - G(\{x\};A)|>\varepsilon\}<\varepsilon. \tag{2.20}$$

Recall that $G(\cdot,\cdot)$ is the limiting distribution given by the Markov renewal theorem (cf. (1.5)).

PROOF. (2.20) follows from (2.19) by the same argument used to prove Corollary 2.2; we omit the details.

Define a new process

$$Z_b=(S_{\tau(b)}-b;\ Y_{\tau(b)});$$

it is clear that $\{Z_b\}_{b=0,1,\dots}$ is Markovian, and takes values in the countable set $\{1,2,\cdots\}\times\mathscr{Y}$. Furthermore the Markov renewal theorem (specifically, (1.5)) implies that $G(\cdot,\cdot)$ is the unique invariant measure for $\{Z_b\}$. Consequently, (2.19) follows by an easy argument from the ergodic theorem (specifically, from (2.10)). We omit the details. (The usefulness of the "first-passage chain" $\{Z_b\}$ in renewal theory is well-known; cf. McDonald (1974) for an extended discussion). □

There are analogous results for the nonarithmetic case. Since (2.19) and (2.20) will only be used in the arithmetic case (cf. Subsection 3A, in the proof of (1.7)), we refrain from stating the nonarithmetic analogues.

For dealing with Markov random walks in the "lattice-nonarithmetic" case (cf. Subsection 3B to follow) certain extensions of the results of Propositions 2.1 and 2.3 and Corollary 2.2 will be needed. These are collected in

PROPOSITION 2.4. *Suppose* (1.1A) *holds, and let $\nu\ge 2$ and $0\le\eta<\nu$ be fixed integers. Then for each initial state $y\in\mathscr{Y}$, each subset $A\subset\mathscr{Y}$ with $0<\pi(A)<\infty$, and each $\varepsilon>0$, there exists an integer $M=M(y,A,\varepsilon,\nu)$ large enough that whenever $a\ge m\ge M$,*

$$P_y\{|E(m^{-1}\sum_{n\ge1}1\{a-m\le S_n\le a;\ Y_n\in A;\ S_n\equiv\eta(\mathrm{mod}\ \nu)\}\mid\mathscr{F}) - (\mu\nu)^{-1}\pi(A)|>\varepsilon\}<\varepsilon. \tag{2.21}$$

Furthermore, for each integer $x\ge0$ there exists $M=M(y,x,A,\varepsilon,\nu)$ such that whenever $a\ge m\ge M$,

$$P_y\{|E(m^{-1}\sum_{a-m\le b\le a;\, b-a\equiv0\,\mathrm{mod}\,\nu}1\{S_{\tau(b)}-b=x;\ Y_{\tau(b)}\in A\}\mid\mathscr{F}) - \nu^{-1}G(\{x\};A)|>\varepsilon\}<\varepsilon. \tag{2.22}$$

This can be proved by arguments very similar to those used in Propositions 2.1 and 2.3 and Corollary 2.4.

3. Multiple coupling: proof of Theorem 1. The technique of "coupling" random processes to prove limit theorems usually proceeds according to the following simple plan. Two processes, usually independent replicas of the same process, but with different initial conditions, are defined on the same probability space. One is in a "steady state" (i.e., its initial distribution is an invariant measure), the other is not. Eventually the processes meet (the coupling time) and coalesce forever after: thus the process which did not begin in steady state nevertheless approaches it.

(The coupling technique as described above was apparently first discovered by Doeblin (1939), who used it to establish various limit theorems for Markov chains on finite state spaces. It was resurrected by Ornstein (1968) in a different context: Ornstein used it to deduce renewal-theoretic results for random walks on $\mathbb{R}$. Simple proofs of the renewal theorem based on the coupling technique may be found in Lindvall (1977) and Athreya, McDonald, and Ney (1978); a more subtle approach to renewal theory based on coupling, with correspondingly greater rewards, is expounded by Ney (1981).)

The proofs of Theorem 1–4 which follow all make use of a coupling construction; this coupling, however, does not weld together two processes, but instead *infinitely many*, each with a different initial condition. The "approach to steady state" comes about by averaging over the ensemble of initial conditions. Apart from the purely technical advantages of this approach, there is an àesthetic one as well, in that the renewal theorem is unmasked as little more than the law of large numbers turned upside down.

There is a certain amount of epsilonics involved even in the proof of Theorem 1, which, as the reader will discover, is still the simplest of the lot. To highlight the essential features of the coupling construction, we will first prove Theorem 1 in the arithmetic case under more restrictive hypotheses, and then indicate the modifications which are needed to remove the restrictions. Finally, we will describe in somewhat less detail the form the coupling construction takes in the nonarithmetic case.

3A. The simplest case. Let $y \in \mathscr{Y}$ be a fixed but typical state in $\mathscr{Y}$ to serve as the initial state of the Markov chain $\{Y_n\}_{n\geq 1}$. Call the times of successive returns to y $N_0, N_1, \cdots$: thus

$$N_0 = 1, \quad N_{k+1} = \min\{n > N_k \colon Y_n = y\}. \tag{3.1}$$

Assume that the underlying probability space is large enough to accommodate two independent copies $\{X^A_{n+1}\}_{n\geq 1}$ and $\{X^B_n\}_{n\geq 1}$ of the (increments in) the Markov random walk, i.e.,

$$\begin{aligned} &P(X^A_{n+1} \in dx^A;\, X^B_{n+1} \in dx^B \mid \mathscr{F};\, X^A_1, \cdots, X^A_n;\, X^B_1, \cdots, X^B_n) \\ &\qquad\qquad = F_{Y_{n+1}}(dx^A)F_{Y_{n+1}}(dx^B) \quad \forall n \geq 0. \end{aligned} \tag{3.2}$$

ASSUMPTIONS 3A.

$$X^A_n \geq 1 \quad \text{and} \quad X^B_n \geq 1 \qquad \forall n = 1, 2, \cdots. \tag{3A.1}$$

(3A.2) Under P_y, the random variables $\sum_{n=1}^{N_1-1} X^A_n$ and $\sum_{n=1}^{N_1-1} (X^A_n - X^B_n)$ have distributions which are supported by $\mathbb{Z}$, but by no proper subgroup of $\mathbb{Z}$.

Assumption (3A.2) is stronger than the arithmetic condition (1.1A). The reason for including it is that the coupling we are about to decribe may fail (the coupling times may be infinite) if (3A.2) does not hold. (I am indebted to Professor Jim Pitman for pointing this out).

Processes $\{S_n^\nu\}_{n\geq 0}$ and coupling times $\sigma(\nu)$ $(\nu = 0, 1, 2, \cdots)$ may now be defined as follows:

$$(3.3)\qquad \begin{aligned} S_0^\nu &= \nu \\ S_{n+1}^0 &= S_n^0 + X_{n+1}^A \\ S_{n+1}^\nu &= \begin{cases} S_n^\nu + X_{n+1}^A & \text{if } \sigma(\nu) \leq n \\ S_n^\nu + X_{n+1}^B & \text{if } \sigma(\nu) > n \end{cases} \\ \sigma(0) &= 0 \\ \sigma(\nu+1) &= \min\{n > \sigma(\nu): S_n^0 = S_n^{\nu+1}\} \\ &= \min\{n > \sigma(\nu): \textstyle\sum_{j=1}^n (X_n^A - X_n^B) = \nu + 1\}. \end{aligned}$$

LEMMA 3.1. *If Assumption* (3A.2) *holds, then for each integer* $\nu > 0$,

$$P_y\{\sigma(\nu) < \infty\} = 1.$$

PROOF. The process $U_k = \sum_{n=1}^{N_k-1} (X_n^A - X_n^B)$ $(k = 0, 1, 2, \cdots)$ is an ordinarily random walk on $\mathbb{Z}$ with mean zero (this because of (3.2) and the Markov property of $\{Y_n\}$; also (1.2)). Consequently, by (3.2A) and the well-known recurrence theorem of Chung, Fuchs, and Ornstein (cf. Proposition 1.2.8 of Spitzer, 1975, or Chung and Ornstein, 1962), $\{U_k\}_{k\geq 0}$ visits each point of $\mathbb{Z}$ infinitely often. The fact that $\{\sigma(\nu) < \infty\}$ follows from this and the definition (3.3). □

The coupling times $\sigma(\nu)$ are obviously stopping times with respect to the filtration

$$\mathscr{G}_n = \sigma((Y_j, X_j^A, X_j^B))_{j=1,2,\cdots,n}.$$

It follows easily from this that each of the processes

$$\{S_n^\nu - \nu\}_{n\geq 1}, \quad \nu = 0, 1, \cdots,$$

has the same marginal distribution (conditional on $\mathscr{F}$). This is one of two obvious but crucial features of the construction (3.3); the second is that the processes $\{S_n^0\}_{n\geq 1}$ and $\{S_n^\nu\}_{n\geq 1}$ "coalesce" at time $\sigma(\nu)$, i.e.,

$$(3.4)\qquad S_n^\nu = S_n^0 \quad \forall n \geq \sigma(\nu),$$

and thus

$$S_n^\nu = S_n^{\nu-1} = \cdots = S_n^0 \quad \forall n \geq \sigma(\nu).$$

PROOF OF (1.6) UNDER ASSUMPTIONS (3A). The fact that $\sigma(\nu) < \infty$ with P_y-probability one guarantees that for large $a \in \mathbb{Z}$, the processes $\{S_n^\nu\}_{n\geq 1}$, $\nu = 0, 1, \cdots, M$ will have coalesced long before any of them gets near the level a, with high P_y-probability. More precisely, for $a, \nu, M \in \mathbb{Z}_+$, let

$$\tau_a(\nu) = \min\{n \geq 0: S_n^\nu > a\}, \tag{3.5}$$

and

$$F_a(M) = \{\sigma(M) < \tau_{a-1}(\nu) \text{ for } \nu = 0, 1, \cdots, M\}; \tag{3.6}$$

then since $P_y\{\sigma(M) < \infty\} = 1$, $\lim_{a\to\infty} P_y(F_a(M)) = 1$ for each fixed $M \in \mathbb{Z}_+$. Thus

$$P(F_a(M) \mid \mathscr{F}) \to_{P_y} 1, \quad \text{as} \quad a \to \infty. \tag{3.7}$$

Now if the processes $\{S_n^\nu\}_{n\geq 1}$ $(\nu = 0, 1, \cdots, M)$ will most likely have coalesced before any of them has reached the level a, (cf. (3.4)–(3.7)) then the expected number of visits to the point a by $\{S_n^0\}_{n\geq 1}$ (conditional on $\mathscr{F}$) should be approximately the same as the expected number of visits to a by $\{S_n^\nu\}_{n\geq 1}$ (conditional on $\mathscr{F}$), for each $\nu = 0, 1, \cdots, M$. In fact, if $A \subset \mathscr{Y}$, $\pi(A) < \infty$, then for any $\nu = 0, 1, \cdots, M$,

$$\begin{aligned} | E(\textstyle\sum_{n\geq 1} 1\{S_n^0 = a,\ Y_n \in A\} \mid \mathscr{F}) - E(\textstyle\sum_{n\geq 1} 1\{S_n^\nu = a,\ Y_n \in A\} \mid \mathscr{F}) \\ \leq 1 - P(F_a(M) \mid \mathscr{F}), \end{aligned}$$

because the difference in the numbers of visits to a by $\{S_n^0\}$ and $\{S_n^\nu\}$, respectively, cannot be greater in absolute value than 1. This obviously follows from Assumption (3A.1); in fact, this is the sole reason for making assumption (3A.1). Averaging the above inequalities for $\nu = 0, 1, \cdots, M$, we obtain

$$\begin{aligned} &| E(\textstyle\sum_{n\geq 1} 1\{S_n^0 = a,\ Y_n \in A\} \mid \mathscr{F}) \\ &\quad - E(\textstyle\sum_{n\geq 1} (M+1)^{-1} \sum_{\nu=0}^M 1\{S_n^\nu = a,\ Y_n \in A\} \mid \mathscr{F}) | \\ &\quad\quad \leq 1 - P(F_a(M) \mid \mathscr{F}). \end{aligned} \tag{3.8}$$

Next we appeal to the fact that each of the processes $\{S_n^\nu - \nu\}_{n\geq 1}$ has the same marginal distribution, conditional on $\mathscr{F}$. This implies that the ($\mathscr{F}$-conditional) probability of a visit to a by $\{S_n^\nu\}_{n\geq 1}$ is the same as the ($\mathscr{F}$-conditional) probability of a visit to $a - \nu$ by $\{S_n^0\}_{n\geq 1}$, in particular,

$$\begin{aligned} &E(\textstyle\sum_{n\geq 1} (M+1)^{-1} \sum_{\nu=0}^M 1\{S_n^\nu = a,\ Y_n \in A\} \mid \mathscr{F}) \\ &\quad = E(\textstyle\sum_{n\geq 1} (M+1)^{-1} \sum_{\nu=0}^M 1\{S_n^0 = a - \nu,\ Y_n \in A\} \mid \mathscr{F}) \\ &\quad = E((M+1)^{-1} \textstyle\sum_{n\geq 1} 1\{a - M \leq S_n^0 \leq a,\ Y_n \in A\} \mid \mathscr{F}). \end{aligned} \tag{3.9}$$

With (3.9), we have essentially reduced the problem of proving (1.6) to an "ergodic" problem, because if M is large, the last expectation in (3.9) involves the behavior of $\{S_n^0\}_{n\geq 1}$ over a long stretch of time. Recall Corollary 2.2 of Section 2: for each $\varepsilon > 0$ there exists M sufficiently large that whenever $a \geq M$,

$$\begin{aligned} P_y\{ | E((M+1)^{-1} \textstyle\sum_{n\geq 1} 1\{a - M \leq S_n^0 \leq a,\ Y_n \in A\} \mid \mathscr{F}) \\ - \mu^{-1}\pi(A) | > \varepsilon\} < \varepsilon. \end{aligned} \tag{3.10}$$

This statement may be thought of as the Law of Large Numbers "turned upside-down", although the reader should recall that its proof in the general case invoked the (unconditional) renewal theorem as well.

The proof of (1.6) is now virtually complete. Combining (3.7)–(3.10), we obtain

$$\limsup_{a\to\infty; a\in\mathbb{Z}} P_y\{|E(\textstyle\sum_{n\ge1} 1\{S_n^0 = a,\ Y_n \in A\} \mid \mathscr{F}) - \mu^{-1}\pi(A)| > \varepsilon\} \le \varepsilon. \tag{3.11}$$

Since $\varepsilon > 0$ can be chosen arbitrarily small, it must be that

$$U(\{a\}, A \mid \mathscr{F}) = E(\textstyle\sum_{n\ge1} 1\{S_n^0 = a,\ Y_n \in A\} \mid \mathscr{F}) \to_{P_y} \mu^{-1}\pi(A) \tag{3.12}$$

as $a \to \infty$ through $\mathbb{Z}$. □

PROOF OF (1.7) UNDER ASSUMPTION (3A.2). The reader should recall that in the proof of (1.6), the assumption (3A.1) was used only to establish (3.8): in fact it was really only needed to assure that the number of visits to a point a would be bounded, thus eliminating a potential uniform integrability problem. For the proof of (1.7) this difficulty does not arise, and so assumption (3A.1) may be dispensed with.

The proof of (1.7) is quite similar to that of (1.6), and certain features require no change at all. In particular we retain the notations $\tau_a(\nu)$ and $F_a(M)$ established in (3.5) and (3.6), and note that (3.7) holds.

On the event $F_a(M)$, the processes $\{S_n^\nu\}$ $(\nu = 0, 1, \cdots, M)$ have coalesced before any of them has exceeded a; consequently

$$S_{\tau_a(\nu)}^\nu - a = S_{\tau_a(0)}^0 - a \quad \text{on} \quad F_a(M), \quad \nu = 1, \cdots, M.$$

Thus for any $x = 1, 2, \cdots$ and any $A \subset \mathscr{Y}$ with $\pi(A) < \infty$,

$$\begin{aligned} &|P\{S_{\tau_a(0)}^0 - a = x;\ Y_{\tau_a(0)} \in A \mid \mathscr{F}\} \\ &\quad - (M+1)^{-1} \textstyle\sum_{\nu=0}^M P\{S_{\tau_a(\nu)}^\nu - a = x;\ Y_{\tau_a(\nu)} \in A \mid \mathscr{F}\}| \\ &\qquad \le 1 - P(F_a(M) \mid \mathscr{F}). \end{aligned} \tag{3.13}$$

Recall once again that the processes $\{S_n^\nu - \nu\}_{n\ge0}$ are marginally identical in law (conditional on $\mathscr{F}$), and notice that the first time S_n^ν exceeds a is the same as the first time $S_n^\nu - \nu$ exceeds $a - \nu$. Therefore

$$\begin{aligned} &(M+1)^{-1} \textstyle\sum_{\nu=0}^M P\{S_{\tau_a(\nu)}^\nu - a = x;\ Y_{\tau_a(\nu)} \in A \mid \mathscr{F}\} \\ &\quad = (M+1)^{-1} \textstyle\sum_{\nu=0}^M P\{S_{\tau_{a-\nu}(0)}^0 - (a-\nu) = x;\ Y_{\tau_{a-\nu}(0)} \in A \mid \mathscr{F}\}. \end{aligned} \tag{3.14}$$

Finally, Proposition 2.3 of Section 2 implies that for each $\varepsilon > 0$ there exists $M \in \mathbb{Z}_+$ sufficiently large that whenever $a \ge M$,

$$\begin{aligned} &P_y\{|(M+1)^{-1} \textstyle\sum_{\nu=0}^M P\{S_{\tau_{a-\nu}(0)}^0 - (a-\nu) = x;\ Y_{\tau_{a-\nu}(0)} \in A \mid \mathscr{F}\} \\ &\qquad - G(\{x\}, A)| > \varepsilon\} < \varepsilon. \end{aligned} \tag{3.15}$$

Combining (3.7) with (3.13)–(3.15), we obtain

$$\limsup_{a\to\infty} P_y\{|P\{S0\tau a(0) - a = x;\ Y\tau a(0) \in A \mid \mathscr{F}\} - G(\{x\}, A)| > \varepsilon\} \le \varepsilon. \tag{3.16}$$

Since $\varepsilon > 0$ was arbitrary, it follows that

$$P\{S^0_{\tau_a(0)} - a = x;\ Y_{\tau_a(0)} \in A \mid \mathscr{F}\} \to_{P_y} G(\{x\}, A) \quad \text{as} \quad a \to \infty \quad \text{through} \quad \mathbb{Z}. \tag{3.17}$$

This proves (1.7) for sets B consisting of only finitely many points. It is entirely routine to deduce that (1.7) holds for infinite sets B as well; we omit the details. □

PROOF OF (1.6) UNDER ASSUMPTION (3A.2). Assumption (3A.1) was only used in establishing the inequality (3.8). Without assumption (3A.1) this inequality may fail, since the Markov random walk $\{S^\nu_n\}_{n\ge1}$ may visit a indefinitely often; however, it is possible to circumvent (3.8), and thus to establish (1.6) without assuming (3A.1).

Fix $\varepsilon > 0$ and $M \in \mathbb{Z}_+$. By the unconditional Markov Renewal Theorem, the inequalities

$$|E_y(\textstyle\sum_{n\ge1} 1\{S^\nu_n = a;\ Y_n \in A\}) - \mu^{-1}\pi(A)| < \varepsilon, \quad \nu = 0, 1, \cdots, M, \tag{3.18}$$

hold for all integers a sufficiently large (and) sets $A \subset \mathscr{Y}$ with $\pi(A) < \infty$. By Corollary 2.2 of Section 2, there is an integer $M = M(\varepsilon)$ sufficiently large that whenever $a \ge M$,

$$P_y\{|E((M+1)^{-1} \textstyle\sum_{n\ge1} 1\{a - M \le S^0_n \le a;\ Y_n \in A\} \mid \mathscr{F}) - \mu^{-1}\pi(A)| > \varepsilon\} < \varepsilon. \tag{3.19}$$

Since the equality (3.9) persists when Assumption (3A.1) is dropped (the original argument made no use of (3A.1), we may rewrite (3.19) as

$$P_y\{|E((M+1)^{-1} \textstyle\sum_{n\ge1} \sum_{\nu=0}^{M} 1\{S^\nu_n = a;\ Y_n \in A\} \mid \mathscr{F}) - \mu^{-1}\pi(A)| > \varepsilon\} < \varepsilon. \tag{3.20}$$

Recall that $P_y(F_a(M)) \to 1$ as $a \to \infty$ (cf. (3.7) and (3.6)); consequently by combining (3.18) and (3.20) we may conclude that for all $a \in \mathbb{Z}_+$ sufficiently large,

$$P_y\{|E((M+1)^{-1} \textstyle\sum_{n\ge1} \sum_{\nu=0} 1\{S^\nu_n = a;\ Y_n \in A\} \cdot 1(F_a(M)) \mid \mathscr{F}) - \mu^{-1}\pi(A)| > 2\,\varepsilon\} < 2\varepsilon. \tag{3.21}$$

Now on the event $F_a(M)$ the random walks $\{S^\nu_n\}$ ($\nu = 0, 1, \cdots, M$) coalesce before reaching a; hence

$$(M+1)^{-1} \textstyle\sum_{n\ge1} \sum_{\nu=0} 1\{S^\nu_n = a;\ Y_n \in A\} \cdot 1(F_a(M)) = \sum_{n\ge1} 1\{S^0_n = a;\ Y_n \in A\} \cdot 1(F_a(M)). \tag{3.22}$$

Thus (3.12) becomes

$$P_y\{|E(\textstyle\sum_{n\ge1} 1\{S^0_n = a;\ Y_n \in A\} \cdot 1(F_a(M)) \mid \mathscr{F}) - \mu^{-1}\pi(A)| > 2\varepsilon\} < 2\varepsilon. \tag{3.23}$$

Using (3.18) (this time only for $\nu = 0$) and the fact that $P_y(F_a(M)) \to 1$ as

$a \to \infty$, we infer from (3.23) that for $a \in \mathbb{Z}_+$ sufficiently large

$$P_y\{|E(\sum_{n\geq 1} 1\{S_n^0 = a;\ Y_n \in A\} \mid \mathscr{F}) - \mu^{-1}\pi(A)| > 4\varepsilon\} < 4\varepsilon.$$

Letting $\varepsilon \downarrow 0$ we find that we have proved (1.6). □

3B. Removing the lattice condition. The assumption (3A.2), as we remarked earlier, is stronger than the arithmetic hypothesis (1.1A) of Theorem 1. Without it, however, the coupling described in the previous subsection may fail: if, for example $P_y\{\sum_{n=1}^{N_1-1} X_n$ is an odd integer$\} = 1$, then Markov random walks started at 0 and 1 respectively can *never* meet, since at times N_j they must have opposite parities. Fortunately there is a modification of the coupling approach which resolves this odd dilemma: we will call this modification a "staggering argument" for reasons which should become clear.

Assume then that under P_y, $\sum_{n=1}^{N_1-1} X_n$ has a distribution which is supported by $\alpha + \beta\mathbb{Z}$ (where $\beta \geq 2$, $0 \leq \alpha < \beta$, and $\alpha, \beta \in \mathbb{Z}$), and by no coarser lattice (a "lattice" being a coset of a subgroup of $\mathbb{Z}$). If $\{X_n^A\}$ and $\{X_n^B\}$ are (conditionally) independent copies of $\{X_n\}$ (i.e., if (3.2) holds) then

$$\sum_{n=1}^{N_1} (X_n^A - X_n^B)$$

has a distribution which is supported by $\beta\mathbb{Z}$ but by no proper subgroup of $\beta\mathbb{Z}$. According to the theorem of Chung, Fuchs, and Ornstein, the process

$$\sum_{n=1}^{N_k-1} (X_n^A - X_n^B), \quad k = 0, 1, \cdots$$

will visit every point of $\beta\mathbb{Z}$ infinitely often, with P_y-probability one.

Define random walks $\{S_n^\nu\}$ and coupling times $\{\sigma(\nu)\}$ as follows:

$$S_0^\nu = \beta\nu, \quad \nu = 0, 1, \cdots;$$

$$S_{n+1}^\nu = \begin{cases} S_n^\nu + X_{n+1}^A & \text{if} \quad n \geq \sigma(\nu) \\ S_n^\nu + X_{n+1}^B & \text{if} \quad n < \sigma(\nu); \end{cases}$$

$$\sigma(0) = 0$$

$$\begin{aligned}\sigma(\nu+1) &= \min\{n > \sigma(\nu)\colon S_n^{\nu+1} = S_n^0\} \\ &= \min\{n > \sigma(\nu)\colon \textstyle\sum_{j=1}^n (X_j^A - X_j^B) = \beta\nu + \beta\}.\end{aligned}$$

(The coupling times $\sigma(\nu)$ are all finite). Notice that

(i) $S_n^\nu = S_n^{\nu-1} = \cdots = S_n^1 = S_n^0\ \forall n \geq \sigma(\nu)$, and
(ii) each of the processes $\{S_n^\nu = \beta\nu\}_{n\geq 0}$ $(\nu \geq 1)$ has the same distribution as $\{S_n^0\}_{n\geq 0}$, conditional on $\mathscr{F}$.

We may now repeat the arguments in (3.7)–(3.9), concluding that for large a

$$\begin{aligned}&E(\textstyle\sum_{n\geq 1} 1\{S_n^0 = a,\ Y_n \in A\} \mid \mathscr{F}) \\ &\quad\approx E(\textstyle\sum_{n\geq 1} (M+1)^{-1} \sum_{\nu=0}^M 1\{S_n^0 = a - \beta\nu,\ Y_n \in A\} \mid \mathscr{F}) \\ &\quad= E((M+1)^{-1} \textstyle\sum_{n\geq 1} 1\{a - M\nu \leq S_n^0 \leq a;\ Y_n \in A;\ S_n^0 - a \equiv 0 \bmod \nu\} \mid \mathscr{F}) \\ &\quad\approx \mu^{-1}\pi(A).\end{aligned}$$

The last approximate equality is a consequence of Proposition 2.4 (compare with the use of Corollary 2.2 in deriving (3.10)). The rest of the argument needed to justify (1.6) is the same as that in (3.11)–(3.12) and (3.18)–(3.23).

The proof of (1.7) follows the same lines. □

3C. The nonarithmetic case. If the increments in the Markov random walk are not valued in a discrete subgroup of $\mathbb{R}$ (or a coset of one), then the coupling scheme described in subsection 3A will generally "fail," since independent walks commencing at distinct points will not visit the same points. Fortunately there is, again, an easy way out of this predicament: instead of insisting that coupling take place at an instant when two random walkers are at the same point, we will allow coupling when the distance between the walkers is less than K^{-1}, where K is a large integer. This "approximate coupling" trick was also used by Lindvall (1977) in his proof of Blackwell's Renewal Theorem. For the problem at hand (to wit, proving Theorem 1) the approximate coupling must be supplemented by another not-so-very-subtle trick: rather than starting the different random walks $\{S_n^\nu\}_{n\ge 0}$ at integer points, as in Subsections 3A and 3B, we must start them at integer multiples of K^{-1}, with K to be eventually made large so as to obtain a good approximation to Lebesgue measure.

There is once again a "lattice problem," which occurs when the distribution of

$$\textstyle\sum_{n=1}^{N_1-1} X_n$$

is concentrated on a lattice $\alpha + \beta\mathbb{Z}$ (recall that $1 = N_0, N_1, \cdots$ are the times of successive visits to state y by $\{Y_n\}_{n\ge 1}$). In the nonarithmetic case α and $\beta > 0$ should be linearly independent over $\mathbb{Z}$. In the lattice nonarithmetic case the "staggering" scheme described in in Subsection 3B must again be used. We will not trouble the reader with the (entirely mundane) details in this case. For the rest of the discussion, then, we will assume

(3C) *the distribution of $\sum_{n=1}^{N_1-1} X_n$ is not supported by any coset of a discrete subgroup of $\mathbb{R}$.*

As in Subsection (3A), let $\{X_n^A\}_{n\ge 1}$ and $\{X_n^B\}_{n\ge 1}$ be independent copies of $\{X_n\}_{n\ge 1}$ (conditional on $\mathscr{F}$): i.e., suppose (3.2) holds. Fix $K \in \mathbb{Z}_+$, and define $\{S_n^\nu\}_{n\ge 0}$, $\nu \ge 0$, and $\sigma(\nu)$, $\nu \ge 0$, by

$$\begin{aligned}
S_0^\nu &= \nu/K \\
S_{n+1}^0 &= S_n^0 + X_{n+1}^A \\
S_{n+1}^\nu &= \begin{cases} S_n^\nu + X_{n+1}^A & \text{if} \quad \sigma(\nu) \le n \\ S_n^\nu + X_{n+1}^B & \text{if} \quad \sigma(\nu) > n \end{cases} \\
\sigma(0) &= 0 \\
\sigma(\nu+1) &= \min\{n > \sigma(\nu) \colon 0 \le S_n^{\nu+1} - S_n^0 < 1/K\} \\
&= \min\{n > \sigma(\nu) \colon \nu/K < \textstyle\sum_{j=1}^{n} (X_n^A - X_n^B) \le (\nu+1)/K\}.
\end{aligned} \tag{3.24}$$

The fact that each $\sigma(\nu) < \infty$ (with P_y-probability one) follows from the nonlattice version of the Chung–Fuchs–Ornstein theorem (thanks to (3C)).

PROOF OF (1.6): A BRIEF SYNOPSIS. This once again rests on two obvious properties of the processes $\{S_n^\nu\}$:

(i) the marginal distribution of $\{S_n^\nu - \nu/K\}_{n\geq 0}$ (conditional on $\mathscr{F}$) does not depend on ν; and
(ii) for each $M \geq 1$, $|S_n^\nu - S_n^0| < 1/K$ whenever $n \geq \sigma(M)$ and $0 \leq \nu < M$.

Let $\tau_a(\nu)$ and $F_a(M)$ be defined by (3.5) and (3.6) (this time for $a \in \mathbb{R}_+$); notice that (3.7) retains its validity. By property (ii),

$$
\begin{aligned}
&(M+1)^{-1} \textstyle\sum_{n\geq 1} \sum_{\nu=0}^{K} 1\{S_n^\nu \in (a + K^{-1}, a+1);\ Y_n \in A\} \\
&\qquad \leq \textstyle\sum_{n\geq 1} 1\{S_n^0 \in (a, a+1);\ Y_n \in A\} \qquad (3.25)\\
&\qquad \leq (M+1)^{-1} \textstyle\sum_{n\geq 1} \sum_{\nu=0}^{K} 1\{S_n^\nu \in (a, a+1+K^{-1});\ Y_n \in A\}
\end{aligned}
$$

on the event $F_a(M)$.

One may now argue as in the proof of (1.6) given in Subsection 3A, under Assumption (3A.2); the two-sided inequality (3.25) replaces (3.22). Using the unconditional Markov renewal theorem, Corollary 2.2 of Section 2, and (3.7), and mimicking the development in (3.18)–(3.21), one obtains

$$P_y\{|E(\textstyle\sum_{n\geq 0} 1\{S_n^0 \in (a, a+1);\ Y_n \in A\} \mid \mathscr{F}) - \mu^{-1}\pi(A)| > 8/K\} < 8/K$$

for all $a \in \mathbb{R}_+$ sufficiently large. Since $K \in \mathbb{Z}_+$ is arbitrary, (1.6) must hold. □

A similar approach works for (1.7).

4. Counterexamples.

EXAMPLE 1. This example illustrates the need for a restriction on the sum of the means over an excursion (cf. (1.9) in Theorem 3). Let G be the probability measure on $\mathscr{Y} = \{1, 2, 3, \cdots\}$ given by

$$G(\{y\}) = Cy^{-3}, \quad \text{where} \quad C^{-1} = \zeta(3) = \textstyle\sum_1^\infty y^{-3},$$

and let $\{Y_n\}_{n\geq 1}$ be i.i.d. with distribution G. Notice that $G = \pi$ is the stationary distribution, and that the recurrence times have exponential moments (hence (1.8) holds).

For each $y \in \mathscr{Y}$ let F_y be the uniform distribution on the interval $[y-1, y+1]$. Then $\mu_y = y$ and $\sigma_y^2 = 1/3$; since (1.8) holds, it follows that (1.10) holds. It is an easy exercise to show that

$$E_y(\textstyle\sum_{j=1}^{N_1-1} \mu_{Y_j})^{2-\delta} < \infty \quad \forall \delta > 0, \quad \text{but} \quad E_y(\sum_{j=1}^{N_1-1} \mu_{Y_j})^2 = \infty.$$

Notice that the Markov random walk specified by $\{Y_n\}$ and $\{F_y\}$ may be written as

$$S_n = \textstyle\sum_{j=1}^n X_j \quad \text{where} \quad X_n = Y_n + Z_n,$$

where $Z_1, Z_2, \cdots$ are i.i.d. uniform on $[-1, 1]$, and are independent of $\mathscr{F} = \sigma(Y_1, Y_2, \cdots)$.

That almost sure convergence in (1.6) and (1.7) fails follows from the divergence of the series

$$\sum_{n=1}^{\infty} P\{Y_n > \sqrt{n \log n}\}.$$

The Borel–Cantelli Lemma implies that $Y_n > \sqrt{n \log n}$ occurs infinitely often, w.p.1, so there exists a sequence of (random) integers J_k, $k = 1, 2, \cdots$, such that $J_k < J_{k+1}$ for each k, and $Y_{J_k} > \sqrt{J_k \log J_k}$ w.p.1. But the law of the iterated logarithm implies that

$$|\sum_{j=1}^{n} Z_j| > \sqrt{2n \log \log n}$$

occurs only finitely often, w.p.1. (since var $Z_j = 1/3 < 1$). Consequently, if

$$u_k = \sum_{\ell=1}^{J_k-1} Y_\ell + [|\sqrt{J_k \log J_k}/2|]$$

and

$$A_k = \{\sum_{j=1}^{n} X_j \in [u_k, u_k + 1] \text{ for some } n\},$$

then $1(A_k) \to 0$ w.p.1, and therefore

$$E_y(\sum_{n\geq 1} 1\{u_k \leq S_n \leq u_k + 1\} \mid \mathscr{F}) \to 0 \quad \text{a.s.},$$

and

$$P_y(S_{\tau(u_k)} - u_k \to \infty \text{ as } k \to \infty \mid \mathscr{F}) = 1. \quad \text{a.s.} \quad \square$$

Example 2. The preceding example illustrated the possibility of creating "predictable gaps" in the conditional renewal measure. If the driving process $\{Y_n\}$ is badly behaved, then it is possible to have "predictable irregularities" in the conditional renewal measure even when the variation in the distributions $\{F_y\}_{y\in\mathscr{Y}}$ is mild (clearly there must be *some* variation in order for the renewal theorem to fail, by Blackwell's Theorem).

Consider the positive recurrent Markov chain $\{Y_n\}_{n\geq 1}$ taking values in $\mathscr{Y} = \{0, 1, 2, \cdots\}$ with initial state $Y_1 = 0$ and transition laws

$$p(y, y-1) = 1 \quad \forall y \geq 1 \qquad p(0, y) = C^{-1}y^{-3}y \quad \forall y \geq 1$$

where $C = \zeta(3)^{-1}$. Notice that for $N = \min\{n > 1: Y_n = 0\}$, $E_y N^2 = \infty$, but $EyN^{2-\delta} < \infty$ for all $\delta > 0$. Let F_0 be the uniform distribution on [1, 2], and let $F_y(y \geq 1)$ be the uniform distribution on [2, 3]. The Markov random walk specified by the transition mechanism $p(\cdot, \cdot)$ and the assignment $y \to F_y$ may be represented by

$$S_n = \sum_{j=1}^{n} X_j \quad \text{where} \quad X_n = 1 + 1\{Y_n \geq 1\} + Z_n$$

and $Z_1, Z_2, \cdots$ are i.i.d., independent of $\mathscr{F} = \sigma(Y_1, Y_2, \cdots)$, uniformly distributed on [0, 1].

The failure of the (conditional) renewal theorem in this example stems from the existence of large excursions in the driving chain $\{Y_n\}$. Let $N_0 = 1$, $N_1, N_2, \cdots$ be the times of successive returns to the initial state 0 by $\{Y_n\}$; then

$$P_0\{N_{k+1} - N_k = n + 1\} = \zeta(3)^{-1}n^{-3}.$$

The Borel–Cantelli Lemma implies that with P_y-probability one, $N_{k+1} - N_k >$

$C_1\sqrt{k \log k}$ infinitely often for each $C_1 > 0$, since $\sum_k (k \log k)^{-1} = \infty$. But by the SLLN, $N_k/k \to E_y N_1 - 1$ almost surely, so it follows that $Y_n > C_2\sqrt{n \log n}$ occurs infinitely often for each $C_2 > 0$, with P_y-probability one. Let $J_k \uparrow$ be an increasing sequence of (random) integers such that $P_0\{Y_{J_k} > \sqrt{J_k \log J_k}\ \forall k\} = 1$.

Now $Z_1, Z_2, \cdots$ are i.i.d. with variance $1/12$, so the law of the iterated logarithm implies that

$$P_y\{|\textstyle\sum_{j=1}^n Z_j - n/2| > \sqrt{n \log \log n} \text{ i.o.}\}$$
$$= P_y\{|S_n - \textstyle\sum_{j=1}^n \mu_{Y_j}| > \sqrt{n \log \log n} \text{ i.o.}\} = 0.$$

Furthermore, starting at time $n = J_k$ the Markov random walk $\{S_m\}$ is obligated to take its next $\sqrt{J_k \log J_k}$ increments from the uniform distribution on [2, 3], since $Y_{J_k} > \sqrt{J_k \log J_k}$. But Blackwell's Theorem will govern the renewal measure in this stretch of time, since conditional on $\mathscr{F}$ the increments from time $n = J_k$ to $n = J_k\sqrt{J_k \log J_k}$ are i.i.d.! Thus if

$$u_k =_\Delta [|\textstyle\sum_{j=1}^{J_k} \mu_{Y_j} + \frac{1}{2} J_k \log J_k|],$$

then

$$E_0(\textstyle\sum_{n\ge 1} 1\{u_k \le S_n < u_k + 1\} \mid \mathscr{F}) \to \frac{1}{\int_2^3 x\,dx} = \frac{2}{5}.$$

Since $5/2 \ne \mu =_\Delta \sum \mu_y \pi(y)$ ($\mu < 5/2$), the conditional renewal measure cannot converge almost surely, in view of Theorem 1. □

5. The rate of coalescence in multiple coupling. In the examples of the preceding section the conditional renewal measures fail to converge because the Markov chains $\{Y_n\}$ wander off on wild excursions before the random walk $\{S_n\}$ can "spread out." This suggests that in the coupling scheme of Section 3 either (i) the law of large numbers (cf. (3.10)) takes effect too slowly, or (ii) the processes $\{S_n^\nu\}$ take too long to coalesce, in order for almost sure convergence in (1.6). In this section we will examine the "coupling rate" for two schemes similar to that described in Section 3; we shall find that the number of steps necessary to weld together $\{S_n^0\}, \{S_n^1\}, \cdots, \{S_n^\nu\}$, where each $\{S_n^j - j\}$ is a replica of $\{S_n^0\}$, is roughly on the order of ν^2.

One might wonder whether a more clever coupling than ours might yield a better rate than ν^2. A moment's reflection, however, will reveal that unless the processes $\{S_n^0\}$ and $\{S_n^\nu\}$ are allowed to couple at different times (i.e., different values of n) no better rate can be achieved, since this would in effect contradict the (local) Central Limit Theorem. As for allowing the random walks to couple at different times—this would be useless for the problem at hand, since *conditional on* $\mathscr{F}$ the processes $\{S_n^\nu\}$ are inhomogeneous in time.

We shall assume throughout this section that for *some* state $y \in \mathscr{Y}$ the distribution F_y *and its symmetrization* are both supported by $\mathbb{Z}$ and by no proper subgroup of $\mathbb{Z}$. This assumption obviates the necessity of "staggering" the processes $\{S_n^\nu\}_{n\ge 1}$ in the construction to follow (cf. Subsection 3B). As in Section 3, let $\{X_n^A\}_{n\ge 1}$ and $\{X_n^B\}_{n\ge 1}$ be conditionally independent copies of the increment sequence: specifically, assume that (3.2) holds. Also, let $N_0 = 1, N_1, N_2, \cdots$ be the instants of successive visits to y by the Markov chain $\{Y_n\}_{n\geq 1}$. According to our assumption concerning the distribution F_y,

$$P_y((X^A_{N_j} - X^B_{N_j}) \in \beta\mathbb{Z} \mid \mathscr{F}) < 1 \quad \forall \beta \geq 2; \tag{5.1}$$

it follows that there is an integer $M < \infty$ sufficiently large that

$$P_y((X^A_{N_j} - X^B_{N_j})1\{|X^A_{N_j} - X^B_{N_j}| \leq M\} \in \beta\mathbb{Z} \mid \mathscr{F}) < 1 \quad \forall \beta \geq 2. \tag{5.2}$$

The coupling scheme to be used in this section differs from that of Section 3 in two respects. First, the processes $\{S^\nu_n\}$ ($\nu \geq 1$) will not proceed independently until coupling, as earlier; in fact, the increments in S^ν_n and S^0_n, respectively, will never be more than M in absolute value. These two modifications allow us to use results for random walks with *bounded* increments, which are more "precise" than the Chung–Fuchs–Ornstein theorem.

Let

$$\xi_j = (X^A_{N_j} - X^B_{N_j})1\{|X^A_{N_j} - X^B_{N_j}| \leq M\}, \quad j = 0, 1, 2, \cdots; \tag{5.3}$$

then conditional on $\mathscr{F}$ (under P_y), $\xi_0, \xi_1, \cdots$ are i.i.d. with symmetric distribution whose support generates $\mathbb{Z}$. Define random processes $\{S^\nu_n\}_{n\geq 0}$, $\nu = 0, 1, \cdots$, and coupling indices $\alpha(\nu)$, by

$$\begin{aligned} S^\nu_0 &= \nu \\ S^0_{n+1} &= S^0_n + X^A_{n+1} \\ S^\nu_{n+1} &= S^\nu_n + X^B_{n+1}, \quad \text{if} \quad n+1 = N_j \quad \text{(some } j\text{)}, \\ &\qquad\qquad\qquad\qquad n < \alpha(\nu), \quad \text{and} \quad \xi_j \neq 0; \\ &\quad\; S^\nu_n + X^A_{n+1}, \quad \text{otherwise}; \\ \alpha(0) &= 0; \quad \text{and} \\ \alpha(\nu + 1) &= \min\{j > \alpha(\nu): S^{\nu+1}_{N_j} = S^0_{N_j}\}. \\ &= \min\{j > \alpha(\nu): \xi_0 + \xi_1 + \cdots + \xi_j = \nu + 1\}. \end{aligned} \tag{5.4}$$

The two critical features of the construction are that (i) the processes $\{S^\nu_n\}$ and $\{S^0_n\}$ coalesce at time $N_{\alpha(\nu)}$, i.e.,

$$S^0_n = S^1_n = \cdots = S^\nu_n \quad \text{on} \quad \{n \geq N_{\alpha(\nu)}\}; \tag{5.5}$$

and (ii) the processes $\{S^\nu_n - \nu\}$ are all marginally the same (in law) conditional on $\mathscr{F}$, i.e.

$$\mathscr{L}_y(\{S^\nu_n - \nu\}_{n\geq 0} \mid \mathscr{F}) = \mathscr{L}_y(\{S^0_n\}_{n\geq 0} \mid \mathscr{F}) \quad \forall \nu = 0, 1, \cdots \tag{5.6}$$

(here $\mathscr{L}_y$ denotes the law of the process under the probability measure P_y). These facts are completely elementary.

Proposition 5.1. *As* $\nu \to \infty$

$$\mathscr{L}_y((\alpha(\nu)/\nu^2) \mid \mathscr{F}\ ) \to_{\text{A.S.}(P_y)} \textit{one-sided stable} - 1/2. \tag{5.7}$$

The proof relies on a potential-theoretic result of Kesten and Spitzer (1963) (cf. Proposition 4, Section 32 of Spitzer (1975); the analogous result for the nonarithmetic case is in Port and Stone (1969)).

THEOREM. *Suppose $\xi_1, \xi_2, \cdots$ are i.i.d. with $E\xi_i = 0$ and $E\xi_i^2 < \infty$; suppose also that the distribution of ξ_i is supported by $\mathbb{Z}$ but by no proper subgroup of $\mathbb{Z}$. Then for each $x \in \mathbb{Z}$ there exists a constant $C_x \in (0, \infty)$ such that*

$$P\{t_x > n\} \sim C_x n^{-1/2} \quad \text{as} \quad n \to \infty,$$

where $t_x = \min\{n \geq 1: \xi_1 + \cdots \xi_n = x\}$.

PROOF OF PROPOSITION 5.1. Recall from the construction (5.4) that

$$\begin{aligned}\alpha(\nu + 1) &= \min\{j > \alpha(\nu): \xi_0 + \cdots + \xi_j = \nu + 1\} \\ &= \min\{j > \alpha(\nu): \xi_{\alpha(\nu)+1} + \xi_{\alpha(\nu)+2} + \cdots + \xi_j = 1\}.\end{aligned}$$

Moreover, conditional on $\mathscr{F}$, the random variables $\xi_1, \xi_2, \cdots$ are i.i.d., with mean zero, finite variance, and a distribution which is supported by $\mathbb{Z}$ but by no proper subgroup (cf. (5.2) and (5.3)). Consequently

$$\alpha(\nu) = \beta_1 + \beta_2 + \cdots + \beta_\nu$$

where $\beta_k = \alpha(k) - \alpha(k - 1)$ are (conditional on $\mathscr{F}$) i.i.d. By the Kesten–Spitzer theorem

$$P\{\beta_1 > n \mid \mathscr{F}\} \sim C_1 n^{-1/2};$$

thus by a standard result in the theory of stable laws (cf. Theorem XIII.6.2 of Feller, 1966),

$$P\{\alpha(\nu)/\nu^2 \leq s \mid \mathscr{F}\} \to G_{1/2}(s)$$

where $G_{1/2}(s)$ is the distribution function for a one-sided stable law of exponent ½. □

6. Almost sure convergence: finite state space. When the state space $\mathscr{Y}$ is finite, it is very difficult for the Markov chain $\{Y_n\}$ to embark on a "wild" excursion, as in the examples of Section 4: there is simply nowhere for it to go. This is reflected in the following large deviation theorem, according to which the empirical distribution of $\{Y_n\}$ approaches the stationary distribution π very quickly.

PROPOSITION 6.1. *If $\mathscr{Y}$ is finite, then for each $\varepsilon > 0$ there exist constants $C_\varepsilon < \infty$ and $\lambda(\varepsilon) > 0$ such that*

$$(6.1) \quad P_y\{\textstyle\sum_{y' \in \mathscr{Y}} |m^{-1} \sum_{n=1}^m 1\{Y_n = y'\} - \pi(y')| > \varepsilon\} \leq C_\varepsilon \cdot \exp(-m\lambda(\varepsilon))$$

for all $m \geq 1$ and each $y \in \mathscr{Y}$.

Much more general results than this are well-known: cf. Donsker and Varadhan (1975, 1976). Proposition 6.1 has a relatively simple elementary proof which will be given in due course.

Proposition 6.1 allows us to obtain a rate of convergence in the law of large numbers for the Markov random walk $\{S_n\}$. This, combined with the result of Section 5 concerning the rate of coupling, will enable us to establish almost sure convergence in (1.6).

Suppose that $S_n = X_1 + \cdots + X_n$ is a Markov random walk with the Markov chain $\{Y_n\}_{n\geq 1}$ as driving process. The conventions (1.1)–(1.3) are assumed to be in force. For the rest of this section it will be understood that the state space $\mathscr{Y}$ of the driving process $\{Y_n\}_{n\geq 1}$ is finite.

LEMMA 6.2. *For each $\varepsilon > 0$ there exist constants $C_\varepsilon < \infty$ and $\lambda(\varepsilon) > 0$ such that*

$$P_y\{P(|S_n - n\mu| > n\varepsilon \mid \mathscr{F}) > \varepsilon\} \leq C_\varepsilon \exp(-n\lambda(\varepsilon)) \tag{6.2}$$

for all $n \geq 1$ and for each $y \in \mathscr{Y}$.

The constants C_ε and $\lambda(\varepsilon)$ may differ from those in (6.1). The proof, which is a relatively straightforward exercise in the use of the SLLN and Proposition 6.1, will be given later in the section.

The almost sure convergence of the conditional renewal measure will only be established for Markov random walks $\{S_n\}$ with strictly positive increments. Thus we will assume for the remainder of this section that each of the distributions F_y, $y \in \mathscr{Y}$ is concentrated on the set of positive integers, i.e.,

$$F_y(\{1, 2, \cdots\}) = 1 \quad \forall y \in \mathscr{Y}. \tag{6.3}$$

The reader should notice that with (6.3) in force, $\{S_n\}_{n\geq 1}$ is strictly increasing in n, and S_n never visits a point a more than once. Consequently the (conditional) renewal measure is just the (conditional) hitting probability function:

$$E_y \textstyle\sum_{n\geq 1} 1\{S_n = a;\ Y_n \in A\} = P_y\{(S_n, Y_n) \text{ visits } \{a\} \times A \text{ for some } n\}. \tag{6.4}$$

LEMMA 6.3. *Fix $\rho > 0$ and $\varepsilon > 0$. Then there exist constants $C(\rho, \varepsilon) < \infty$ and $\lambda(\rho, \varepsilon) > 0$ such that for each $y \in \mathscr{Y}$ and $A \subset \mathscr{Y}$, and each $n \geq 1$,*

$$\begin{aligned} P_n\{\max\nolimits_{0\leq a\leq n^\rho} | E(n^{-1} \textstyle\sum_{m\geq 1} 1\{a \leq S_m \leq a + n;\ Y_m \in A\} \mid \mathscr{F}) \\ - \mu^{-1}\pi(A) | > \varepsilon\} \leq C(\varepsilon, \rho)\exp(-n\lambda(\varepsilon, \rho)). \end{aligned} \tag{6.5}$$

Consequently,

$$\begin{aligned} \max\nolimits_{0\leq a\leq n^\rho} | E(n^{-1} \textstyle\sum_{m\geq 1} 1\{a \leq S_m \leq a + n;\ Y_m \in A\} \mid \mathscr{F}) \\ - \mu^{-1}\pi(A) | \to 0 \end{aligned} \tag{6.6}$$

almost surely (relative to P_y).

The proof of this lemma is a long but straightforward affair in which the rates (6.1) and (6.2) are translated into corresponding rates for the renewal measure. We defer the details until later in the section.

PROOF OF THEOREM 2. We will consider only the convergence of the (conditional) renewal measure; the proof of almost sure convergence in (1.7) is virtually identical. Furthermore, we will assume that the initial state y is such that neither F_y nor its symmetrization is supported by a proper subgroup of $\mathbb{Z}$. This allows us to call on our results concerning the coupling scheme established in (5.4).

REMARKS.

(1) In order to prove that

$$E_y(\textstyle\sum_{n\geq 1} 1\{S_n = a;\ Y_n \in A\} \mid \mathscr{F}) \to \mu^{-1}\pi(A) \quad (a \to \infty) \tag{6.7}$$

almost surely (P_y) for *every* $y \in \mathscr{Y}$ it suffices to establish it for *any particular* value of y. This is because $\{Y_n\}_{n\geq 1}$ visits every point of $\mathscr{Y}$ and any "initial segment" of the process (S_n, Y_n) cannot affect the validity of (6.7). We leave it to the reader to supply a formal proof.

(2) If there is no point $y \in \mathscr{Y}$ for which (5.1) holds, then the coupling scheme of (5.4) must be modified to incorporate an appropriate "staggering," as in Section 3B. In all other respects the argument given below applies verbatim.

The first phase of the proof is to retrace the line of reasoning in Section 3A, especially the relations (3.8) and (3.9). This leads to

$$\begin{aligned} &|E(\textstyle\sum_{n\geq 1} 1\{S_n^0 = a;\ Y_n \in A\} \mid \mathscr{F}) \\ &\quad - E(\textstyle\sum_{n\geq 1} (M+1)^{-1} 1\{a - M \leq S_n^0 \leq a;\ Y_n \in A\} \mid \mathscr{F})| \\ &\qquad \leq 1 - P(\tilde{F}_a(M) \mid \mathscr{F}), \end{aligned} \tag{6.8}$$

where

$$\tilde{F}_a(M) = \{N_{\alpha(M)} < \tau_{a-1}(\nu) \text{ for each } \nu = 0, 1, \cdots, M\} \tag{6.9}$$

and

$$\tau_a(\nu) = \min\{n \geq 1 \colon S_n^\nu > a\} \tag{6.10}$$

(the coupling index $\alpha(\nu)$ is defined in (5.4)). Notice that (6.8) is valid for all values of M and a.

Define events $\Gamma_1(m)$ and $\Gamma_2(m)$ as follows:

$$\Gamma_1(m) = \{\min\nolimits_{0\leq\nu\leq m} \tau_{\mu m^3}(\nu) < m^3/2\} \quad \text{(early crossing)}, \tag{6.11}$$

and

$$\Gamma_2(m) = \{N_{\alpha(m+1)} \geq m^3/2\} \quad \text{(late coupling)}. \tag{6.12}$$

Clearly

$$1 - P(\tilde{F}_a(m+1) \mid \mathscr{F}) \leq P(\Gamma_1(m) \mid \mathscr{F}) + P(\Gamma_2(m) \mid \mathscr{F})$$

for every integer a such that $a \geq \mu m^3$. By (6.6) of Lemma 6.3,

$$\begin{aligned} \max\nolimits_{\mu m^3 \leq a \leq \mu(m+1)^3} &|E(\textstyle\sum_{n\geq 1}(m+1)^{-1} 1\{a - m \leq S_n^0 \leq a;\ Y_n \in A\} \mid \mathscr{F}) \\ &\quad - \mu^{-1}\pi(A)| \to 0 \quad \text{almost surely} \quad (P_y) \end{aligned}$$

as $m \to \infty$; consequently to prove (6.7) it suffices, in view of (6.8), to show that

$$P(\Gamma_1(m) \mid \mathscr{F}) \to 0 \quad \text{almost surely} \quad (P_y), \tag{6.13}$$

and

$$P(\Gamma_2(m) \mid \mathscr{F}) \to 0 \quad \text{almost surely} \quad (P_y). \tag{6.14}$$

PROOF OF (6.14). Recall that $N_0 = 1, N_1, N_2, \cdots$ are the times of successive returns to state y by $\{Y_n\}_{n\geq 1}$. By the ergodic theorem

$$\frac{N_{[m^{5/2}]}}{m^{5/2}} \to \pi(y)^{-1} \quad \text{A.S.} \quad (P_y);$$

by Proposition 5.1,

$$P\{\alpha(m+1) > m^{5/2} \mid \mathscr{F}\} \to 0 \quad \text{A.S.} \quad (P_y);$$

therefore

$$P(N_{\alpha(m+1)} \geq \pi(y)^{-1} m^{5/2} \mid \mathscr{F}) \to 0 \quad \text{A.S.} \quad (P_y).$$

PROOF OF (6.13). Recall that on the event $\{n \geq N_{\alpha(\nu)}\}$ $S_n^0 = S_n^1 = \cdots = S_n^\nu$ (cf. (5.5)); Moreover on $\{n < N_{\alpha(\nu)}\}$ $S_n^\nu = S_n^{\nu+k} - k$ for $k = 1, 2, \cdots$ (cf. (5.4)). Consequently

$$\begin{aligned}\Gamma_1(m) &= \{\min_{0\leq\nu\leq m} \tau_{\mu m^3}(\nu) < m^3/2\} \\ &= \{S^m_{[m^3/2]} > \mu m^3 \text{ or } S^0_{[m^3/2]} > \mu m^3\},\end{aligned}$$

since among those path $\{S_n^1\}, \cdots \{S_n^m\}$ which have not coalesced with $\{S_n^0\}$ by time $[m^3/2]$, S_n^m has the largest value at $n = [m^3/2]$. It follows that

$$P(\Gamma_1(m) \mid \mathscr{F}) \leq 2P(S^0_{[m^3/2]} > \mu m^3 - m \mid \mathscr{F}). \tag{6.15}$$

By Lemma 6.2 and the Borel–Cantelli Lemma,

$$P(S^0_{[m^3/2]} > \mu m^3 - m \mid \mathscr{F}) \to 0 \quad \text{A.S.} \quad (P_y). \quad \square$$

PROOF OF PROPOSITION 6.1. According to the ergodic theorem (strong law) for Markov chains, the empirical distribution converges to the stationary distribution, w.p.1 (regardless of the initial state). Since there are only finitely many possible initial states, the convergence is uniform over all initial distributions: in particular, for each $\delta > 0$ there exists a $k = k(\delta)$ such that for every $y \in \mathscr{Y}$,

$$P_y\{\textstyle\sum_{y'\in\mathscr{Y}} | k^{-1} \sum_{j=1}^k 1\{Y_j = y'\} - \pi(y') | > \delta\} < \delta. \tag{6.16}$$

To prove (6.1) we will break up the time axis into "blocks" of length k. There will be a lot of "good" blocks, i.e., blocks for which the empirical distribution is close to the stationary distribution; and a small proportion of "bad" blocks, in which the empirical distribution is not so close to the stationary distribution. Let

$$Z_n = 1\{\delta \leq \textstyle\sum_{y'\in\mathscr{Y}} | k^{-1} \sum_{j=(n-1)k+1}^{nk} 1\{Y_j = y'\} - \pi(y') | \leq 2\}: \tag{6.17}$$

then Z_n is the indicator of the event that the nth block is bad. Notice that

$$(6.18)\qquad P_y(Z_n = 1 \mid Y_1, Y_2, \cdots, Y_{(n-1)k+1}) < \delta \quad \forall y \in \mathscr{Y}, \ n \geq 1$$

by (6.16). Notice also that the measure of discrepancy between the empirical distribution and the stationary distribution is the total variation distance, which is never greater than 2.

If $\delta > 0$ is sufficiently small, say $\delta < \varepsilon/5$, and if m is large, then a large (bigger than ε) discrepancy between the empirical distribution of the first mk observations and the stationary distribution can occur only if a large proportion of $Z_1, Z_2, \cdots, Z_m$ assume the value 1. (This is because of the triangle inequality for the total variation distance.) Furthermore, the empirical distribution can change by at most $1/m$ between $n = mk$ and $n = (m+1)k$ (in total variation). Consequently if $1/m < \delta$, *and* $0 \leq n < k$, then

$$(6.19)\qquad \begin{aligned} &\{\textstyle\sum_{y' \in \mathscr{Y}} \mid (mk+n)^{-1} \sum_{j=1}^{mk+1} 1\{Y_j = y'\} - \pi(y') \mid > 5\delta\} \\ &\quad \subset \{\textstyle\sum_{y' \in \mathscr{Y}} \mid (mk)^{-1} \sum_{j=1}^{mk} 1\{Y_j = y'\} - \pi(y') \mid > 4\delta\} \\ &\quad \subset \{m^{-1} \textstyle\sum_{i=1}^{m} Z_i > 3\delta/2\}. \end{aligned}$$

The exponential rate (6.1) therefore follows from (6.19), (6.18), and

LEMMA 6.4. *Suppose $Z_1, Z_2, \cdots$ are 0–1 valued random variables, adapted to $\mathscr{A}_1 \subset \mathscr{A}_2 \subset \cdots$. If*

$$P(Z_{n+1} = 1 \mid \mathscr{A}_n) \leq \rho$$

for all $n = 0, 1, \cdots$ ($\mathscr{A}_0 = \{\varnothing, \Omega\}$) then for each $\varepsilon > 0$ there exists a constant $\lambda(\varepsilon) > 0$ such that

$$P\{m^{-1} \textstyle\sum_{n=1}^{m} Z_n > \rho + \varepsilon\} \leq \exp(-m\lambda(\varepsilon))$$

for all m.

PROOF. By the Markov inequality

$$P\{m^{-1} \textstyle\sum_{n=1}^{m} Z_n > \rho + \varepsilon\} \leq \dfrac{E(1+s)^{\sum_{n=1}^{m} Z_n}}{(1+s)^{m(\rho+\varepsilon)}} \leq \dfrac{(1+s\rho)^m}{(1+s)^{(\rho+\varepsilon)m}}$$

for all $s > 0$, m. For each $\varepsilon > 0$ there exists an $s > 0$ for which

$$1 + s\rho < (1+s)^{\rho+\varepsilon}$$

(calculus!) so the desired inequality holds with $e^{-\lambda(\varepsilon)} = (1+s\rho)/(1+s)^{\rho+\varepsilon}$. □

PROOF OF LEMMA 6.2. For $y' \in \mathscr{Y}$ let

$$M_n(y') = \textstyle\sum_{j=1}^{n} 1\{Y_j = y'\} \quad \text{and} \quad S_n(y') = \textstyle\sum_{j=1}^{n} X_j 1\{Y_j = y'\}.$$

Notice that $S_n(y')$ is (conditional on $\mathscr{F}$) the sum of $M_n(y')$ i.i.d. random variables each with distribution $F_{y'}$. Consequently, by SLLN there exists $m(y')$ sufficiently large that

$$(6.20)\qquad P_y\{\mid S_n(y') - \mu_{y'} M_n(y') \mid > \varepsilon M_n(y') \mid \mathscr{F}\} \cdot 1\{M_n(y') \geq m(y')\} < \varepsilon.$$

Now in order for $|S_n - n\mu| > 2\varepsilon n$ to occur, it must be that either

$$|S_n(y') - \mu_{y'}M_n(y')| > \varepsilon M_n(y') \text{ for some } y' \in \mathscr{Y},$$

or

$$|n\mu - \textstyle\sum_{y'\in\mathscr{Y}} \mu_{y'}M_n(y')| > n\varepsilon.$$

Let n be sufficiently large that $n\pi(y') > 2m(y')$ for every $y' \in \mathscr{Y}$; then by (6.20)

$$\begin{aligned} &P_y\{P(|S_n - n\mu| > n \cdot (2\varepsilon) \mid \mathscr{F}) > 2\varepsilon \mid \mathscr{Y} \mid\} \\ &\quad \le P_y\{|n\mu - \textstyle\sum_{y'\in\mathscr{Y}} \mu_{y'}M_n(y')| > n\varepsilon\} \\ &\qquad + P_y\{|M_n(y') - n\pi(y')| > n\pi(y')/2 \text{ for some } y' \in \mathscr{Y}\}. \end{aligned} \tag{6.21}$$

(Here $|\mathscr{Y}|$ denotes the cardinality of $\mathscr{Y}$). Both probabilities on RHS (6.21) are exponentially decaying in n, by Proposition 6.1; the inequality (6.2) follows. □

PROOF OF LEMMA 6.3. It suffices to consider only (6.5), since (6.6) follows from (6.5) by a trivial application of the Borel–Cantelli Lemma.

We will use Proposition 6.1 and Lemma 6.2 to establish exponential bounds for six separate probabilities:

$$P_y\{P(\tau(n^\rho + n) > n^{\rho+1} \mid \mathscr{F}) > \delta\} \le C\exp(-n\lambda) \tag{6.22}$$

$$P_y\{\max_{0\le a\le n^\rho} P(S_{\tau(a)} - a > \delta n \mid \mathscr{F}) > \delta\} \le C\exp(-n\lambda) \tag{6.23}$$

$$\begin{aligned} &P_y\{\max_{0\le m\le n^{\rho+1}} P(S_{m+[n\mu^{-1}(1-2\delta)]} - S_m > n(1-\delta) \mid \mathscr{F}) > \delta\} \\ &\qquad \le C\exp(-n\lambda) \end{aligned} \tag{6.24}$$

$$P_y\{\max_{0\le m\le n^{\rho+1}} P(S_{m+[n\mu^{-1}(1+\delta)]} - S_m < n \mid \mathscr{F}) > \delta\} \le C\exp(-n\lambda) \tag{6.25}$$

$$\begin{aligned} &P_y\{\max_{0\le m\le n^{\rho+1}} \textstyle\sum_{y'\in\mathscr{Y}} |(n\mu^{-1}(1-2\delta))^{-1} \sum_{1\le j\le n\mu^{-1}(1-2\delta)} 1\{Y_j = y'\} \\ &\qquad - \pi(y')| > \delta\} \le C\exp(-n\lambda) \end{aligned} \tag{6.26}$$

$$\begin{aligned} &P_y\{\max_{0\le m\le n^{\rho+1}} \textstyle\sum_{y'\in\mathscr{Y}} |(n\mu^{-1}(1+\delta))^{-1} \sum_{1\le j\le n\mu^{-1}(1+\delta)} 1\{Y_j = y'\} \\ &\qquad - \pi(y')| > \delta\} \le C\exp(-n\lambda). \end{aligned} \tag{6.27}$$

Here $\tau(a) = \min\{n: S_n > a\}$; the constants C, λ may depend on δ but not n; and the inequalities hold simultaneously for all $y \in \mathscr{Y}$. The exponential rate (6.5) follows from (6.22)–(6.27): it is a routine matter to verify that with $6\delta + \mu\delta((1-2\delta)^{-1} + (1+\delta)) < \varepsilon$,

$$\text{LHS}(6.5) \le \textstyle\sum_{j=22}^{27} \text{LHS}(6.j).$$

PROOFS OF (6.22)–(6.27).

(6.22): Notice that since $S_m \uparrow$,

$$P(\tau(n^\rho + n) > n^{\rho+1} \mid \mathscr{F}) \le P(S_{[n^{\rho+1}]} \le n^\rho + n \mid \mathscr{F}),$$

so (6.22) follows from (6.2).

(6.23): Since $\{S_m\}$ cannot visit any integer more than once, it follows that for any $a \in \mathbb{Z}^+$

$$P(S_{\tau(a)} - a > \delta n \mid \mathscr{F}) \le \sum_{y' \in \mathscr{Y}} \sum_{\ell=0}^{a} F_{y'}((\delta n + a - \ell, \infty))$$

$$\le \sum_{y' \in \mathscr{Y}} \sum_{\ell=0}^{\infty} F_{y'}((\delta n + \ell, \infty)) \to 0 \quad \text{as} \quad n \to \infty,$$

since each $F_{y'}$ has finite first moment.

(6.24)–(6.25): It follows from (6.2) that with $n_0 = [n\mu^{-1}(1 - 2\delta)]$,

$$\text{LHS}(6.24) \le n^{\rho+1} \max_{y' \in \mathscr{Y}} P_{y'}\{P(|S_{n_0} - \mu n_0| > n_0\delta/2 \mid \mathscr{F})\} \le C \exp(-n\lambda)$$

for certain constants $C < \infty$, $\lambda > 0$ ((6.25) is similar).

(6.26)–(6.27): These are almost immediate from Proposition 6.1. □

7. Almost sure convergence: infinite state space. The argument in the preceding section hinged on the exponential estimate of Proposition 6.1 for the rate of convergence of the empirical distribution of the Markov chain. In infinite state spaces such estimates are not generally available; even when they are, almost sure convergence of the renewal measure may fail (cf. Example 1, Section 4). This suggests that almost sure convergence depends not on the degree of "mixing" in the Markov chain alone, but also on the nature of the assignment $y \to F_y$. In this section we shall establish almost sure convergence of the renewal measure under the hypotheses of Theorem 3, to wit, that each $X_n > 0$, and

$$E_y\left(\sum_{n=1}^{N_1-1} \mu_{Y_n}\right)^2 < \infty, \tag{7.1}$$

$$E_y\left(\sum_{n=1}^{N_1-1} \sigma^2_{Y_n}\right) < \infty, \tag{7.2}$$

and

$$E_y N_1^2 < \infty. \tag{7.3}$$

The proof of Theorem 4 is similar (although not identical) and will therefore be omitted.

The basic strategy of the proof is no different from that of Section 6: balance the speed of convergence in the Law of Large Numbers (actually the erogdic theorem for the first-passage process) against the rate $O(n^2)$ of coalescence in the coupling scheme of Section 5. The balance here is considerably more delicate than in the finite state space case, however. The second moment hypotheses (7.1)–(7.3) give *precisely* the right speed of convergence in the Law of Large Numbers to compensate for the $O(n^2)$ rate of coalescence. (Indeed, the examples of Section 4 demonstrate that if (7.2) holds, then (7.1) is a "minimal" moment condition for almost sure convergence in (1.6), in the sense that $E_y(\sum_{n=1}^{N_1-1} \mu_{Y_n})^{2-\varepsilon} < \infty$ is insufficient to guarantee A.S. convergence.) The essential tool for obtaining rates of convergence from (7.1)–(7.3) is the Hsu–Robbins Theorem:

THEOREM (Hsu and Robbins, 1948). *Suppose* $\xi_1, \xi_2, \cdots$ *are i.i.d. with*

$E\xi_i = 0$ and var $\xi_i < \infty$. *Then*

$$\textstyle\sum_{n=1}^{\infty} P\{|\xi_1 + \cdots + \xi_n| > \varepsilon n\} < \infty$$

for every $\varepsilon > 0$.

Throughout this section we will assume that assumptions (6.1)–(6.3) are in force, that all of the distributions $F_{y'}$, $y' \in \mathscr{Y}$ are supported by the set of positive integers $\{1, 2, \cdots\}$, and also that for the given initial point y, the symmetrization of F_y is supported by no proper subgroup of $\mathbb{Z}$ (this last is only so that the coupling construction of Section 5 may be used without modification). As usual, $1 = N_0, N_1, \cdots$ are the times of successive returns to the initial state y; we will sometimes write $N(j)$ instead of N_j, for typographical reasons. The process $\{S_n\}$ will be Markov random walk with $\{Y_n\}$ as its driving process; the processes $\{S_n^\nu\}$, $\nu = 0, 1, \cdots$ will be as in (5.4). Keep in mind that conditional on $\mathscr{F}$ the processes $\{S_n\}$ and $\{S_n^\nu - \nu\}$ are identical in law.

Define first passage indices $\beta(m)$ for the process of accumulated means:

(7.4) $$\beta(m) = \min\{k: \textstyle\sum_{n=1}^{N(k)} \mu_{Y_n} > K_1 m^2 - 2K_2 m\}.$$

Here K_1 and K_2 are large but fixed constants which will be further specified later. Notice, however, that with P_y-probability one,

(7.5) $$\textstyle\sum_{n=1}^{N(k)} \mu_{Y_n} < K_1 m^2 - K_2 m$$

for all sufficiently large m.

PROOF. By the ergodic theorem $\beta(m)/m^2 \to K_1\pi(y)/\mu < \infty$. The increments $\xi_k = \sum_{n=N(k)+1}^{N(k+1)} \mu_{Y_n}$ are i.i.d., with finite second moment. Now if $\xi_1, \xi_2, \cdots$ are i.i.d. with finite second moment, then for any constant $K^* < \infty$

$$\max_{k \le K^* m^2} |\xi_k|/m \to 0$$

almost surely by the Borel–Cantelli Lemma, since $E\xi_1^2 < \infty$ implies

$$\textstyle\sum_k P\{|\xi_k| > \varepsilon k^{1/2}\} < \infty$$

which implies $P\{|\xi_k|/k^{1/2} \le \varepsilon \text{ eventually}\} = 1$. Thus

$$m^{-1} \textstyle\sum_{n=N(\beta(m)-1)+1}^{N(\beta(m))} \mu_{Y_n} \to 0 \quad \text{a.s. } (P_y),$$

which implies (7.5). □

LEMMA 7.1. *Suppose g: $\mathscr{Y} \to [0, \infty)$ is a function for which*

(7.6) $$\bar{g} =_\Delta \textstyle\sum_{y' \in \mathscr{Y}} g(y')\pi(y') < \infty$$

and

(7.7) $$E_y\left(\textstyle\sum_{n=1}^{N_1 - 1} g(Y_n)\right)^2 < \infty.$$

Then for each real $\gamma > 0$ and each integer $K > 0$,

(7.8) $$\max_{n: 0 \le n - N(\beta(m)) \le Km} |(\gamma m)^{-1} \textstyle\sum_{0 \le j < \gamma m} g(Y_{n+j}) - \bar{g}| \to 0$$

almost surely (P_y), as $m \to \infty$.

PROOF. The plan is to partition the time interval $N(\beta(m)) \le n \le N(\beta(m) + Km)$ (notice that $N(\beta(m)) + Km \ge N(\beta(m) + Km))$ into subintervals $N(\beta(m) + (k-1)m(\varepsilon)) \le n \le N(\beta(m) + km(\varepsilon))$ (where $m(\varepsilon) = [\varepsilon m]$ and $\varepsilon > 0$ is a small fixed constant, and $k = 1, 2, \cdots, Km/m(\varepsilon)$), and show that

$$(7.9) \quad M_m(g) =_\Delta \max_{1\le k\le Km/m(\varepsilon)} |\, \pi(y)m(\varepsilon)^{-1} \textstyle\sum_{j=N(\beta(m)+(k-1)m(\varepsilon))}^{N(\beta(m)+km(\varepsilon))-1} g(Y_j) - \bar{g} \,| \to 0$$

almost surely (P_y), as $m \to \infty$. The convergence (7.8) will then be deduced from (7.9).

By (7.6) and (7.7), the random variables

$$\textstyle\sum_{j=N(i)}^{N(i+1)-1} g(Y_j), \quad i = 0, 1, \cdots$$

are i.i.d. with finite second moment; consequently the Hsu–Robbins theorem implies that

$$\textstyle\sum_{r=1}^{\infty} P_y\{\, |\, r^{-1} \sum_{j=N(0)}^{N(r)-1} g(Y_j) - \bar{g}\pi(y)^{-1} \,| > \delta\} < \infty$$

for each $\delta > 0$. Now any given integer $r \ge 1$ appears at most $2/\varepsilon$ times in the sequence $\{m(\varepsilon)\}_{m>1/\varepsilon}$, and $Km/m(\varepsilon) \le 2K\varepsilon^{-1}$ for all $m > 2/\varepsilon$, so

$$\begin{aligned}
&\textstyle\sum_{m>2/\varepsilon} P_y\{M_m(g) > \delta\} \\
&\quad \le \textstyle\sum_{m>2/\varepsilon} (2K\varepsilon^{-1}) P_y\{\, |\, \pi(y)m(\varepsilon)^{-1} \sum_{j=N(\beta(m))}^{N(\beta(m)+m(\varepsilon))-1} g(Y_j) - \bar{g} \,| > \delta\} \\
&\quad \le \textstyle\sum_{r=1}^{\infty} (2K\varepsilon^{-1})(2\varepsilon^{-1}) P_y\{\, |\, r^{-1} \sum_{j=N(0)}^{N(r)-1} g(Y_j) - \bar{g}\pi(y)^{-1} \,| > \pi(y)^{-1}\delta\} < \infty.
\end{aligned}$$

The convergence (7.9) therefore follows from the Borel–Cantelli Lemma.

NOTE. The fact that $\sum_{j=N(\beta(m))}^{N(\beta(m)+m(\varepsilon))-1} g(Y_j)$ has the same distribution as $\sum_{j=N(0)}^{N(m(\varepsilon))-1} g(Y_j)$ follows from the fact that the random times $\{\beta(m)\}$ are stopping times with respect to the filtration $\mathscr{F}_{N(n)} = \sigma(Y_1, Y_2, \cdots, Y_{N(n)})$, $n = 0, 1, \cdots$.

To complete the proof of the lemma we must show that (7.8) follows from (7.9). First apply (7.9) with the function $g \equiv 1$; this gives

$$[N(\beta m + km(\varepsilon)) - N(\beta(m))]m^{-1} \to k\, \varepsilon\, \pi(y)^{-1}$$

uniformly for $k = 1, 2, \cdots, km/m(\varepsilon)$, with P_y-probability one, as $m \to \infty$ (keep in mind that for fixed $\varepsilon > 0$, $k > 0$, $Km/m(\varepsilon)$ is bounded as $m \to \infty$). It now follows that for each n, $N(\beta(m)) \le n \le N(\beta(m)) + Km$, the number of points $\{N(\beta m + km(\varepsilon)) \colon k = 0, 1, \cdots, Km/m(\varepsilon)\}$ which lie in the interval $[n, n + \gamma m)$ differs from $\gamma\pi(y)\varepsilon^{-1}$ by no more than 2, for m large (with P_y-probability one). Consequently for each g satisfying (7.6) and (7.7), the result (7.9) implies

$$\begin{aligned}
&\limsup_{m\to\infty} \max_{n: 0\le n-N(\beta(m))\le Km} |\, (\gamma m)^{-1} \textstyle\sum_{0\le j<\gamma m} g(Y_{j+n}) - \bar{g} \,| \\
&\qquad\qquad \le 2\bar{g}/(\gamma\pi(y)\varepsilon^{-1} - 2) \quad \text{a.s.} \quad (P_y).
\end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, (7.8) follows. □

LEMMA 7.2. *For each real $\gamma > 0$, $\delta > 0$, and each integer $K > 0$,*

$$(7.10) \quad P(\max_{n: 0\le n-N(\beta(m))\le Km} |\, (\gamma m)^{-1} \textstyle\sum_{0\le j<\gamma m} X_{n+j} - \mu \,| > \delta \mid \mathscr{F}) \to 0$$

almost surely (P_y), as $m \to \infty$.

PROOF. Recall that conditional on $\mathscr{F}$ the increments X_n are independent with distributions F_{Y_n}, for $n = 1, 2, \cdots$. Consequently by Chebyshev's inequality

$$P(|(\gamma_m)^{-1} \sum_{0\le j<\gamma m} X_{n+j} - \mu| > \delta \mid \mathscr{F})$$
$$\le 1\{|(\gamma m)^{-1} \sum_{0\le j<\gamma m} \mu_{Y_{n+j}} - \mu| > \delta/2\} + (\gamma m)^{-2} \sum_{0\le j<\gamma m} \sigma^2_{Y_{n+j}} \cdot (4/\delta^2) \quad \text{a.s. } (P_y),$$

for each integer $n \ge 1$. Using (7.8) first for $g(y) = \mu_y$ and then for $g(y) = \sigma_y^2$, we conclude that

$$\max_{n:0\le n-N(\beta(m))\le Km} P(|(\gamma m)^{-1} \sum_{0\le j<\gamma m} X_{n+j} - \mu| > \delta \mid \mathscr{F}) \to 0 \tag{7.11}$$

almost surely (P_y), as $m \to \infty$.

The deduction of (7.10) from (7.11) is similar to the argument that showed (7.8) follows from (7.9). Choose $\varepsilon > 0$ small and break up the interval $[N(\beta(m)), N(\beta(m)) + (K + \gamma)m]$ into subintervals of approximate length εm: specifically, let

$$n_j(m) = N(\beta(m)) + jm(\varepsilon), \quad j = 1, 2, \cdots, \frac{[(K+\gamma)m]}{m(\varepsilon)} + 1$$

where $m(\varepsilon) = [\varepsilon m]$. Since $[(K + \gamma)m]/m(\varepsilon) \le 2(K + \gamma)\varepsilon^{-1}$ for all $m \ge 2\varepsilon^{-1}$, it follows from (7.11) that

$$P(\max_{1\le j\le[(K+\gamma)m]/(m(\varepsilon))+1} |(\varepsilon m)^{-1} \textstyle\sum_{i=0}^{m(\varepsilon)} X_{n_j(m)-i} - \mu| > \delta \mid \mathscr{F}) \to 0 \tag{7.12}$$

almost surely (P_y), as $m \to \infty$. Now for any integer n such that $0 \le n - N(\beta(m)) \le Km$, the number of points $n_j(m)$ $(j = 1, 2, \cdots)$ which lie between n and $n + \gamma m$ differs from γ/ε by at most 2, for all m sufficiently large. Averaging first over those intervals $[n_j(m), n_{j+1}(m))$ which intersect $[n, n + \gamma m)$, second over those which are entirely contained in $[n, n + \gamma m)$, and performing an obvious bracketing maneuver, we find that

$$\{\max_{n:0\le n-N(\beta(m))\le Km} |(\gamma m)^{-1} \textstyle\sum_{0\le i<\gamma m} X_{n+i} - \mu| > 2\delta\}$$
$$\subset \{\max_{1\le j\le[(K+\gamma)m]/m(\varepsilon)+1} |(\varepsilon m)^{-1} \textstyle\sum_{i=0}^{m(\varepsilon)} X_{n_j(m)+i} - \mu| > \delta\},$$

provided $\varepsilon > 0$ is sufficiently small. Therefore (7.10) follows from (7.12). □

LEMMA 7.3. *For each subset A of the state space $\mathscr{Y}$ and each integer $K_1 > 0$,*

$$\max_{K_1 m^2\le a\le K_1(m+1)^2} |E(m^{-1} \textstyle\sum_{n\ge1} 1\{a - m \le S_n \le a;\ Y_n \in A\} \mid \mathscr{F}) - \mu^{-1}\pi(A)| \to 0 \tag{7.13}$$

almost surely (P_y), as $m \to \infty$.

PROOF. Choose K_2 large, and let $\beta(m)$ be defined by (7.4). By Chebyshev's inequality, the ergodic theorem, and (7.5),

$$P(|S_{N(\beta(m))} - \textstyle\sum_{n=1}^{N(\beta(m))} \mu_{Y_n}| \ge K_2 m \mid \mathscr{F}) \le (\textstyle\sum_{n=1}^{N(\beta(m))} \sigma^2_{Y_n})(K_2 m)^{-2} \to K_1\mu^{-1}\sigma^2 K_2^{-2} \quad \text{a.s. } (P_y) \text{ as } m \to \infty,$$

where $\sigma^2 = \sum_{y' \in \mathscr{Y}} \sigma_{y'}^2 \pi(y')$. It therefore follows from (7.5) that

$$P(K_1 m^2 - 3K_2 m < S_{N(\beta m)} < K_1 m^2 \mid \mathscr{F}) \geq 1 - \varepsilon \quad \text{a.s.} \quad (P_y) \tag{7.14}$$

for all sufficiently large m, provided $K_1 \mu^{-1} \sigma^2 K_2^{-2} < \varepsilon$.

Define events $\Gamma_i(m)$, $i = 1, \cdots, 4$, by

$$\Gamma_1(m) = \{K_1 m^2 - 3K_2 m < S_{N(\beta(m))} < K_1 m^2\},$$

$$\Gamma_2(m) = \{\max_{n: 0 \leq n - N(\beta(m)) \leq Km} | m^{-2} \textstyle\sum_{0 \leq j < m\mu^{-1}\varepsilon} X_{n+j} - \varepsilon | < \varepsilon\},$$

$$\Gamma_3(m) = \{\max_{n: 0 \leq n - N(\beta(m)) \leq Km} | m^{-1} \textstyle\sum_{0 \leq j < m\mu^{-1}(1-\varepsilon)} X_{n+j} - (1 - \varepsilon) | < \varepsilon\},$$

and

$$\Gamma_4(m) = \{\max_{n: 0 \leq n - N(\beta(m)) \leq Km} | m^{-1} \textstyle\sum_{0 \leq j < m\mu^{-1}(1+\varepsilon)} X_{n+j} - (1 + \varepsilon) | < \varepsilon\},$$

where $K > 2K_1 m + 3K_2 M$. Since each increment of S_n is at least one, the number of n for which $K_1 m^2 - 3K_2 m < n \leq K_1(m + 1)^2$ is no more than Km; moreover, on $\Gamma_1(m)$ all of these n lie between $N(\beta(m))$ and $N(\beta(m)) + Km$. On $\Gamma_1(m) \cap \Gamma_2(m)$,

$$S_{\tau(a)} - a \leq 2\varepsilon m$$

for every a, $K_1 m^2 \leq a < K_1(m + 1)^2$, where $\tau(a) = \min\{n: S_n > a\}$. Consequently, on $\bigcap_{i=1}^4 \Gamma_i(m)$,

$$| m^{-1} \textstyle\sum_{n \geq 1} 1\{a - m \leq S_n \leq a\} - \mu^{-1} | \leq 2\mu^{-1}\varepsilon, \tag{7.15}$$

for $K_1 m^2 \leq a \leq K_1(m + 1)^2$.

Now from Lemma 7.2 and inequality (7.14) it follows that for all sufficiently large m,

$$P(\textstyle\bigcap_{i=1}^4 \Gamma_i(m) \mid \mathscr{F}) \geq 1 - 2\varepsilon \quad \text{a.s.} \quad (P_y). \tag{7.16}$$

Since $0 \leq m^{-1} \sum_{n \geq 1} 1\{a - m \leq S_n \leq a\} \leq (m + 1)/m$, (7.15) and (7.16) together imply that a.s. (P_y),

$$\max_{K_1 m^2 \leq a \leq K_1(m+1)^2} | E(m^{-1} \textstyle\sum_{n \geq 1} 1\{a - m \leq S_n \leq a\} \mid \mathscr{F}) - \mu^{-1} | \leq (2\mu^{-1} + 4)\varepsilon$$

for all m sufficiently large. Since $\varepsilon > 0$ is arbitrary, (7.13) follows in the special case $A = \mathscr{Y}$.

The proof of (7.13) for arbitrary $A \subset \mathscr{Y}$ uses (7.16) and Lemma 7.1 with $g = 1_A$. According to Lemma 7.1,

$$m^{-1} \textstyle\sum_{0 \leq j < m\mu^{-1}(1-\varepsilon)} 1_A(Y_{n+j}) \geq \mu^{-1}\pi(A) - 2\varepsilon$$

and

$$m^{-1} \textstyle\sum_{0 \leq j < m\mu^{-1}(1+\varepsilon)} 1_A(Y_{n+j}) \leq \mu^{-1}\pi(A) + 2\varepsilon$$

for all n, $N(\beta(m)) \leq n \leq N(\beta(m)) + Km$, provided m is sufficiently large. Therefore, on $\bigcap_{i=1}^4 \Gamma_i(m)$,

$$| m^{-1} \textstyle\sum_{n \geq 1} 1\{a - m \leq S_n \leq a;\ Y_n \in A\} - \mu^{-1}\pi(A) | \leq 2\varepsilon, \quad \forall a,\ K_1 m^2 \leq a \leq K_1(m + 1)^2. \tag{7.17}$$

Combining (7.16) and (7.17), we obtain a.s. (P_y),

$$\max_{K_1 m^2 \le a \le K_1(m+1)^2} | E(m^{-1} \sum_{n\ge 1} 1\{a - m \le S_n \le a;\ Y_n \in A\} \mid \mathscr{F}) - \mu^{-1}\pi(A) | \le 6\varepsilon$$

for all m sufficiently large. Since $\varepsilon > 0$ is arbitrary, (7.13) follows. □

Proof of Theorem 3. As in the proof of Theorem 2 we will only consider the convergence of the (conditional) renewal measure. Again we will assume that the distribution F_y associated with the initial state y has the property that its symmetrization is supported by no proper subgroup of the integers $\mathbb{Z}$; thus the coupling results of Section 5 are applicable.

As in the proofs of Theorems 1 and 2 the first step is to use the coupling construction (specifically, the properties (5.5) and (5.6)) to conclude

$$\begin{aligned} &| E(\textstyle\sum_{n\ge 1} 1\{S_n^0 = a;\ Y_n \in A\} \mid \mathscr{F}) \\ &\quad - E((m+1)^{-1} \textstyle\sum_{n\ge 1} 1\{a - m \le S_n^0 \le a;\ Y_n \in A\} \mid \mathscr{F}) | \\ &\qquad \le 1 - P(\tilde{F}_a(m) \mid \mathscr{F}), \end{aligned} \tag{7.18}$$

where

$$\tilde{F}_a(m) = \{N(\alpha(m)) < \tau_{a-1}(\nu) \text{ for each } \nu = 0, 1, \cdots, m\}, \tag{7.19}$$

and

$$\tau_a(\nu) = \min\{n: S_n^\nu > a\} \tag{7.20}$$

($\alpha(m)$ is defined by (5.4)). This holds for all values of m and a in $\mathbb{Z}$.

Define events $\Gamma_1(m)$ and $\Gamma_2(m)$ by

$$\Gamma_1(m) = \{\min_{0\le\nu\le m} \tau_{K_1 m^2 - 1}(\nu) \le K m^2\}, \quad \text{and} \quad \Gamma_2(m) = \{N(\alpha(m)) \ge K m^2\};$$

then

$$1 - P(\tilde{F}_a(m) \mid \mathscr{F}) \le P(\Gamma_1(m) \mid \mathscr{F}) + P(\Gamma_2(m) \mid \mathscr{F}) \tag{7.21}$$

for $K_1 m^2 \le a \le K_1(m+1)^2$. We will show that if $K_1 \gg K \gg 0$, then each of the last two conditional probabilities is small.

First notice that for any $\varepsilon > 0$ there is a $K^* > 0$ sufficiently large that with P_y-probability one,

$$P(\alpha(m) \ge K^* m^2 \mid \mathscr{F}) < \varepsilon$$

for all m large: this follows from Proposition 5.1. Now $\alpha(m) \to \infty$ as $m \to \infty$, so the ergodic theorem implies that

$$\frac{N(\alpha(m))}{\alpha(m)} \to \frac{1}{\pi(y)}$$

almost surely (P_y). Consequently if $K = 2K^*\pi(y)^{-1}$, then

$$P(N(\alpha(m)) \ge K m^2 \mid \mathscr{F}) = P(\Gamma_2(m) \mid \mathscr{F}) < \varepsilon \tag{7.22}$$

for all sufficiently large m, with P_y-probability one.

Next, recall that on $\{n \geq N(\alpha(\nu))\}$, $S_n^0 = S_n^1 = \cdots = S_n^\nu$ (*cf.* (5.5)); moreover on $\{n < N(\alpha(\nu))\}$, $S_n^\nu = S_n^{\nu+k} - k$ for $k \geq 1$ (cf. (5.4)). Therefore

$$\Gamma_1(m) = \{\max(S_{Km^2}^0, S_{Km^2}^m) \geq K_1 m^2\}$$

(cf. the proof of (6.13)); it follows that

$$P(\Gamma_1(m) \mid \mathscr{F}) \leq 2P(S_{Km^2}^0 \geq K_1 m^2 - m \mid \mathscr{F}) < \varepsilon \tag{7.23}$$

for all large m, a.s. (P_y), provided K_1/K is sufficiently large (by Chebyshev's Inequality and the ergodic theorem).

Combining (7.18), (7.21), (7.22), (7.23), and Lemma 7.3, we obtain

$$\limsup_{m\to\infty} \max_{K_1 m^2 \leq a \leq K_1(m+1)^2} | E(\textstyle\sum_{n\geq 1} 1\{S_n^0 = a;\ Y_n \in A\} \mid \mathscr{F}) - \mu^{-1}\pi(A) | \leq 2\varepsilon \quad \text{a.s.} \quad (P_y).$$

Since $\varepsilon > 0$ is arbitrary, it follows that

$$\lim_{a\to\infty} E(\textstyle\sum_{n\geq 1} 1\{S_n = a;\ Y_n \in A\} \mid \mathscr{F}) = \mu^{-1}\pi(A)$$

with P_y probability one. □

8. Concluding remarks. The semi-Markov process is but one of many stochastic processes in which two or more distinct "random mechanisms" are at work. Queueing models typically incorporate separate arrival and service processes; models of population growth often postulate a "random environment" within which a random evolution occurs; statistical models of survival studies commonly include separate arrival, censoring, and response processes. Often, it would seem, one would like to know whether limit theorems for such processes hold conditional on the output of one (or more) of the random mechanisms. In a typical problem involving a "random environment" one would imagine the environment as "fixed," having been hammered out by a dice-throwing Hephaestus at the beginning of time: it is in this *particular* environment, rather than across the ensemble of all possible environments, that one would like to understand the behavior of an evolving population. In problems of statistical inference (such as survival studies) one should (according to R. A. Fisher) generally try to make one's inference conditional on the observed values of the "ancillary" statistics (such as the arrival times).

The coupling techniques we have introduced in this paper may be of use in establishing conditional limit theorems even when Theorems 1–4 are irrelevant. We hope to demonstrate this in forthcoming articles. But we also hope that the reader will appreciate that embedded Markov random walks are as common as E. Coli, and hence that our theorems may be of broad applicability.

As an example, consider the discrete time $G \mid G \mid 1$ queue. Tasks arrive at a server at times $A_n = \alpha_1 + \cdots + \alpha_n$, where the interarrival times $\{\alpha_j\}$ are i.i.d. from a distribution F on the positive integers. To each task is attached a service time: assume that the service times $\{\xi_n\}$ are i.i.d. from a distribution G on the positive integers, and are independent of the interarrival times. Let Z_t be the number of tasks in the queue at time t. It is well-known that if $\mu_F > \mu_G$ (here μ_F and μ_G are the means of F and G respectively) and if supp F, supp G both generate

$\mathbb{Z}$, then the queue approaches an equilibrium, i.e.,

$$Z_t \to_{\mathscr{D}} H \quad \text{as} \quad t \to \infty \quad \text{through} \quad \mathbb{Z}$$

for some distribution H on the nonnegative integers.

Let

$$\mathscr{F}^{\alpha} = \sigma(\alpha_1, \alpha_2, \cdots) \quad \text{and} \quad \mathscr{F}^{\xi} = \sigma(\xi_1, \xi_2, \cdots).$$

A moment's reflection will reveal that conditional on $\mathscr{F}^{\alpha}$, Z_t does not generally converge in law as $t \to \infty$: in eras when arrivals are very sparse, the queue will tend to empty out. However,

THEOREM. *Suppose F and G have exponentially decaying tails, that $\mu_F > \mu_G$, and that both* supp(F) *and* supp(G) *generate the group $\mathbb{Z}$. Then*

$$\mathscr{L}(Z_t \mid \mathscr{F}^{\xi}) \to H \quad a.s.$$

as $t \to \infty$. Suppose $T(t)$ is the accumulated service time up to time t, i.e.,

$$T(t) = \sum_{j=1}^{N(t)+1} \xi_j \quad \textit{where} \quad N(t) = \#\{\textit{tasks completed by time } t\},$$

and $\tau(t) = \inf\{s\colon T(s) \ge t\}$. Then there exists a distribution $\tilde{H}$ on the nonnegative integers such that

$$\mathscr{L}(Z_{\tau(t)} \mid \mathscr{F}^{\alpha}) \to \tilde{H} \quad a.s.\ as \quad t \to \infty.$$

To prove this one need only identify appropriate embedded Markov random walks and apply Theorem 3 (the hypotheses on the tails of F and G will make verification of (1.8)–(1.10) routine; they could be weakened). For the first limit, let Y_n^0 consist of the number of tasks in the queue, their service times, and the time since the last arrival at the time the nth task is completed; let $Y_n = (Y_n^0, Y_{n+1}^0)$; and let X_n be the amount of time which passes between the time the nth task is completed and the time the $(n+1)$st task is completed. For the second limit, let Y_n^0 consist of the number of tasks in the queue, together with the amount of service time remaining on the task being processed at the time of the nth arrival; let $Y_n = (Y_n^0, Y_{n+1}^0)$; and let X_n be the amount of service time which accumulates between the nth and $(n+1)$st arrival. The details of the argument are left to the reader.

REFERENCES

ATHREYA, K., MCDONALD, D. and NEY, P. (1978a). Limit theorems for semi-Markov processes and renewal theory for Markov chains. *Ann. Probab.* **6** 788–797.

ATHREYA, K., MCDONALD, D. and NEY, P. (1978b). Coupling and the renewal theorem. *Amer. Math. Monthly* **85** 809–814.

CHUNG, K. L. (1961). *Markov Chains with Stationary Transition Probabilities.* Springer-Verlag, Heidelberg.

CHUNG, K. L. and ORNSTEIN, D. (1962). On the recurrence of sums of random variables. *Bull. Amer. Math. Soc.* **68** 30–32.

CINLAR, E. (1969). On semi-Markov processes on arbitrary state spaces. *Proc. Cambridge Phil. Soc.* **66** 381–392.

CINLAR, E. (1974). Periodicity in Markov renewal theory. *Adv. Appl. Probab.* **6** 61–78.

DERMAN, C. (1954). A solution to a set of fundamental equations in Markov chains. *Proc. Amer. Math. Soc.* **5** 332–334.

DOEBLIN, W. (1938). Exposé de la théorie des chaînes simples constantes de Markov à un nombre fini d'ētats. *Rev. Math. de l'Union Interbalkanique* **2** 77–105.

DONSKER, M. and VARADHAN, S. R. S. (1975). Asymptotic evaluation of certain Markov process expectations for large time—I. *Comm. Pure Appl. Math.* **27** 1–47.

DONSKER, M. and VARADHAN, S. R. S. (1976). Asymptotic evaluation of certain Markov process expectations for large time—III. *Comm. Pure Appl. Math.* **29** 389–461.

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

HSU, P. L. and ROBBINS, H. (1948). Complete convergence and the law of large numbers. *Proc. Nat. Acad. Sci. USA* **33** 25–31.

JACOD, J. (1971). Théorème de renouvellement et classification pour les chains semi-Markoviennes. *Ann. Inst. H. Poincaŕe Sect. B* **7** 83–129.

KESTEN, H. (1974). Renewal theory for Markov chains. *Ann. Probab.* **3** 355–387.

KESTEN, H. and SPITZER, F. (1963). Ratio theorems for random walk. *J. Analyse Math.* **11** 285–322.

LÉVY, P. (1954). Processes semi-Markoviens. *Proc. Int. Cong. Math.* **3** 416–426.

LINDVALL, T. (1977). A probabilistic proof of Blackwell's theorem. *Ann. Probab.* **5** 482–485.

MCDONALD, D. (1975). Renewal theorem and Markov chains. *Ann. Inst. H. Poincaŕe Sect B* **11** 187–197.

MCDONALD, D. (1978). On semi-Markov and semi-regenerative processes I. *Z. Warsch. verw. Gebiete* **42** 261–277.

NEY, P. (1981). A refinement of the coupling method in renewal theory. *Stochastic Process. Appl.* **11** 11–26.

OREY, S. (1959). Recurrent Markov chains. *Pacific J. Math.* **9** 805–827.

OREY, S. (1961). Change of time scale for Markov processes. *Trans. Amer. Math. Soc.* **99** 384–390.

ORNSTEIN, D. (1969). Random walk I. *Trans. Amer. Math. Soc.* **138** 1–43.

PORT, S. and STONE, C. (1969). Potential theory of random walks on Abelian groups. *Acta. Math.* **122** 19–114.

PYKE, R. (1961). Markov renewal processes with finitely many states. *Ann. Math. Statist.* **32** 1231–1242.

SMITH, W. (1955). Regenerative stochastic processes. *Proc. Royal Statist. Soc. A* **232** 6–31.

SMITH, W. (1961). On some general renewal theorems for nonidentically distributed random walks. *Proc. 4th Berk. Symp. II* 467–514.

SPITZER, F. (1976). *Principles of Random Walk,* 2nd ed. Springer-Verlag, Heidelberg.

DEPARTMENT OF STATISTICS
COLUMBIA UNIVERSITY
NEW YORK, NEW YORK 10027