Z. Wahrscheinlichkeitstheorie verw. Gebiete 29, 193–227 (1974)


# Uniform Rates of Convergence for Markov Chain Transition Probabilities

J. W. Pitman

## 1. Introduction and Statement of Results

Let $P$ be a Markov matrix indexed by a countable set of states $J$, and denote the $n$-th iterate of $P$ by

$$(1.1)\qquad P^n=(p_{ij}^{(n)}),\quad i,j\in J,\ n=0,1,2,\ldots.$$

If $P$ is irreducible and aperiodic then according to a well known theorem of Kolmogorov [14]

$$(1.2)\qquad \lim_{n\to\infty} p_{ij}^{(n)}=1/m_{jj},$$

where $m_{jj}$ denotes the expected recurrence time of state $j$, and $1/m_{jj}$ is taken to be zero in the transient and null recurrent cases when $m_{jj}=\infty$. In his book [1] Breiman gives an elegant proof of this theorem in the positive recurrent case which is based on the simple probabilistic device of comparing the progress of two independent Markov chains with the same transition probabilities $P$ but different initial distributions. It is shown here that this device can in fact be used to cover the null recurrent case too, and more importantly that in the positive recurrent case it is possible to further exploit the idea to obtain a number of new and powerful refinements of the main limit theorem. The central results will shortly be stated in Theorems 1 and 2 below. These theorems provide new results on the rate of convergence in (1.2) for Markov chains with infinite state space. The results generalise the work of Feller [5] and Karlin [9] on the rate of convergence of renewal sequences, as well as extending results from the potential theory of positive recurrent chains due to Kemeny, Snell and Knapp [10].

To state the results we first require some notation. Let $\lambda=(\lambda_j)$ be a probability distribution on $J$, to be thought of as an initial distribution, and set

$$p_{\lambda j}^{(n)}=(\lambda P^n)_j=\sum_{i\in J}\lambda_i\, p_{ij}^{(n)}.$$

Here, as everywhere in the sequel, it is tacitly assumed that $j$ ranges over the state space $J$ and $n$ ranges over the set $\mathbb{N}$ of non-negative integers. For the initial distribution $\delta_i$ on $J$ which attributes probability one to the single state $i$ in $J$ we shall always write simply $i$ instead of $\delta_i$ in subscripts. Let us suppose that on a suitable probability space $(\Omega,\mathscr{F},\mathbb{P}_\lambda)$ a Markov chain $(X_n)$ has been constructed with state space $J$, initial distribution $\lambda$ and stationary transition probabilities $P$. We then have that

$$p_{\lambda j}^{(n)}=\mathbb{P}_\lambda(X_n=j),$$

so that $\lambda P^n$ is the distribution of $X_n$.

A bounded signed measure $\nu$ on the power set of $J$ will be referred to simply as a *signed measure on* $J$, such a signed measure being determined by its values $\nu_j=\nu\{j\}$ on single point sets $\{j\}\subset J$. The total variation norm of $\nu$ is denoted by $\|\nu\|$:

$$\|\nu\|=\sum_{j\in J}|\nu_j|.$$

We shall only consider the norms of signed measures $\nu$ on $J$ with total mass $\nu(J)$ equal to zero (e.g. differences between probabilities), and then also

$$\|\nu\|=2\sup_{H\subset J}|\nu(H)|. \tag{1.3}$$

For the rest of the introduction let us suppose that our given Markov matrix $P$ is irreducible, aperiodic and positive recurrent. By virtue of dominated convergence and Scheffé's theorem [18] the limit theorem (1.2) can be restated for this case in the following way: for all initial probability distributions $\lambda$ on $J$,

$$\lim_{n\to\infty}\|\lambda P^n-\pi\|=0, \tag{1.4}$$

where $\pi=(\pi_j)$ is the (unique) invariant probability measure for $P$ given by $\pi_j=1/m_{jj}$. From (1.4) and the triangle inequality for $\|\ \|$ we have also that for any pair of initial probabilities $\lambda$ and $\mu$,

$$\lim_{n\to\infty}\|\lambda P^n-\mu P^n\|=0, \tag{1.5}$$

(1.4) being the special case of (1.5) when $\mu=\pi$. It may be noted that the sequences of norms appearing in (1.4) and (1.5) in fact decrease monotonically to zero, since

$$\lambda P^n-\mu P^n=(\lambda-\mu)P^n$$

and $P$ is norm decreasing on the space of signed measures with total variation norm. It is the rate at which these norms decrease to zero which is the main concern of the two theorems stated below. Also considered in these theorems are the measures $\lambda\sum_{k=1}^{n}P^k$ associated with initial distributions $\lambda$ on $J$. Recalling our Markov chain $(X_n)$ defined on $(\Omega,\mathscr{F},\mathbb{P}_\lambda)$ with initial distribution $\lambda$ and transition probabilities $P$, we have that just as $\lambda P^n$ is the distribution of $X_n$, so $\lambda\sum_{k=1}^{n}P^k$ is the "expected occupation time measure" which attributes to a subset $H$ of $J$ the expected number of times $k$ with $1\leq k\leq n$ that $X_k\in H$, that is to say

$$\mathbb{E}_\lambda\sum_{k=1}^{n}1_H(X_n),$$

where $\mathbb{E}_\lambda$ denotes expectation with respect to $\mathbb{P}_\lambda$ and $1_H$ is the indicator function of the subset $H$ of $J$.

Let $T_j$ denote the first passage time of $(X_n)$ to the state $j$ in $J$:

$$T_j=\inf\{n\colon n>0,\ X_n=j\},$$

and set

$$m_{\lambda j}=\mathbb{E}_\lambda T_j=\sum_{i\in J}\lambda_i m_{ij}. \tag{1.6}$$

We are supposing that $P$ is irreducible and positive recurrent, and thus for all $i$ and $j$ in $J$ we have

$$m_{ij} < \infty \tag{1.7}$$

(see Freedman [8], 1.8). In Theorem 1 we shall be imposing the condition on an initial distribution $\lambda$ that $m_{\lambda j}$ be finite for all $j$ in $J$. It is an easy consequence of (1.7) that this is in fact the case as soon as $m_{\lambda j}$ is finite for some $j$ in $J$, and in particular the condition is satisfied whenever $\lambda$ has finite support.

**Theorem 1.** *Suppose that $P$ is an irreducible, aperiodic and positive recurrent Markov matrix with countable state space $J$, and let $\lambda$ and $\mu$ be two initial distributions on $J$ such that for all $j$ in $J$ both $m_{\lambda j}$ and $m_{\mu j}$ are finite. Then both*

$$\lim_{n\to\infty} n \|\lambda P^n - \mu P^n\| = 0, \tag{1.8}$$

*and*

$$\sum_{n=1}^{\infty} \|\lambda P^n - \mu P^n\| < \infty; \tag{1.9}$$

*furthermore there is a signed measure $v$ on $J$ with total mass zero such that*

$$\lim_{n\to\infty} \left\| \left[ \lambda \sum_{k=1}^{n} P^k - \mu \sum_{k=1}^{n} P^k \right] - v \right\| = 0, \tag{1.10}$$

*and $v$ is given by*

$$v_j = (m_{\mu j} - m_{\lambda j})/m_{jj}, \qquad j \in J. \tag{1.11}$$

Proof of this theorem is the subject of Section 3. To illustrate the strength of the assertions of convergence in norm it is perhaps worth spelling out in detail some of the statements of Theorem 1. Thus according to (1.8) we have

$$\lim_{n\to\infty} n \sum_{j\in J} |p_{\lambda j}^{(n)} - p_{\mu j}^{(n)}| = 0, \tag{1.12}$$

or, using (1.3), for every subset $H$ of $J$

$$\lim_{n\to\infty} n |\lambda P^n(H) - \mu P^n(H)| = 0,$$

the convergence being uniform over all subsets. Naturally similar amplifications can be made corresponding to (1.9) and (1.10), and for that matter any of the other assertions of convergence in norm which follow later. Elaborating on the statements (1.9) and (1.10) we see that if $m_{\lambda j}$ and $m_{\mu j}$ are both finite, $j \in J$, then

$$\sum_{n=1}^{\infty} (p_{\lambda j}^{(n)} - p_{\mu j}^{(n)}) = (m_{\mu j} - m_{\lambda j})/m_{jj}, \qquad j \in J, \tag{1.13}$$

and the series are uniformly absolutely convergent over $j$ in $J$, indeed

$$\sum_{j\in J} \sum_{n=1}^{\infty} |p_{\lambda j}^{(n)} - p_{\mu j}^{(n)}| < \infty. \tag{1.14}$$

Apart from the assertions of uniform and absolute convergence the result (1.13) can be found in Kemeny, Snell and Knapp [10] where it is also shown that

$\sum_{j\in J}|v_j|$ is finite, a fact implicit here in the statement that $v$ is a bounded signed measure. However as far as I am aware both (1.12) and the assertion that the series in (1.13) are absolutely convergent are new even in the simplest case with $\lambda=\delta_i$ and $\mu=\delta_k$ for particular states $i$ and $k$ in $J$, when the results invariably hold and reduce to statements purely concerning the limit behaviour of the $n$-step transition probabilities $p_{ij}^{(n)}$ and $p_{kj}^{(n)}$, $j\in J$ (cf. (5.14) below).

Obviously a most important special case of Theorem 1 will arise on making the simplifying substitution of the invariant probability $\pi$ for the second initial distribution $\mu$, but the statement of both this and another important corollary of Theorem 1 are being left until after the statement of the more general Theorem 2 below (see the special cases $r=2$ of Corollary 1 and $r=1$ of Corollary 2).

For $r=1,2,\ldots$ let $m_{\lambda j}^{(r)}$ denote the $r$-th moment of the first passage time to $j$ for initial distribution $\lambda$:

$$m_{\lambda j}^{(r)}=\mathbb{E}_\lambda(T_j)^r=\sum_{i\in J}\lambda_i\, m_{ij}^{(r)}.$$

If $m_{jj}^{(r)}$ is finite for some $j$ in $J$ then $m_{jj}^{(r)}$ must be finite for all $j$ in $J$ and we say that $P$ has recurrence times with finite $r$-th moment (see Freedman [8], 2.9). In this case

$$m_{ij}^{(r)}<\infty \tag{1.15}$$

for all $i$ and $j$ in $J$, and for a general initial distribution $\lambda$ on $J$ either $m_{\lambda j}^{(r)}$ is finite for all $j$ in $J$ or $m_{\lambda j}^{(r)}$ is finite for no $j$ in $J$. In particular the former case obtains if $\lambda$ has finite support. Theorem 1 is the special case $r=1$ of the following more general theorem:

**Theorem 2.** *Let $r$ be a positive integer and suppose that $P$ is an irreducible and aperiodic Markov matrix with countable state space $J$ and recurrence times with finite $r$-th moment. Let $\lambda$ and $\mu$ be two initial distributions on $J$ such that both $m_{\lambda j}^{(r)}$ and $m_{\mu j}^{(r)}$ are finite for $j$ in $J$. Then*

$$\lim_{n\to\infty} n^r\,\|\lambda P^n-\mu P^n\|=0, \tag{1.16}$$

$$\sum_{n=1}^{\infty} n^{r-1}\,\|\lambda P^n-\mu P^n\|<\infty, \tag{1.17}$$

*and*

$$\lim_{n\to\infty} n^{r-1}\left\|\left[\lambda\sum_{k=1}^{n}P^k-\mu\sum_{k=1}^{n}P^k\right]-v\right\|=0, \tag{1.18}$$

*where $v$ is the signed measure on $J$ given by* (1.11).

This theorem is proved in Section 4. There is a useful alternative way of expressing the condition that $m_{\lambda j}^{(r)}$ be finite. Fix $j$ in $J$, let ${}^jP$ denote the sub-stochastic matrix obtained from $P$ by replacing the $j$-th column by zeros, and define a matrix ${}^jU$ by

$${}^jU=\sum_{n=0}^{\infty}({}^jP)^n, \tag{1.19}$$

so that ${}^jU$ is the matrix with $(i,k)$-th entry equal to

$$\mathbb{E}_i\,[\text{Number of times in state } k \text{ before first passage to } j],$$

which is of course finite under our assumption that $P$ is irreducible. It is a consequence of Proposition (4.6) in this paper that for any positive integer $r$ and any initial distribution $\lambda$ an $J$

$$\mathbb{E}_\lambda T_j(T_j+1)\dots(T_j+r-1)=r!\,\lambda({}^jU)^r 1, \tag{1.20}$$

where on the right hand side 1 denotes the function on $J$ which takes the constant value one. Thus $m_{\lambda j}^{(r)}$ is finite if and only if $\lambda({}^jU)^r 1$ is finite, and this reduces the condition to one which involves only the first moments of first passage times, since with our implicit assumption that $P$ is positive recurrent we have

$${}^jU_{ik}=(m_{ij}+m_{jk}-m_{ik})/m_{kk}+\delta_{ik}-\delta_{jk},$$

where $\delta_{gh}=1$ if $g=h$, 0 otherwise (cf. Chung [3], I.11). Considering now the invariant probability $\pi$ for $P$, it is well known and easy to prove that

$$\delta_j({}^jU)=m_{jj}\pi=\pi/\pi_j, \tag{1.21}$$

and putting this in the moment identity (1.20) above for $\lambda=\delta_j$, $r=2$, we obtain after rearrangement the identity

$$m_{\pi j}=(m_{jj}^{(2)}+m_{jj})/2m_{jj}$$

(cf. Chung [3], I.11); furthermore for $r=2, 3, \dots$ we see that $m_{\pi j}^{(r-1)}$ is finite if and only if $m_{jj}^{(r)}$ is finite, i.e. $P$ has recurrence times with finite $r$-th moment (Kemeny, Snell and Knapp [10], 9.65). Thus taking $\mu=\pi$ in Theorem 2 gives us the following corollary:

**Corollary 1.** *Suppose that $P$ is irreducible, aperiodic, and positive recurrent with invariant probability $\pi$, and that $P$ has recurrence times with finite r-th moment for some integer $r\geqq 2$. Let $\lambda$ be an initial distribution such that $m_{\lambda j}^{(r-1)}$ is finite for j in J (in particular this will be the case if $\lambda$ is bounded by a multiple of $\pi$). Then*

$$\lim_{n\to\infty} n^{r-1}\|\lambda P^n-\pi\|=0, \tag{1.22}$$

$$\sum_{n=1}^{\infty} n^{r-2}\|\lambda P^n-\pi\|<\infty, \tag{1.23}$$

*and*

$$\lim_{n\to\infty} n^{r-2}\left\|\lambda\sum_{k=1}^{n} P^k-n\pi-v'\right\|=0, \tag{1.24}$$

*where $v'$ is the signed measure on J with total mass zero which is given by*

$$v'_j=[(m_{jj}^{(2)}+m_{jj})/2m_{jj}-m_{\lambda j}]/m_{jj}, \quad j\in J. \tag{1.25}$$

In particular, we deduce from the corollary with $r=2$ (i.e. from (1.13) really) that if $P$ has recurrence times with finite second moment and $m_{\lambda j}$ is finite, $j\in J$, then for $j\in J$

$$\sum_{n=1}^{\infty}(p_{\lambda j}^{(n)}-\pi_j)=\lim_{n\to\infty}\left[\sum_{k=1}^{n}p_{\lambda j}^{(k)}-n\pi_j\right]=v'_j, \tag{1.26}$$

and the series on the left is uniformly absolutely convergent over $j$ in $J$, indeed

$$\sum_{j\in J}\sum_{n=1}^{\infty}|p_{\lambda j}^{(n)}-\pi_j|<\infty. \tag{1.27}$$

The case of (1.26) with $\lambda=\delta_j$ was proved by Feller in [5] using power series methods. Also implicit in Feller's work is the result

$$\lim_{n\to\infty} n\,|p_{jj}^{(n)}-\pi_j|=0,$$

which is of course implied by (1.19) with $r=2$. The corresponding special cases with $r$ greater than 2 can be found in a renewal theoretic context in Karlin's paper [9], but the uniformity assertions are new.

If $P$ has recurrence times with finite $r$-th moment and $\lambda$ is such that $m_{\lambda j}^{(r)}$ is finite then it is easy to see that $m_{\lambda P,j}^{(r)}$ is also finite; thus taking $\mu=\lambda P$ in Theorem 2 gives us rates for the convergence of $\lambda P^n(I-P)$ to the zero signed measure $\pi(I-P)$.

**Corollary 2.** *Let $r$ be a positive integer and suppose that $P$ is irreducible and aperiodic with recurrence times having finite r-th moment. Then for all initial distributions $\lambda$ on $J$ such that $m_{\lambda j}^{(r)}$ is finite for $j$ in $J$, both*

$$\lim_{n\to\infty} n^r\,\|\lambda P^n(I-P)\|=0, \tag{1.28}$$

*and*

$$\sum_{n=0}^{\infty} n^{r-1}\,\|\lambda P^n(I-P)\|<\infty. \tag{1.29}$$

In particular, from the case $r=1$ of Corollary 2 we see if $P$ is irreducible, aperiodic and positive recurrent then for any initial distribution $\lambda$ on $J$ such that $m_{\lambda j}$ is finite for $j$ in $J$, the sequences

$$(p_{\lambda j}^{(n)})_{n\in\mathbb{N}},\quad j\in J$$

have uniformly bounded variations

$$V_{\lambda j}=\sum_{n=0}^{\infty}|p_{\lambda j}^{(n)}-p_{\lambda j}^{(n+1)}|,$$

and indeed

$$\sum_{j\in J}|V_{\lambda j}|<\infty.$$

The only result in this direction which seems to have been obtained previously is that $V_{jj}$ is finite for each $j$ in $J$ (see Kingman [13], 1.6(iv), where this result is seen to follow from the original Erdös-Feller-Pollard proof of the renewal theorem using Wiener's theorem on the reciprocal of an absolutely convergent Fourier series).

These then are the main results of the paper. For each of the above results concerning the convergence of $\lambda P^n$ for measures $\lambda$ on $J$ there is a corresponding dual result on the convergence of $P^n f$ for functions $f$ on $J$, and putting the results for measures and functions together it is possible to obtain a very complete description of the way in which $\lambda P^n f$ converges to $\pi f$ for varying probabilities $\lambda$ and functions $f$ on the state space. These matters are discussed in Section 5 after a reformulation of the main theorems in terms of the way the transition matrix $P$

acts on signed measures of total mass zero, a reformulation which shows up the connection between the present results and the potential theory of recurrent Markov chains due to Kemeny, Snell and Knapp. At the end of Section 5 there is a discussion of the lack of converses to the main theorems. Finally, in Section 6 we consider the implications of our results for renewal theory. The uniformity assertions of the Markov chain theorems are not of any interest in this context, but it is still found that the present results extend the theorems of Feller and Karlin on the rates of convergence of delayed renewal sequences (i.e. sequences of transition probabilities $(p_{\lambda j}^{(n)})_{n\in\mathbb{N}}$). However the results for zero delay renewal sequences (i.e. sequences of diagonal transition probabilities $(p_{jj}^{(n)})_{n\in\mathbb{N}}$) are not as sharp as those obtained by Stone in [19] using Fourier analytic techniques.

## 2. The Bivariate Chain

In this section the central theme of the paper is developed. This is the analysis of the limit behaviour of the $n$-step transition probabilities of a Markov chain by comparison of the progress of two independent Markov chains with the same transition probabilities but different initial distributions. We take as given a Markov matrix $P$ with countable state space $J$ and use the notation developed in the introduction.

The Markov matrix $P$ is *irreducible* if for all states $i$ and $j$ in $J$ there is an $n>0$ such that $p_{ij}^{(n)}>0$. A state $j$ is said to be *transient* if $\mathbb{P}_j(T_j<\infty)<1$, *recurrent* if $\mathbb{P}_j(T_j<\infty)=1$, *null recurrent* if $m_{jj}=\mathbb{E}_j T_j$ is infinite, and *positive recurrent* if $m_{jj}$ is finite. A measure $\pi$ on $J$ is said to be *invariant* for $P$ if $\pi P=\pi$. The following proposition summarises most of what will be required from the elementary theory of Markov chains. These facts will be used in the proof of the main limit theorem (2.8) below, so proofs are indicated which do not rely on this theorem.

(2.1) **Proposition.** *Suppose that $P$ is irreducible.*

(i) *Either all states are transient, all states are null recurrent or all states are positive recurrent.*

(ii) *Let $\lambda$ be a probability on $J$, $j$ a state in $J$. The states are recurrent if and only if the series*

$$\sum_{n=0}^{\infty} p_{\lambda j}^{(n)}$$

*is divergent, and in this case*

$$\mathbb{P}_\lambda(T_j<\infty)=1.$$

(iii) *The states are positive recurrent if and only if there exists an invariant probability measure $\pi$ for $P$; if it exists this $\pi$ is unique and given by $\pi_j=1/m_{jj}$, $j\in J$.*

*Proof.* For part (ii) and the fact that the states are either all transient or all recurrent see Freedman [8], Section 1.5. For (i) it remains to show that if the states are recurrent then they are either all null recurrent or else all positive recurrent. This will now be proved together with (iii). Suppose that there is a state $i$ in $J$ which is positive recurrent. Then an invariant measure $\psi$ for $P$ may be defined by setting $\psi(H)$ to be the expected number of times that a Markov chain with initial distribution $\delta_i$ and transitions $P$ visits the set $H\subset J$ before the time $T_i$ of its first return to $i$ (see [8], 2.19). Evidently $\psi(J)=m_{ii}$ and thus $\psi/m_{ii}$ is an invariant probability for $P$. On the other hand, if there exists an invariant probability $\pi$ for $P$ then all states must be recurrent by the criterion of (ii) with $\lambda=\pi$, and since irreducibility implies $\pi_j>0$, $j\in J$, it follows by applying an elementary identity due to Kac ([8], 2.46) to the stationary process of occurrences of a given state $j$ derived from a stationary Markov chain with initial distribution $\pi$ and transitions $P$ that

(2.2) $$m_{jj}=1/\pi_j, \quad j\in J,$$

so that all states are positive recurrent and $\pi$ is unique as specified.

By virtue of part (i) of the above proposition an irreducible Markov matrix $P$ is itself described as transient, null recurrent or positive recurrent according to the behaviour of its states.

A state $j$ is said to be *aperiodic* if the greatest common divisor of the set

$$\{n: n>0, p_{jj}^{(n)}>0\}$$

is one. If $P$ is irreducible then either all or no states are aperiodic, and so this property too is attributed to $P$. If $P$ is irreducible and aperiodic then for each pair of states $i$ and $j$ in $J$ there exists an $n^*\in N$ such that

(2.3) $$p_{ij}^{(n)}>0 \quad \text{for all } n>n^*$$

(see [8], Section 1.4).

We now turn to the consideration of two independent Markov chains $(X_n)$ and $(X_n')$ defined on the same probability space, both with stationary transition probabilities $P$, but with different initial distributions $\lambda$ and $\mu$. Setting $Z_n=(X_n, X_n')$, it is clear that the bivariate process $(Z_n)$ is a Markov chain with countable state space $J\times J$, initial distribution $\lambda\times\mu$ and stationary transition probabilities $\tilde{P}$ given by

(2.4) $$\tilde{p}_{(i,j),(k,l)}^{(n)}=p_{ik}^{(n)}\,p_{jl}^{(n)}, \quad i,j,k,l\in J.$$

It is of vital importance to the analysis which follows that many properties of $P$ are inherited by $\tilde{P}$. Typically we have the simple Proposition (2.5) below; some more subtle relationships between $P$ and $\tilde{P}$ will be brought out later.

(2.5) **Proposition.** *Suppose that $P$ is irreducible and aperiodic. Then $\tilde{P}$ is also irreducible and aperiodic; if furthermore $P$ is positive recurrent then so is $\tilde{P}$, and if $P$ is transient then so is $\tilde{P}$.*

*Proof.* Suppose $P$ to be irreducible and aperiodic. That $\tilde{P}$ is irreducible and aperiodic follows immediately from (2.4) and (2.3). If $P$ is also positive recurrent then (2.1)(iii) shows that $\tilde{P}$ must be positive recurrent too, for if $\pi$ is the invariant probability for $P$ then $\pi\times\pi$ is evidently invariant for $\tilde{P}$. Finally, $\tilde{P}$ is obviously transient if $P$ is transient.

There is a remarkable gap in Proposition (2.5): $\tilde{P}$ need not inherit from $P$ the property of null recurrence. Indeed if $P$ is null recurrent then $\tilde{P}$ may be either null recurrent or transient, though it obviously could not be positive recurrent. Examples of both possibilities are readily constructed from the $P$'s corresponding to suitable null recurrent one (respectively, two) dimensional random walks, for which $\tilde{P}$ can be made to correspond to a null recurrent two dimensional random walk (respectively, transient four dimensional random walk).

From now on it will be assumed that the bivariate process $(Z_n)$ has in fact been constructed in the usual way as a sequence of $J \times J$ valued random variables $Z_n = (X_n, X'_n)$ defined on a measurable space $(\tilde{\Omega}, \tilde{\mathscr{F}})$ such that for each probability measure $\eta$ on $J \times J$ there is a probability measure $\tilde{\mathbb{P}}_\eta$ on $\tilde{\mathscr{F}}$ under which $(Z_n)$ is a Markov chain with initial distribution $\eta$ and stationary transition probability matrix $\tilde{P}$ given by (2.4). Let $\lambda$ and $\mu$ be two probabilities on $J$. Then under $\tilde{\mathbb{P}}_{\lambda \times \mu}$ the processes $(X_n)$ and $(X'_n)$ are two independent Markov chains, both with state space $J$ and transition matrix $P$, but with initial distributions $\lambda$ and $\mu$ respectively.

Let us now take an arbitrary state $i$ in $J$, to be considered fixed throughout the discussion, and let $T_{(i,i)}$ (or simply $T$) denote the first passage time of $(Z_n)$ to $(i, i)$:

$$T_{(i,i)} = T = \inf\{n\colon n > 0, X_n = X'_n = i\},$$

where $T$ takes the value $\infty$ if $(Z_n)$ never hits $(i, i)$. The point of the construction of the bivariate process is this: under $\tilde{\mathbb{P}}_{\lambda \times \mu}$ the distribution of $X_n$ is $\lambda P^n$ and the distribution of $X'_n$ is $\mu P^n$; but once $T$ has occurred the evolution of the two processes is distributionally the same since both processes start afresh at time $T$ in state $i$, and thus any difference between the distribution $\lambda P^n$ of $X_n$ and the distribution $\mu P^n$ of $X'_n$ can only arise from the set $\{T > n\}$ in the underlying probability space $\tilde{\Omega}$ on which $T$ happens after time $n$. To be precise, we recall that $\|v\|$ denotes the total variation of a signed measure $v$ on $J$; then with no extra assumptions whatsoever on the nature of the transition matrix $P$ indexed by $J$ we have the following inequality:

(2.6) **Lemma.** *Let $\lambda$ and $\mu$ be two probabilities on $J$, $i$ a state in $J$. Then*

$$\|\lambda P^n - \mu P^n\| \leqq 2\tilde{\mathbb{P}}_{\lambda \times \mu}(T_{(i,i)} > n).$$

*Proof.* For $j$ in $J$ we have

$$\begin{aligned} p_{\lambda j}^{(n)} = \tilde{\mathbb{P}}_{\lambda \times \mu}(X_n = j) &= \sum_{m=1}^{n} \tilde{\mathbb{P}}_{\lambda \times \mu}(T = m, X_n = j) + \tilde{\mathbb{P}}_{\lambda \times \mu}(T > n, X_n = j) \\ &= \sum_{m=1}^{n} \tilde{\mathbb{P}}_{\lambda \times \mu}(T = m)\, p_{ij}^{(n-m)} + \tilde{\mathbb{P}}_{\lambda \times \mu}(T > n, X_n = j), \end{aligned}$$

by the Markov property of $(Z_n)$. But equally

$$p_{\mu j}^{(n)} = \tilde{\mathbb{P}}_{\lambda \times \mu}(X'_n = j) = \sum_{m=1}^{n} \tilde{\mathbb{P}}_{\lambda \times \mu}(T = m)\, p_{ij}^{(n-m)} + \tilde{\mathbb{P}}_{\lambda \times \mu}(T > n, X'_n = j),$$

and thus

$$p_{\lambda j}^{(n)} - p_{\mu j}^{(n)} = \tilde{\mathbb{P}}_{\lambda \times \mu}(T > n, X_n = j) - \tilde{\mathbb{P}}_{\lambda \times \mu}(T > n, X'_n = j). \tag{2.7}$$

Certainly then

$$|p_{\lambda j}^{(n)} - p_{\mu j}^{(n)}| \leqq \tilde{\mathbb{P}}_{\lambda \times \mu}(T > n, X_n = j) + \tilde{\mathbb{P}}_{\lambda \times \mu}(T > n, X'_n = j),$$

and adding this inequality over all $j$ in $J$ yields the inequality of the lemma.

As we shall soon see the simple inequality of Lemma (2.6) permits the decisive resolution of a great many questions concerning the convergence of the $n$-step transition probabilities of the original Markov chain with transitions $P$. All that is left to do to obtain limit theorems for $P^n$ is to ascertain how the transition matrix $P$ and the initial distributions $\lambda$ and $\mu$ influence the decay of the tail probabilities $\tilde{\mathbb{P}}_{\lambda\times\mu}(T>n)$ of the first passage time $T$ of the bivariate process. Already from Proposition (2.5) we have enough to establish the main limit theorem:

(2.8) **Theorem** *(Kolmogorov). Let $P$ be an irreducible and aperiodic Markov matrix with countable state space $J$. Then for all initial distributions $\lambda$ on $J$*

$$\lim_{n\to\infty} p_{\lambda j}^{(n)} = 1/m_{jj}, \tag{2.9}$$

*where $m_{jj}$ is the expected recurrence time of state $j$ and $1/m_{jj}$ is taken to be zero in the transient and null recurrent cases when $m_{jj}=\infty$. If $P$ is positive recurrent then furthermore*

$$\lim_{n\to\infty} \|\lambda P^n - \pi\| = 0, \tag{2.10}$$

*where $\pi=(\pi_j)=(1/m_{jj})$ is the unique invariant probability for $P$.*

*Proof.* If $P$ is transient then (2.9) is an obvious consequence of (2.1)(ii). If $P$ is positive recurrent then $\tilde{P}$ is recurrent by Proposition (2.5) so that given two initial distributions $\lambda$ and $\mu$ on $J$ and a state $i$ in $J$ we have from (2.1)(ii) applied to $\tilde{P}$ that

$$\tilde{\mathbb{P}}_{\lambda\times\mu}(T_{(i,i)}<\infty)=1. \tag{2.11}$$

Thus $\lim_{n\to\infty} \tilde{\mathbb{P}}_{\lambda\times\mu}(T_{(i,i)}>n)=0$ and the inequality of Lemma (2.6) gives us

$$\lim_{n\to\infty} \|\lambda P^n - \mu P^n\| = 0. \tag{2.12}$$

The results (2.9) and (2.10) now follow on taking $\mu$ to be the unique invariant probability $\pi$ for $P$ (see (2.1)(iii)).

Lastly, if $P$ is null recurrent then either $\tilde{P}$ is transient, in which case by (2.1)(ii) and (2.4) we have

$$\sum_{n=0}^{\infty} \tilde{p}_{\lambda\times\lambda,(j,j)}^{(n)} = \sum_{n=0}^{\infty} (p_{\lambda j}^{(n)})^2 < \infty,$$

so that quite obviously $\lim_{n\to\infty} p_{\lambda j}^{(n)}=0$, or else $\tilde{P}$ is null recurrent, in which case (2.11) and (2.12) still hold and we can deduce (2.9) by the following argument due to Orey [15]: Suppose that there is an initial distribution $\lambda$ such that $p_{\lambda j}^{(n)}$ does not converge to zero for all $j$ in $J$. Then by the usual diagonal argument there exists an increasing sequence of positive integers $(k_n)$ and a sub-probability measure $\psi=(\psi_j)$ on $J$ such that

$$\lim_{n\to\infty} p_{\lambda j}^{(k_n)} = \psi_j, \quad j\in J, \tag{2.13}$$

with $0<\psi(J)\leq 1$, this last inequality being a consequence of Fatou's Lemma. Now using (2.13) and (2.12) with $\mu=\lambda P$ it is easy to see that $\psi$ must be invariant for $P$ so that $\psi/\psi(J)$ is an invariant probability for $P$. But according to (2.1)(iii) there exists no such invariant probability for null recurrent $P$, and we have a contradiction. This completes the proof of the theorem.

Orey showed in [15] that (2.12) in fact holds for all irreducible, aperiodic and recurrent transition matrices $P$ (see also Freedman [8], 1.13), but the argument here only works when $\tilde{P}$ is recurrent. It might be thought that the present argument could be repaired by replacing $T$ by the earlier time $T^*$ when $(X_n)$ and $(X_n')$ first meet anywhere in $J$, but unfortunately there even exist null recurrent $P$ for which $T^*$ can be infinite with positive probability (cf. Freedman [8], 1.13).

In the proof of the theorem for positive recurrent $P$ we used only the recurrence of $\tilde{P}$, disregarding the actual positive recurrence of $\tilde{P}$ which is assured by (2.5). In the next section this positive recurrence of $\tilde{P}$ is exploited in conjunction with the inequality (2.6) to obtain the sharper statements of Theorem 1, while in Section 4 we obtain the further refinements of Theorem 2 by using the fact that $\tilde{P}$ inherits from $P$ the property of having recurrence times with finite $r$-th moments, $r=1, 2, \ldots$. These results are only of interest for positive recurrent Markov chains with infinite state space $J$, since if $J$ is finite they are eclipsed by the fact that convergence of the transition probabilities occurs geometrically fast (see Kendall [11]). This fact too can be deduced from the key inequality (2.6), for if the state space $J$ is finite then so is the state space $J\times J$ of the bivariate chain, and it then follows from the irreducibility and recurrence of $\tilde{P}$ that there exist constants $c$ and $\rho$ with $0<c<\infty$ and $0<\rho<1$ such that for any initial distribution $\eta$ on $J\times J$

$$\tilde{\mathbb{P}}_\eta(T>n)\leqq c\,\rho^n, \qquad n\in N$$

(see Freedman [8], 1.9, (79)). Thus immediately from (2.6) with $\mu=\pi$ we have that for any initial distribution $\lambda$ on $J$

$$\|\lambda P^n-\pi\|\leqq 2c\,\rho^n, \qquad n\in N.$$

It might also be possible to study geometric ergodicity of recurrent Markov chains with infinite state space in this way, but the method does not seem to be immediately rewarding (cf. [11], [12], [20]).

We conclude this section with a derivation of the Erdös-Feller-Pollard renewal theorem [4] from Kolmogorov's theorem. That this can be done is well known, but the argument here is not the usual one and we shall wish to refer back to it later.

(2.14) **Theorem** *(Erdös-Feller-Pollard). Let $(f_n)$ be a probability distribution on the positive integers with g.c.d. $\{n: f_n>0\}=1$, and let $(u_n)$ be the associated aperiodic renewal sequence:*

$$u_0=1; \qquad u_n=\sum_{m=1}^{n} f_m\,u_{n-m}, \qquad n=1,2,\ldots.$$

*Let $\mu=\sum_{n=1}^{\infty} n f_n$. Then*

$$\lim_{n\to\infty} u_n=1/\mu,$$

*where $1/\mu$ is taken to be zero if $\mu=\infty$.*

*Proof.* Let $M$ be the supremum of the set of integers $n$ for which $f_n>0$, and consider the state space $J$ comprising all integers $j$ with $0\leqq j<M$. Define a transition

matrix $P$ indexed by $J$ as follows:

$$p_{0j}=f_{j+1}, \quad 0\leqq j<M,$$

and for $0<i<M$

$$\begin{aligned} p_{ij}&=1 \quad \text{if } j=i-1, \\ &=0 \quad \text{otherwise.} \end{aligned}$$

Obviously $P$ is irreducible, and $P$ is recurrent and aperiodic since for initial distribution $\delta_0$ the distribution of the first passage time to 0 is just $(f_n)$. Furthermore it is easy to see that $p_{00}^{(n)}=u_n$ and thus the proof is completed by applying (2.9) to this transition matrix $P$ with $\lambda=\delta_0$ and $j=0$.

The $P$ used in the above proof is actually the transition matrix of the forward recurrence time Markov chain associated with a discrete time regenerative phenomenon with renewal sequence $(u_n)$ (see Kingman [13], 1.6, (xiv)). Most frequently it is observed that the renewal theorem can be derived by applying Kolmogorov's theorem to another Markov chain associated with a regenerative phenomenon, the backward recurrence time process, which also has a state with recurrence time distribution $(f_n)$. However we shall see later that the forward recurrence time chain considered here is neatly tailored to suit the applications to renewal sequences of the various refinements of Kolmogorov's Theorem which we will shortly be proving.

## 3. Proof of Theorem 1

Suppose throughout this section that $P$ is an irreducible, aperiodic and positive recurrent Markov matrix with countable state space $J$, and that $\lambda$ and $\mu$ are two initial distributions on $J$. We shall always be working within the framework developed in the previous section, but for the sake of brevity the tildes have been dropped from the previous notation $(\tilde{\Omega}, \tilde{\mathscr{F}}, \tilde{\mathbb{P}}_\eta)$: on an underlying measure space $(\Omega, \mathscr{F})$ it is assumed that for each probability distribution $\eta$ on $J\times J$ we have a measure $\mathbb{P}_\eta$ on $\mathscr{F}$ such that a $J\times J$ valued process $(Z_n)=((X_n, X_n'))$ defined on $(\Omega, \mathscr{F})$ is a Markov chain with initial distribution $\eta$ and stationary transition probabilities $\tilde{P}$ given by (2.4). In particular we have that under $\mathbb{P}_{\lambda\times\mu}$ the processes $(X_n)$ and $(X_n')$ are independent Markov chains with state space $J$, transition probabilities $P$ and initial distributions $\lambda$ and $\mu$ respectively. An arbitrary state $i$ in $J$ is taken as fixed from now on, and as in the last section $T$ is the random variable defined on $(\Omega, \mathscr{F})$ to be the first passage time of the bivariate process $(Z_n)$ to $(i, i)$, that is to say the first time that both $(X_n)$ and $(X_n')$ are simultaneously in state $i$.

The hypothesis of Theorem 1 is that $\lambda$ and $\mu$ are probabilities on $J$ such both $m_{\lambda i}$ and $m_{\mu i}$ are finite, but for the moment let us further suppose that

$$\mathbb{E}_{\lambda\times\mu} T<\infty. \tag{3.1}$$

(It may be observed that the quantity $\mathbb{E}_{\lambda\times\mu} T$ is simply a first passage moment for the bivariate chain bearing the same relation to $\tilde{P}$ as $m_{\lambda i}$ does to $P$, and were it not for the clumsiness of the notation we might well write something like $\tilde{m}_{\lambda\times\mu,(i,i)}$ instead of $\mathbb{E}_{\lambda\times\mu} T$.) Most of this section will in fact be spent showing that (3.1) is actually implied by finiteness of $m_{\lambda i}$ and $m_{\mu i}$, but once granted (3.1) the proof of Theorem 1 is easy.

Indeed, according to Lemma (2.6) we have

(3.2) $$\|\lambda P^n - \mu P^n\| \leqq 2\, \mathbb{P}_{\lambda \times \mu}(T > n),$$

but with $\mathbb{E}_{\lambda \times \mu} T$ finite we have

$$\lim_{n \to \infty} n\, \mathbb{P}_{\lambda \times \mu}(T > n) = 0,$$

hence also

$$\lim_{n \to \infty} n\, \|\lambda P^n - \mu P^n\| = 0,$$

which is the first assertion (1.8) of Theorem 1. Again, adding the inequality (3.2) over $n$ gives us

$$\sum_{n=0}^{\infty} \|\lambda P^n - \mu P^n\| \leqq 2 \sum_{n=0}^{\infty} \mathbb{P}_{\lambda \times \mu}(T > n) = 2\, \mathbb{E}_{\lambda \times \mu} T,$$

and thus with (3.1) we have

(3.3) $$\sum_{n=1}^{\infty} \|\lambda P^n - \mu P^n\| < \infty,$$

which is just the second assertion (1.9) of Theorem 1. Finally, letting $\nu_n$ denote the signed measure $\sum_{k=1}^{n} (\lambda - \mu) P^k$ of total mass zero, it is an immediate consequence of (3.3) that $(\nu_n)$ is in fact a Cauchy sequence in the Banach space of signed measures on $J$ with total variation norm $\| \;\|$, and thus $(\nu_n)$ must indeed converge to some signed measure $\nu$ on $J$ with total mass zero. It only remains to evaluate this limit $\nu$, and this is effected by Proposition (3.5) below.

In the proof of Proposition (3.5) and again later we will make essential use of the version of Wald's identity now stated as a lemma:

(3.4) **Lemma.** *Let $\mathbb{P}$ be a probability on $(\Omega, \mathscr{F})$, and suppose that on the probability space $(\Omega, \mathscr{F}, \mathbb{P})$ there is defined a sequence $(Y_n)$ of independent and identically distributed non-negative random variables with $\mathbb{P}(Y_1 > 0) > 0$, as well as a non-negative integer valued random variable $N$ such that for each $n = 1, 2, \ldots$ the event $\{N < n\}$ is independent of $Y_n$. Then*

$$\mathbb{E} \sum_{n=1}^{N} Y_n = (\mathbb{E}\, Y_1)(\mathbb{E}\, N),$$

*where $\mathbb{E}$ denotes expectation with respect to $\mathbb{P}$, and a vacuous sum is taken to be zero.*

*Proof.* See [8], 1.8, (71).

(3.5) **Proposition.** *If $\mathbb{E}_{\lambda \times \mu} T$ is finite then*

$$\sum_{n=1}^{\infty} (p_{\lambda i}^{(n)} - p_{\mu i}^{(n)}) = (m_{\mu i} - m_{\lambda i})/m_{ii},$$

*and the series is absolutely convergent.*

*Proof.* We already know from (3.3) that the series is absolutely convergent; that it converges to the stated limit is a well known fact which can be deduced from the main limit Theorem (2.8) by dominated convergence arguments (see Kemeny, Snell and Knapp [10], 9.48 and 9.50), but it is nonetheless interesting to consider the following alternative evaluation of the sum which is more in the spirit of the present approach through the bivariate chain. From adding the identity (2.7) over all $n \geqq 1$ we have

$$(3.6) \qquad \sum_{n=1}^{\infty} (p_{\lambda i}^{(n)} - p_{\mu i}^{(n)}) = \sum_{n=1}^{\infty} \mathbb{P}_{\lambda \times \mu}(T > n, X_n = i) - \sum_{n=1}^{\infty} \mathbb{P}_{\lambda \times \mu}(T > n, X_n' = i)$$

since both series on the right hand side are bounded above by

$$\sum_{n=0}^{\infty} \mathbb{P}_{\lambda \times \mu}(T > n) = \mathbb{E}_{\lambda \times \mu} T.$$

But let $N$ be the random variable defined on $(\Omega, \mathscr{F})$ as the number of times that $X_n = i$, $0 < n < T$, and define $N'$ similarly in terms of $(X_n')$. Then (3.6) just states that

$$(3.7) \qquad \sum_{n=1}^{\infty} (p_{\lambda i}^{(n)} - p_{\mu i}^{(n)}) = \mathbb{E}_{\lambda \times \mu} N - \mathbb{E}_{\lambda \times \mu} N'.$$

To evaluate the right hand side of (3.7) let us set $T_0 = 0$ and define random variables $0 < T_1 < T_2 \ldots$ on $(\Omega, \mathscr{F})$ as the successive positive times that $X_n = i$:

$$T_m = \inf\{n \colon n > T_{m-1}, X_n = i\}, \qquad m = 1, 2, \ldots;$$

we set $R_m = T_{m+1} - T_m$, and make similar definitions for $T_m'$ and $R_m'$ in terms of $(X_n')$. Now evidently

$$T = T_1 + \sum_{m=1}^{N} R_m = T_1' + \sum_{m=1}^{N'} R_m',$$

whence

$$(3.8) \qquad \mathbb{E}_{\lambda \times \mu} T = m_{\lambda i} + m_{ii} \mathbb{E}_{\lambda \times \mu} N = m_{\mu i} + m_{ii} \mathbb{E}_{\lambda \times \mu} N',$$

where we have used the fact that $\mathbb{E}_{\lambda \times \mu} T_1$ is simply $m_{\lambda i}$, and applied Wald's identity (3.4) to the sequence of random variables $R_1, R_2, \ldots$ which are independent and identically distributed under $\mathbb{P}_{\lambda \times \mu}$ with mean $m_{ii}$ (cf. [8], 1.3, (31) and (32)), with the random time $N$ which is such that under $\mathbb{P}_{\lambda \times \mu}$ the event $\{N < m\}$ is independent of $R_m$ by the strong Markov property of the bivariate chain at its stopping time $T_m$, together of course with similar considerations for the primed quantities. But this identity (3.8) gives us exactly what is needed for (3.5) in (3.6):

$$\mathbb{E}_{\lambda \times \mu} N - \mathbb{E}_{\lambda \times \mu} N' = (m_{\mu i} - m_{\lambda i})/m_{ii}.$$

This now completes the proof of Theorem 1 under the additional assumption (3.1) that $\mathbb{E}_{\lambda \times \mu} T$ is finite. It only remains to show that this assumption is in fact implied by the hypothesis of the theorem that both $m_{\lambda i}$ and $m_{\mu i}$ are finite. This is the substance of the following proposition, whose proof takes up the remainder of the section.

(3.9) **Proposition.** *Let $\eta$ be a probability on $J \times J$ with marginal distributions $\lambda$ and $\mu$ on $J$. Then $\mathbb{E}_\eta T$ is finite if and only if both $m_{\lambda i}$ and $m_{\mu i}$ are finite.*

*Proof.* One way is quite trivial: if $\mathbb{E}_\eta T$ is finite then certainly both $m_{\lambda i}$ and $m_{\mu i}$ are finite, indeed less than or equal to $\mathbb{E}_\eta T$, since for the first passage time $T_1$ of the marginal chain $(X_n)$ to the state $i$ we have certainly $T_1 \leqq T$, thus

$$m_{\lambda i} = \mathbb{E}_\eta T_1 \leqq \mathbb{E}_\eta T,$$

with similar considerations for $m_{\mu i}$. It is however the converse implication which is required to establish (3.1), and for this part it seems to be necessary to build up the proof gradually. Completion of the proof is therefore deferred until after the development of some preliminary results.

Let us first observe that since $P$ is irreducible and positive recurrent we do at least know that $\mathbb{E}_\eta T$ is finite whenever $\eta$ has finite support (see (1.7)), so in particular if $\lambda$ is a probability on $J$ with finite support then $\mathbb{E}_{\delta_i \times \lambda} T$ is finite. We thus deduce from (3.5) that if $\lambda$ has finite support then

(3.10) $$\sum_{n=1}^{\infty} (p_{ii}^{(n)} - p_{\lambda i}^{(n)}) = (m_{\lambda i} - m_{ii})/m_{ii},$$

and the series is absolutely convergent. This is enough to establish the following two lemmas:

(3.11) **Lemma.** *The identity* (3.10) *holds for an arbitrary initial distribution* $\lambda$ *on* $J$ *provided the series is interpreted as diverging to* $+\infty$ *if* $m_{\lambda i} = \infty$.

*Remark.* It may be noted that we are not yet asserting the actual absolute convergence of the series in (3.10) when $m_{\lambda i}$ is finite. This has to wait until Lemma (3.15) below when it may be deduced from (3.5). The trouble is that for the partial sums of the absolute series there is no obvious bound analogous to (3.12) below which would make a dominated convergence argument work directly from the absolute convergence of (3.10) for $\lambda$ with finite support.

*Proof.* For a probability $\gamma$ on $J$ let $S_\gamma^k$ denote the $k$-th partial sum $\sum_{n=1}^{k} (p_{ii}^{(n)} - p_{\gamma i}^{(n)})$. A simple first entrance to $i$ argument shows that for any probability $\gamma$ on $J$ and for all $k$ we have

(3.12) $$-1 \leqq S_\gamma^k \leqq m_{\gamma i}$$

(see [8], Lemma 9.48). Now

$$S_\lambda^k = \sum_{j \in J} \lambda_j S_j^k,$$

and we know from (3.10) that

$$\lim_{k \to \infty} S_j^k = (m_{ji} - m_{ii})/m_{ii},$$

with $|S_j^k| \leqq m_{ji}$ by (3.12) and $\sum_{j \in J} \lambda_j m_{ji} = m_{\lambda i}$. Thus if $m_{\lambda i}$ is finite we deduce by dominated convergence that

$$\lim_{k \to \infty} S_\lambda^k = \sum_{j \in J} \lambda_j (m_{ji} - m_{ii})/m_{ii} = (m_{\lambda i} - m_{ii})/m_{ii}.$$

If on the other hand $m_{\lambda i}$ is infinite, we have from Fatou's lemma that

$$\liminf_{k\to\infty} S_\lambda^k \geqq \sum_{j\in J} \lambda_j \lim_{k\to\infty} S_j^k = \sum_{j\in J} \lambda_j (m_{ji} - m_{ii})/m_{ii} = \infty,$$

which is to say that the series (3.10) diverges to $+\infty$.

We shall also employ the following interesting consequence of (3.10):

(3.13) **Lemma.** *Let $(u_n)$ be an aperiodic renewal sequence associated with a distribution $(f_n)$ on the positive integers which has finite mean. Then for any sequence $(a_n)$ of real numbers the series*

$$\sum_{n=0}^{\infty} a_n \quad \textit{and} \quad \sum_{n=0}^{\infty} a_n u_n$$

*converge and diverge together.*

*Proof.* Consider (3.10) for the particular Markov matrix $P$ that was used in the proof of the renewal theorem (2.14), with $i=0$, $\lambda=\delta_1$. In this case we have

$$p_{00}^{(n)} = u_n, \qquad p_{10}^{(n)} = u_{n-1}, \qquad n \geqq 1,$$

and thus the absolute convergence of the series in (3.10) gives us

$$\sum_{n=1}^{\infty} |u_n - u_{n-1}| < \infty \tag{3.14}$$

(see also Kingman [13], 1.6, (iv)). The assertion of the lemma is now a well known property of any sequence $(u_n)$ which satisfies (3.14) and has a non zero limit (see Ferrar [7], Theorem 25 A).

We are now in a position to establish the special case of Proposition (3.9) when $\eta$ is of the form $\lambda \times \delta_i$ for some probability $\lambda$ on $J$. This is Lemma (3.15).

(3.15) **Lemma.** *Let $\lambda$ be a probability on $J$.*

$$\textit{If} \quad m_{\lambda i} < \infty \quad \textit{then} \quad \mathbb{E}_{\lambda\times\delta_i} T < \infty.$$

*Proof.* Suppose that $m_{\lambda i} < \infty$. Now Lemma (3.11) above is expressed in terms of the original irreducible, aperiodic and positive recurrent Markov matrix $P$, but it applies equally well to the derived matrix $\tilde{P}$ which has these same properties (Proposition (2.5)). Thus to establish that $\mathbb{E}_{\lambda\times\delta_i} T$ is finite it suffices to demonstrate the convergence of the series

$$\sum_{n=1}^{\infty} [\tilde{p}_{(i,i),(i,i)}^{(n)} - \tilde{p}_{\lambda\times\delta_i,(i,i)}^{(n)}] = \sum_{n=1}^{\infty} (p_{ii}^{(n)} - p_{\lambda i}^{(n)})\, p_{ii}^{(n)} \tag{3.16}$$

(see (2.4)). But since $m_{\lambda i} < \infty$ we have from Lemma (3.11) (applied straightforwardly now to $P$), that the series

$$\sum_{n=1}^{\infty} (p_{ii}^{(n)} - p_{\lambda i}^{(n)})$$

is convergent, and furthermore $(p_{ii}^{(n)})$ is the aperiodic renewal sequence associated with the first return distribution of the state $i$ which has mean $m_{ii}$. Thus taking $a_n = p_{ii}^{(n)} - p_{\lambda i}^{(n)}$ and $u_n = p_{ii}^{(n)}$ in (3.13) we deduce that the series (3.16) is convergent and (3.15) follows.

At last we can prove Proposition (3.9) in its full generality:

*Proof of Proposition* (3.9). Suppose $\eta$ is a probability on $J \times J$ with marginals $\lambda$ and $\mu$ on $J$ such that both $m_{\lambda i}$ and $m_{\mu i}$ are finite. We wish to show that $\mathbb{E}_\eta T$ is finite. Now as in the proof of (3.5) let $T_1'$ be the first positive time that $X_n' = i$, and let $\alpha$ denote the $\mathbb{P}_\eta$ distribution of $X_{T_1'}$. We have that $\mathbb{E}_\eta T_1' = m_{\mu i} < \infty$ and from the strong Markov property of the bivariate chain at the stopping time $T_1'$ we get

$$\mathbb{E}_\eta T = m_{\mu i} + \mathbb{E}_{\alpha \times \delta_i} T.$$

But now by (3.15) it suffices to show that $m_{\alpha i} < \infty$, and to this end we consider the random time $W$ when $(X_n)$ first hits $i$ after $T_1'$, the first positive time that $(X_n')$ hits $i$. By using the strong Markov property at $T_1'$ again we have

$$\mathbb{E}_\eta W = m_{\mu i} + m_{\alpha i},$$

so all we have to do now is show that $\mathbb{E}_\eta W < \infty$. But letting $M$ denote the number of times $n$ that $X_n = i$, $0 < n \leqq T_1'$, we find by applying Wald's identity (3.4) as in the proof of (3.5) that

$$\mathbb{E}_\eta W = m_{\lambda i} + m_{ii} \mathbb{E}_\eta M \leqq m_{\lambda i} + m_{ii} m_{\mu i} < \infty,$$

where we have used the fact that $M \leqq T_1'$ so that $\mathbb{E}_\eta M \leqq \mathbb{E}_\eta T_1' = m_{\mu i}$. This concludes the proof.

## 4. Proof of Theorem 2

We shall be proceeding very much as in the proof of Theorem 1 in Section 3, and we retain the general framework and notation of that section. Let $r$ be a positive integer. For Theorem 2 we suppose that $P$ is an irreducible and aperiodic Markov matrix which has recurrence times with finite $r$-th moments, and that $\lambda$ and $\mu$ are two initial distributions on the state space $J$ such that both $m_{\lambda i}^{(r)}$ and $m_{\mu i}^{(r)}$ are finite. We shall devote ourselves to the task of establishing that these assumptions in fact imply

$$\mathbb{E}_{\lambda \times \mu} T^r < \infty, \tag{4.1}$$

where $T$ is still the first passage time of the bivariate process to $(i, i)$. Once granted (4.1) the reader will easily deduce Theorem 2 from the fundamental inequality (2.6) by making the obvious extension of the arguments used at the beginning of the last section for the case $r = 1$.

We begin with a general result concerning moments of first passage times which we will later apply to both the bivariate chain and its marginal chains. To formulate this result we suppose that we are given an arbitrary Markov matrix $P^*$ indexed by a countable set of states $K$, and that on our underlying measurable space $(\Omega, \mathscr{F})$ we have defined for each initial distribution $\kappa$ on $K$ a probability $\mathbb{P}_\kappa^*$ on $\mathscr{F}$ under which a sequence of $K$ valued random variables $(Y_n)$ defined on $(\Omega, \mathscr{F})$ is a Markov chain with initial distribution $\kappa$ and stationary transition probabilities $P^*$.

In addition to defining the probabilities $\mathbb{P}_\kappa^*$ on $(\Omega, \mathscr{F})$ for probabilities $\kappa$ on $K$ it is convenient also to define a measure $\mathbb{P}_\kappa^*$ on $(\Omega, \mathscr{F})$ corresponding to a measure

$\kappa=(\kappa_k)$ on $K$ even when $\kappa$ is not a probability. We define this measure $\mathbb{P}_\kappa^*$ by

$$\mathbb{P}_\kappa^* = \sum_{k\in K} \kappa_k \mathbb{P}_\kappa^*.$$

Let $W$ be an extended positive integer valued random variable defined on $(\Omega, \mathscr{F})$, to be thought of as a random time. The *pre-$W$ occupation measure for $(Y_n)$ under $\mathbb{P}_\kappa^*$* is the measure $\upsilon$ on $K$ defined by

$$(4.2)\qquad \upsilon(H)=\mathbb{E}_\kappa^* \sum_{n=0}^{W-1} 1_H(Y_n), \qquad H\subset K,$$

where $1_H$ is the indicator function of the set $H\subset K$ and $\mathbb{E}_\kappa^*$ denotes integration with respect to $\mathbb{P}_\kappa^*$. Thus for a probability $\kappa$ on $K$, $\upsilon(H)$ is the expected number of visits of $(Y_n)$ to be the set $H$ before time $W$ for initial distribution $\kappa$. That $\upsilon$ is indeed a measure is apparent from the alternative expression

$$(4.3)\qquad \upsilon(H)=\sum_{n=0}^{\infty} \mathbb{P}_\kappa^*(Y_n\in H,\ W>n), \qquad H\subset K$$

which is easily derived from the definition. We note in particular that $\upsilon(K)=\mathbb{E}_\kappa^* W$.

We need finally some notation to describe the higher moments of a random time $W$. For extended non-negative integers $n$ let $[n]_0=1$ and for $r=1, 2, \ldots$ let

$$[n]_r=n(n+1)\ldots(n+r-1).$$

If $\mathbb{P}$ is a measure on $(\Omega, \mathscr{F})$ and $\mathbb{E}$ denotes integration with respect to $\mathbb{P}$, then of course

$$(4.4)\qquad \mathbb{E}[W]_r=\sum_{n=1}^{\infty}[n]_r\,\mathbb{P}(W=n),$$

and for $r\geqq 1$ we also have the useful identity

$$(4.5)\qquad \mathbb{E}[W]_r=r\sum_{n=1}^{\infty}[n]_{r-1}\,\mathbb{P}(W\geqq n).$$

(4.6) **Proposition.** *Suppose that $W$ is the first passage time of the Markov chain $(Y_n)$ to some set of states $G$ contained in $K$. Let $\kappa$ be a measure on $K$ and let $\upsilon$ be the pre-$W$ occupation measure for $(Y_n)$ under $\mathbb{P}_\kappa^*$ defined by* (4.2). *Then for $r=1, 2, \ldots$ we have*

$$\mathbb{E}_\kappa^*[W]_r=r\,\mathbb{E}_\upsilon^*[W]_{r-1}.$$

*Proof.* If $\mathbb{P}_\kappa^*(W=\infty)>0$ then the identity holds trivially since both sides are infinite. If on the other hand $\mathbb{P}_\kappa^*(W=\infty)=0$ then we will show that in fact

$$(4.7)\qquad \mathbb{P}_\kappa^*(W\geqq n)=\mathbb{P}_\upsilon^*(W=n), \qquad n=1, 2, \ldots.$$

Multiplying this identity by $r[n]_{r-1}$ and adding over $n$ now yields the desired result by virtue of (4.4) and (4.5). To prove (4.7) we define a matrix $Q$ to be the transition matrix $P^*$ with all columns corresponding to states in $G$ replaced by zeros, and set $R=P^*-Q$. Now it is easy to see that for any measure $\nu$ on $K$ we have for $A \subset K$,

(4.8) $$\mathbb{P}_\nu^*(W=n, X_n \in A) = \nu Q^{n-1} R 1_A, \qquad n=1,2,\ldots,$$

(4.9) $$\mathbb{P}_\nu^*(W>n, X_n \in A) = \nu Q^n 1_A, \qquad n=0,1,2,\ldots.$$

But now we have

$$\begin{aligned}
\mathbb{P}_\kappa^*(W \geqq n) &= \sum_{m=n}^{\infty} \mathbb{P}_\kappa^*(W=m) = \sum_{m=n}^{\infty} \kappa Q^{m-1} R 1 && \text{by (4.8)}\\
&= \left(\sum_{m=0}^{\infty} \kappa Q^m\right) Q^{n-1} R 1 \\
&= \upsilon Q^{n-1} R 1 && \text{by (4.3) and (4.9)}\\
&= \mathbb{P}_\upsilon^*(W=n) && \text{by (4.8)}.
\end{aligned}$$

In particular we deduce at once from Proposition (4.6) the identity (1.20) which was stated in the introduction, and we also have Corollary (4.10) below which plays a vital role in our proof of Theorem 2. A number of other results related to the identity (4.6) can be found in [16] and [17].

(4.10) **Corollary.** *Suppose as for* (4.6) *that $W$ is the first passage time of the Markov chain $(Y_n)$ to some subset of its state space and that $\upsilon$ is the pre-$W$ occupation measure for $(Y_n)$ under $\mathbb{P}_\kappa^*$. Then for each $r=0, 1, \ldots$ we have*

$$\mathbb{E}_\kappa^* W^{r+1} < \infty \quad \textit{if and only if} \quad \mathbb{E}_\upsilon^* W^r < \infty.$$

*Proof.* This is immediate from (4.6) and the trivial inequalities

$$n^r \leqq [n]_r \leqq r!\, n^r, \qquad n, r \in N.$$

Let us return now to the consideration of our irreducible, aperiodic and positive recurrent Markov matrix $P$ with invariant probability $\pi$. As observed in the introduction we have immediately from (4.10) and (1.21) that for each $r=1, 2, \ldots,$

(4.11) $$m_{ii}^{(r+1)} < \infty \quad \text{if and only if} \quad m_{\pi i}^{(r)} < \infty.$$

Considering again the bivariate chain $((X_n, X_n'))$ defined on $(\Omega, \mathscr{F}, \mathbb{P}_\eta)$ with initial distribution $\eta$ and transitions $\tilde{P}$ derived from $P$, we recall that $T$ is the first passage time of the bivariate chain to $(i, i)$. We have the following lemma:

(4.12) **Lemma.** *Let $\eta$ be a measure on $J \times J$ with marginals $\eta_1$ and $\eta_2$ on $J$ such that both $m_{\eta_1 i}$ and $m_{\eta_2 i}$ are finite, and let $\varphi$ denote the pre-$T$ occupation measure for $((X_n, X_n'))$ under $\mathbb{P}_\eta$. Then the measure $\varphi$ on $J \times J$ has marginals $\varphi_1$ and $\varphi_2$ on $J$ such that there exist finite non-negative constants $c_1$ and $c_2$ for which*

(4.13) $$\varphi_1 = \alpha_1 + c_1 \pi, \qquad \varphi_2 = \alpha_2 + c_2 \pi,$$

*where $\pi$ is the invariant probability for $P$, $\alpha_1 = \eta_1({}^iU)$ is the occupation measure for the first marginal chain $(X_n)$ prior to its first passage to $i$, and $\alpha_2 = \eta_2({}^iU)$ is defined similarly in terms of $(X_n')$.*

*Proof.* We prove the lemma just for $\eta$ a probability on $J \times J$ since the result for an arbitrary measure $\eta$ follows at once. Now it is clear that the first marginal

$\varphi_1$ of $\varphi$ is simply the pre-$T$ occupation measure for the first marginal chain $(X_n)$ under $\mathbb{P}_\eta$. That is to say, if for $H \subset J$ we let $V_H$ be the number of times $n$ that $X_n \in H$, $0 \leqq n < T$, then

$$\varphi_1(H) = \mathbb{E}_\eta V_H.$$

Now as in the proof of (3.5) let $T_0 = 0$ and let $0 < T_1 < T_2 < \cdots$ be the successive positive times $n$ that $X_n = i$, and for $k = 0, 1, \ldots$ let $V_H^k$ be the number of times $n$ that $X_n \in H$ for $T_k \leqq n < T_k$, with $N$ being the number of times $n$ that $X_n = i$ for $0 < n < T$. Now clearly

$$V_H = V_H^0 + \sum_{k=1}^{N} V_H^k,$$

so that we have

$$\varphi_1(H) = \mathbb{E}_\eta V_H = \mathbb{E}_\eta V_H^0 + \mathbb{E}_\eta \sum_{k=1}^{N} V_H^k,$$

whence

(4.14) $$\varphi_1(H) = \alpha_1(H) + (\mathbb{E}_\eta N)\, \pi(H)/\pi\{i\}, \qquad H \subset J$$

by the definition of $\alpha_1$ and the now familiar application of Wald's identity (3.4), this time to the sequence $V_H^1, V_H^2, \ldots$ of independent and identically distributed random variables with mean $\pi(H)/\pi\{i\}$ (see [8], 1.3, (31), and (1.21)), with the random time $N$ which is such that for $k = 1, 2, \ldots$ the event $N < k$ is independent of $V_H^k$ by the strong Markov property of the bivariate chain across its stopping time $T_k$. We note that $\mathbb{E}_\eta N$ is finite since $N \leqq T$ and $\mathbb{E}_\eta T$ is finite by (3.9) and our assumption that both $m_{\eta_1 i}$ and $m_{\eta_2 i}$ are finite, and thus with (4.14) we have proved the first relation of (4.13) with the constant $c_1$ equalling $(\mathbb{E}_\eta N)/\pi_i$, and the result for the second marginal follows symmetrically.

We come now to the result which completes our proof of Theorem 2. For a different formulation see (6.10)(ii) below.

(4.15) **Proposition.** *Let $r$ be a positive integer and suppose that $P$ is an irreducible and aperiodic Markov matrix which has recurrence times with finite $r$-th moments. Let $\eta$ be a measure on $J \times J$ with marginals $\eta_1$ and $\eta_2$ on $J$. Then $\mathbb{E}_\eta T^r$ is finite if and only if both $m_{\eta_1 i}^{(r)}$ and $m_{\eta_2 i}^{(r)}$ are finite.*

*Proof.* Just as was the case with Proposition (3.9), it is obvious that if $\mathbb{E}_\eta T^r$ is finite then both $m_{\eta_1 i}^{(r)}$ and $m_{\eta_2 i}^{(r)}$ must be finite, without any assumptions whatsoever on $P$. For the converse part we have to prove that given an irreducible and aperiodic Markov matrix $P$ the following statement is true for each $r = 1, 2, \ldots$:

(4.16) If $P$ has recurrence times with finite $r$-th moments then $\mathbb{E}_\eta T^r$ is finite for all measures $\eta$ on $J \times J$ such that both $m_{\eta_1 i}^{(r)}$ and $m_{\eta_2 i}^{(r)}$ are finite.

We proceed by induction. We know that (4.16) holds for $r = 1$, since this is just the content of Proposition (3.9), so let us make the inductive hypothesis:

(4.17) The statement (4.16) holds for a particular $r$.

We wish to show that (4.17) implies that (4.16) also holds with $r$ replaced by $r + 1$, so we suppose that

(4.18) $P$ has recurrence times with finite moment of order $r + 1$, and

(4.19) $\eta$ is a measure on $J \times J$ with marginals $\eta_1$ and $\eta_2$ on $J$ such that both $m_{\eta_1 i}^{(r+1)}$ and $m_{\eta_2 i}^{(r+1)}$ are finite.

We will establish that

$$\mathbb{E}_\eta T^{(r+1)} < \infty. \tag{4.20}$$

Now since $T$ is the first passage time to $(i, i)$ of the bivariate chain it follows from (4.10) that to prove (4.20) it is sufficient to show that

$$\mathbb{E}_\varphi T^r < \infty, \tag{4.21}$$

where $\varphi$ is the pre-$T$ occupation measure for the bivariate chain under $\mathbb{P}_\eta$. But by (4.18) the recurrence times of $P$ certainly have finite moment of order $r$, and thus by our inductive hypothesis (4.17) we see that it is sufficient for (4.21) to establish that the measure $\varphi$ on $J \times J$ has marginals $\varphi_1$ and $\varphi_2$ on $J$ such that both $m_{\varphi_1 i}^{(r)}$ and $m_{\varphi_2 i}^{(r)}$ are finite. But since (4.19) implies that $m_{\eta_1 i} = m_{\eta_1 i}^{(1)}$ is finite we have according to Lemma (4.12) that

$$\varphi_1 = \alpha_1 + c_1 \pi$$

where $\alpha_1$ is the occupation measure prior to the first passage time to $i$ for the marginal chain $(X_n)$ with initial distribution $\eta_1$, $c_1$ is a finite constant, and $\pi$ is the invariant probability for $P$. Thus by linearity we have

$$m_{\varphi_1 i}^{(r)} = m_{\alpha_1 i}^{(r)} + c_1 m_{\pi i}^{(r)}.$$

But $m_{\alpha_1 i}^{(r)}$ is finite by (4.10) and the assumption (4.19) that $m_{\eta_1 i}^{(r+1)}$ is finite, and $m_{\pi i}^{(r)}$ is finite by (4.11) and (4.18). We thus deduce that $m_{\varphi_1 i}^{(r)}$ is finite, and by symmetry so too is $m_{\varphi_2 i}^{(r)}$, so that (4.21) and thence (4.20) follow by the inductive hypothesis (4.17) and the argument by induction is complete.

Taking $\eta = \lambda \times \mu$ in Proposition (4.16) we see at last that our assumption (4.1) that $\mathbb{E}_{\lambda \times \mu} T^r$ is finite is indeed implied by the hypotheses of Theorem 2 that $P$ has recurrence times with $r$-th moments and that both $m_{\lambda j}^{(r)}$ and $m_{\mu j}^{(r)}$ are finite for $j$ in $J$. Our proof of Theorem 2 is thus complete.

Finally, there is an obvious corollary of Proposition (4.16) which is worth stating:

(4.22) **Corollary.** *Suppose that $P$ is irreducible and aperiodic. Then so too is $\tilde{P}$, and for each $r = 1, 2, \ldots$ $\tilde{P}$ inherits from $P$ the property of having recurrence times with finite $r$-th moments.*

*Proof.* Take $\eta = \delta_i \times \delta_i$ in (4.16).

## 5. Results for Signed Measures and Duals for Functions

Theorems 1 and 2 in the Introduction are concerned with the difference in behaviour between two recurrent Markov chains with the same transition matrix $P$ but different initial distributions $\lambda$ and $\mu$. However, these initial distributions $\lambda$ and $\mu$ really only enter the picture through their difference $\lambda - \mu$, and thus inasmuch as any signed measure of total mass zero is a constant multiple of the difference between two probabilities, these theorems are really concerned with the way $P$ acts on signed measures of total mass zero. In Theorem (5.2) below we reformulate Theorems 1 and 2 from this point of view to obtain what is actually a slight strengthening of the original theorems. Even though casting our results in terms of signed measures seems at first to be losing much of the probabilistic content of the original theorems, when we go on to consider the duals of these results for signed measures it turns out that we again have results of probabilistic significance, and putting things together at the end in Corollary (5.13) we find that for a positive recurrent $P$ with invariant probability $\pi$ we can give a very complete description of the way in which $\lambda P^n f$ converges to $\pi f$ for varying initial distributions $\lambda$ and functions $f$ on the state space. The section is concluded with a discussion of the lack of converses to the main theorems.

Suppose throughout that $P$ is an irreducible, aperiodic and recurrent Markov matrix indexed by the countable state space $J$. The reformulation of Theorems 1 and 2 which is stated below as Theorem (5.2) is perhaps best regarded as a refinement for chains having recurrence times with finite $r$-th moments of the following result due to Orey which is valid even for null-recurrent chains: for all signed measures $\varphi$ on $J$ with total mass $\varphi 1$ equal to zero,

$$\lim_{n\to\infty} \|\varphi P^n\| = 0. \tag{5.1}$$

This is just a restatement of the result remarked upon at the end of the proof of (2.8). Let us agree to say that a random variable which is finite with probability one has finite moment of order zero (arbitrarily setting $\infty^0 = \infty$), and then Orey's result becomes the case $r=0$ of the theorem stated below. The reader should refer back to (1.19) for the definition of the matrix ${}^jU$ associated with a state $j$ in $J$.

(5.2) **Theorem.** *Let $r$ be a non-negative integer and suppose that $P$ has recurrence times with finite moment of order $r$. Then for all signed measures $\varphi$ on $J$ with total mass $\varphi 1$ equal to zero and $|\varphi|({}^jU)^r 1$ finite for some (and hence every) state $j$ in $J$, we have*

$$\lim_{n\to\infty} n^r \|\varphi P^n\| = 0; \tag{5.3}$$

*if $P$ is positive recurrent ($r \geqq 1$) then also*

$$\sum_{n=0}^{\infty} n^{r-1} \|\varphi P^n\| < \infty, \tag{5.4}$$

*and as $n$ tends to infinity the signed measure $\sum_{k=0}^{n-1} \varphi P^k$ converges in total variation norm to the signed measure $\psi = -\varphi C$ on $J$, where $C$ is the matrix with entries*

$$\begin{aligned} c_{ij} &= m_{ij}/m_{jj}, & i &\neq j \\ &= 0, & i &= j, \end{aligned}$$

*indeed*

$$\lim_{n\to\infty} n^{r-1} \left\| \sum_{k=0}^{n-1} \varphi P^k - \psi \right\| = 0; \tag{5.5}$$

*the limit signed measure $\psi$ has total mass $\psi 1$ equal to zero, and $\varphi$ may be recovered from $\psi$ through the identity*

$$\varphi=\psi(I-P). \tag{5.6}$$

*Remarks.* (i) We note that by the assumption that $P$ has recurrence times with finite moment of order $r$, the hypothesis that $|\varphi|({}^jU)^r 1$ is finite holds if and only if $m^{(r)}_{|\varphi|j}=\sum_{i\in j}|\varphi_i|\, m^{(r)}_{ij}$ is finite, and that this condition is satisfied if either $\varphi$ has finite support or if $P$ also has recurrence times with finite moment of order $r+1$ and $|\varphi|$ is bounded by a multiple of the invariant probability $\pi$ (see (1.20) and (1.21)).

(ii) To fit in with the notation of recurrent potential theory the sums converging to $\psi$ are from 0 to $n-1$ here, not from 1 to $n$ as they were in Theorem 1. This creates the identity (5.6) and the zeros along the diagonal of the $C$ matrix.

*Proof.* For $r=0$ the theorem states nothing more than (5.1). For $r\geqq 1$ define probabilities $\lambda$ and $\mu$ on $J$ by $\lambda=\varphi^+/(\varphi^+ 1)$, $\mu=\varphi^-/(\varphi^- 1)$. The results of the theorem can now be simply read off from Theorems 1 and 2 by the linearity of the operators $P^n$, the final observation (5.6) coming from the fact that $P$ is a contraction on the space of signed measures so that

$$\begin{aligned}\psi(I-P)&=\lim_{n\to\infty}\varphi\sum_{k=0}^{n-1}P^k(I-P)\\&=\varphi-\lim_{n\to\infty}\varphi P^n=\varphi,\end{aligned}$$

where the limits refer to convergence in $\|\ \|$, and the last step follows from (5.1).

Theorems 1 and 2 as stated originally are essentially just the above result for $r\geqq 1$ expressed for the difference $\lambda-\mu$ of two probabilities $\lambda$ and $\mu$ on $J$. However, we see now that rather than requiring as we did in Theorem 1 that both

$$m_{\lambda j}=\sum_{i\in J}\lambda_i m_{ij} \quad \text{and} \quad m_{\mu j}=\sum_{i\in J}\mu_i m_{ij}$$

be finite, it is really only necessary to assume that

$$\sum_{i\in J}|\lambda_i-\mu_i|\, m_{ij}$$

is finite, provided that we replace $m_{\mu j}-m_{\lambda j}$ by

$$\sum_{i\in J}(\mu_i-\lambda_i)\, m_{ij},$$

with similar remarks applying to Theorem 2.

As far as I know the results of Theorem (5.2) are new apart from the case $r=0$ and the assertion of convergence of $\sum_{n=0}^{\infty}\varphi P^n$ for positive recurrent $P$. This latter is essentially just Theorem (9.50) of Kemeny, Snell and Knapp [10], with pointwise convergence of the signed measures strengthened to convergence in norm. In the potential theoretic terminology of [10], the signed measure $\psi=\sum_{n=0}^{\infty}\varphi P^n$ is the *potential* of the *charge* $\varphi$. Such limits also exist under certain circumstances for null-recurrent $P$ (see [10]), but consideration of such matters seems to be beyond the scope of the present methods.

We come now to the fact that each of our results concerning the convergence of the signed measures $\varphi P^n$ and $\sum_{k=0}^{n-1} \varphi P^k$ has a dual concerning the convergence of the functions $P^n f$ and $\sum_{k=0}^{n-1} P^k f$ for real valued functions $f$ on the state space $J$. For background on this duality theory of recurrent Markov chains the reader is referred to Section 6.2 of [10], but for present purposes it suffices to take from this source the following facts: corresponding to our irreducible, aperiodic and recurrent Markov matrix $P=(p_{ij})_{i,j\in J}$ with invariant measure $\pi=(\pi_j)$ there is a dual Markov matrix $\hat{P}=(\hat{p}_{ij})_{i,j\in J}$ defined by

$$\hat{p}_{ij}=(\pi_j/\pi_i)\, p_{ji};$$

this Markov matrix $\hat{P}$ is also irreducible, aperiodic and recurrent with the same invariant measure $\pi$, and each state $j$ in $J$ has the same recurrence time distribution under $\hat{P}$ as it does under $P$, so that in particular $\hat{P}$ has recurrence times with $r$-th moment finite if and only if $P$ does too. In our notation a quantity with a hat relates to $\hat{P}$ in the same way as the quantity without a hat would relate to $P$: thus $\hat{m}_{ij}$ is the mean first passage time from $i$ to $j$ for a Markov chain with transitions $\hat{P}$, and so on.

Let us now consider the dual of Theorem (5.2). Attention is naturally restricted to functions $f$ on $J$ which are integrable with respect to $\pi$, and we denote by $\|f\|_\pi$ the usual $L^1$ norm of $f$ in the space of all $\pi$-integrable functions on $J$: $\|f\|_\pi=\pi\,|f|$.

(5.7) **Theorem.** (*Dual of* (5.2).) *Let $r$ be a non-negative integer and suppose that $P$ has recurrence times with finite moment of order $r$. Then for all real valued functions $f$ on $J$ with $\pi$-integral $\pi f$ equal to zero and $\pi({}^jU)^r\,|f|$ finite for some (and hence every) state $j$ in $J$, we have*

$$\lim_{n\to\infty} n^r\,\|P^n f\|_\pi=0; \tag{5.8}$$

*if $P$ is positive recurrent ($r\geqq 1$) then also*

$$\sum_{n=0}^{\infty} n^{r-1}\,\|P^n f\|_\pi<\infty, \tag{5.9}$$

*and as $n$ tends to infinity the function $\sum_{k=0}^{n-1} P^k f$ converges in $\|\ \|_\pi$ to the function $g=-Gf$, where $G$ is the matrix with entries*

$$\begin{aligned} g_{ij}&=\hat{m}_{ji}/m_{jj}=[\pi({}^iU)]_j, && j\neq i\\ &=0, && j=i, \end{aligned}$$

*indeed*

$$\lim_{n\to\infty} n^{r-1}\left\|\sum_{k=0}^{n-1} P^k f-g\right\|_\pi=0; \tag{5.10}$$

*the limit function* $g$ *is such that* $\pi g=0$, *and* $f$ *may be recovered from* $g$ *through the identity*

$$f=g(I-P). \tag{5.11}$$

*Remarks.* (i) We note that because of the assumption that $P$ has recurrence times with finite moment of order $r$, the condition that $\pi({}^J U)^r|f|$ be finite is satisfied if either $f$ has finite support or if the recurrence times of $P$ also have finite moment of order $r+1$ and $f$ is bounded.

(ii) It is easy to derive the following explicit formulae for $\hat{m}_{ji}$ in terms of first passage moments of the original chain:

$$\begin{aligned}\hat{m}_{ji}&=m_{ij}+\sum_{k\in J}(m_{ki}-m_{kj})/m_{kk}\\&=m_{ij}+m_{\pi i}-m_{\pi j}\\&=m_{ij}+\frac{1}{2}\left(\frac{m_{ii}^{(2)}}{m_{ii}}-\frac{m_{jj}^{(2)}}{m_{jj}}\right),\end{aligned}$$

where the first formula holds in any positive recurrent chain but the last two are only valid if recurrence times have finite second moment.

*Proof.* One simply applies Theorem (5.2) to $\hat{P}$ with the signed measure $\varphi=(\varphi_j)=(\pi_j f_j)$ of total mass zero, for a trite calculation shows that $\|\varphi\hat{P}^n\|=\|P^n f\|_\pi$, and one finds that

$$|\varphi|({}^J\hat{U})^r 1=\pi({}^J U)^r|f| \quad \text{since} \quad (({}^J\hat{U})^r)_{ik}=\frac{\pi_k}{\pi_i}(({}^J U)^r)_{ki}$$

(see [10], Section 6.2).

The probabilistic content of the theorem above is this: if $P$ has recurrence times with finite $r$-th moment and $f$ has $\pi$-integral zero and $\pi({}^J U)^r|f|$ finite, then for all probabilities $\varphi$ on $J$ (or even signed measures $\varphi$) which are bounded by a multiple of $\pi$ (e.g. $\varphi$ with finite support),

$$\lim_{n\to\infty} n^r\varphi P^n f=0, \tag{5.12}$$

and for fixed $f$ the convergence is uniform over all $\varphi$ bounded by a particular multiple of $\pi$. (For a probability measure $\varphi$ on $J$ the quantity $\varphi P^n f$ is of course the expectation of $f(X_n)$ for a Markov chain $(X_n)$ with initial distribution $\varphi$ and stationary transition probabilities $P$.) One derives (5.12) from (5.8) by simply observing that if $|\varphi|\leqq c\pi$ then

$$|\varphi P^n f|\leqq|\varphi|\,|P^n f|\leqq c\pi|P^n f|=c\,\|P^n f\|_\pi,$$

and naturally enough the other statements in Theorem (5.7) can be given similar interpretations. It is interesting to observe that the statement above concerning the convergence in (5.12) is the dual of a similar amplification of (5.3) which does not however admit quite such an immediate probabilistic interpretation: if $P$ has recurrence times with finite $r$-th moment and $\varphi$ is a signed measure with total mass zero with $|\varphi|({}^J U)^r 1$ finite, then (5.12) holds for all functions $f$ on $J$ which are bounded (by a multiple of 1), and for fixed $\varphi$ the convergence is uniform over uniformly bounded $f$. The greatest value of these results concerning the convergence (5.12) is that for positive recurrent chains with invariant probability $\pi$ they give two dual criteria for rates of convergence of $\lambda P^n h$ to $\pi h$ for probabilities $\lambda$ and $\pi$-integrable functions $h$: one can either take $f=h-1(\pi h)$ in (5.12) with $\varphi=\lambda$ and use the first result above for functions with $\pi$-integral zero, or one can take $f=h$ and $\varphi=\lambda-\pi$ and use the second result for signed measures of total mass zero, though to obtain a useful result in either way one ends up wanting $\pi({}^jU)^r 1$ to be finite, which is tantamount to requiring that the recurrence times actually have finite moments of order $(r+1)$. We have in fact the following corollary of Theorem (5.2) and its dual (5.7):

(5.13) **Corollary.** *Let $r$ be a positive integer and suppose that $P$ is positive recurrent with invariant probability $\pi$ and recurrence times with finite moment of order $r$. Suppose that either*

(i) *$\lambda$ is a probability on $J$ such that $\lambda({}^jU)^{r-1}1$ is finite for some $j$ in $J$ and $f$ is a bounded function on $J$, or*

(ii) *$\lambda$ is a probability on $J$ bounded by a multiple of $\pi$ and $f$ is a function such that $\pi({}^jU)^{r-1}|f|$ is finite for some $j$ in $J$.*

*(In particular both conditions are satisfied if $\lambda$ is bounded by a multiple of $\pi$ and $f$ is bounded.)*

*Then*

$$\lim_{n\to\infty} n^{r-1}(\lambda P^n f-\pi f)=0,$$

*and if $r\geqq 2$ then also*

$$\sum_{n=0}^{\infty} n^{r-2}|\lambda P^n f-\pi f|<\infty$$

*and*

$$\lim_{n\to\infty} n^{r-2}\left|\lambda\left(\sum_{k=1}^{n} P^k\right)f-n\pi f-(\pi-\lambda)Df\right|=0,$$

*where $D$ is the matrix with entries $d_{ij}=m_{ij}/m_{jj}$.*

*In case* (i) *for fixed $\lambda$ the convergence is uniform over uniformly bounded $f$, while in case* (ii) *for fixed $f$ the convergence is uniform over all $\lambda$ bounded by the same multiple of $\pi$.*

*Remark.* Note that the sums in the last part are from 1 to $n$ again. This simplifies the expression for the limit signed measure $(\pi-\lambda)D$ which has $j$-th component

$$\begin{aligned}[(\pi-\lambda)D]_j&=(m_{\pi j}-m_{\lambda j})/m_{jj}\\&=[(m_{jj}^{(2)}+m_{jj})/2m_{jj}-m_{\lambda j}]/m_{jj}.\end{aligned}$$

*Proof.* As outlined above.

It may be observed that under condition (i) the corollary is just another way of expressing Corollary 2 in the introduction, and the result under condition (ii) is dual. We mention one final corollary of Theorems (5.2) and (5.7) which explicitly concerns the limiting behaviour of the $n$-step transition probabilities $p_{ij}^{(n)}$ and illustrates how these theorems really do tell us more than Corollary (5.13).

(5.14) **Corollary.** *Let $r$ be a positive integer and suppose that $P$ has recurrence times with $r$-th moments finite. Then for all states $i, j, k$ and $l$ in $J$,*

$$\lim_{n\to\infty} n^r (p_{ij}^{(n)}/\pi_j - p_{kl}^{(n)}/\pi_l) = 0$$

*and*

$$\sum_{n=1}^{\infty} n^{r-1} |p_{ij}^{(n)}/\pi_j - p_{kl}^{(n)}/\pi_l| < \infty.$$

*Remark.* The point is that from (5.13) we only know that

$$\lim_{n\to\infty} n^{r-1} (p_{ij}^{(n)}/\pi_j - 1) = 0.$$

*Proof.* Taking $\varphi = \delta_i - \delta_k$ and $f = \delta_j/\pi_j$ in (5.12) and using the second criterion for convergence derived from (5.3) we have

$$\lim_{n\to\infty} n^r (p_{ij}^{(n)}/\pi_j - p_{kj}^{(n)}/\pi_j) = 0,$$

while taking $\varphi = \delta_k$, $f = \delta_j/\pi_j - \delta_l/\pi_l$ in (5.12) and using the first criterion derived from (5.8) we have

$$\lim_{n\to\infty} n^r (p_{kj}^{(n)}/\pi_j - p_{kl}^{(n)}/\pi_l) = 0.$$

The first result above now follows on adding and the second result is proved similarly from (5.4) and (5.9).

We conclude this section by examining the possibility of obtaining converses to our theorems above. Let us consider Theorem (5.2). The strongest conclusion of the theorem is evidently the convergence of the series of norms (5.4), and one might at least hope that this implied the hypotheses of the theorem. However any such hopes are dashed by the counterexample considered below. This example provides an irreducible, aperiodic and recurrent Markov matrix $P$ which has recurrence times with finite moments of all orders, together with a signed measure $\varphi$ of total mass zero such that

$$\sum_{i\in J} |\varphi_i| \, m_{ij} = \infty, \qquad j \in J,$$

but such that

$$\varphi P^n = 0, \qquad n \geqq 1.$$

(5.15) *Example.* Let $Q = (q_{ij})$ be the Markov matrix with state space the non-negative integers $\mathbb{N}$ which was used in the proof of the renewal Theorem (2.14) for the distribution $(f_k) = (1/2^k)$:

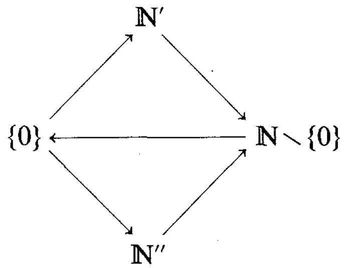
$$q_{0j} = 1/2^{j+1}, \qquad j \geqq 0,$$

$$q_{j,j-1} = 1, \qquad j \geqq 1.$$

It is obvious that $Q$ is irreducible and aperiodic, and that $Q$ has recurrence times with finite moments of all orders. Let $\mathbb{N}'$ and $\mathbb{N}''$ be two further copies of the non-negative integers $\mathbb{N}$, with non-negative integers $j$ in $\mathbb{N}$ corresponding to $j'$ in $\mathbb{N}'$ and $j''$ in $\mathbb{N}''$. Now on the countable state space $J = \mathbb{N} \cup \mathbb{N}' \cup \mathbb{N}''$ define a Markov matrix $P$ by

$$\begin{aligned} &p_{0j}=0, && j\in\mathbb{N}\\ &p_{0j'}=p_{0j''}=\tfrac{1}{2}q_{0j}=1/2^{j+2}, && j\in\mathbb{N}\\ &p_{j'j}=p_{j''j}=1, && j\in\mathbb{N}\\ &p_{j,j-1}=1, && j\in\mathbb{N}\setminus\{0\}. \end{aligned}$$

Roughly speaking, a Markov chain with transitions $P$ may be thought of as a Markov chain with transitions $Q$ which has been altered by putting in shunts through one or other of the extra copies of $\mathbb{N}$ at each transition away from zero of the original chain: each transition $0\to j$ is altered with probability half to $0\to j'\to j$ and with probability half to $0\to j''\to j$, the alterations being made independently of each other and of the original chain. We have the following communication diagram for $P$ which makes it obvious that $P$ is irreducible and aperiodic, and that $P$ too has recurrence times with finite moments of all orders:




We now take a distribution $(a_j)$ on $\mathbb{N}$ such that $\sum_{j=1}^{\infty} j a_j=\infty$, say $a_0=0$, $a_j=1/j(j+1)$, $j\geqq 1$, and define a signed measure $\varphi$ on $J$ with total mass zero by

$$\begin{aligned} &\varphi_j=0, && j\in\mathbb{N},\\ &\varphi_{j'}=a_j, && j'\in\mathbb{N}',\\ &\varphi_{j''}=-a_j, && j''\in\mathbb{N}''. \end{aligned}$$

Then since evidently

$$m_{j'0}=m_{j''0}=m_{j0}+1=j+1, \qquad j\in\mathbb{N}\setminus\{0\},$$

we have

$$\sum_{i\in J}|\varphi_i|\, m_{i0}=2\sum_{j=0}^{\infty} a_j(j+1)=2\sum_{j=1}^{\infty}\left(\frac{1}{j}\right)=\infty, \tag{5.16}$$

while $\varphi P$ is quite obviously the zero signed measure.

The Markov chain constructed above also provides a counterexample to an erroneous result of Kemeny, Snell and Knapp (Theorem (9.53) in [10], the assertion that $g=-Gf$). Their claim is equivalent by duality to the assertion that if $\beta$ is a signed measure of total mass zero then $\beta$ is the potential of the charge $\alpha=\beta(I-P)$, and that $\beta=-\alpha C$ where $C$ was defined in Proposition (5.2). The first assertion is certainly correct, but the second is not, as we see by taking $P$ as in the example above with $\beta=\varphi$, when $\alpha=\varphi(I-P)=\varphi$ too, but $\alpha C=\varphi C$ is undefined by (5.16).

The error in their proof occurs in the second last sentence, where it is tacitly assumed that $A({}^0Nf)$ equals $(A^0N)f$. However, only the first of these products is necessarily well defined, and the associative rule cannot therefore be invoked.

Confining attention to the case $r=1$ of Theorem (5.2) it is worth observing that the kind of behaviour in Example (5.15) is only possible if $\sum_{i\in J}\varphi_i^+ m_{ij}$ and $\sum_{i\in J}\varphi_i^- m_{ij}$ are both infinite. If only one of these quantities is infinite, say the first, then it is easy to deduce from Lemma (3.11) that for all $j$ in $J$

$$\lim_{n\to\infty}\left(\sum_{k=0}^{n-1}\varphi P^k\right)_j=-\infty,$$

hence certainly

$$\sum_{n=0}^{\infty}\|\varphi P^n\|=\infty.$$

However whether or not this implies

$$\limsup_{n\to\infty} n\,\|\varphi P^n\|>0$$

I do not know. It is also possible to make similar observations for $r=2, 3, \ldots$, and one can thus formulate partial converses to Theorems 1 and 2 and Corollary 1 in the Introduction, but this is left to the reader.

## 6. Applications to Renewal Theory

Let $Y_0, Y_1, \ldots$ be a sequence of independent $\mathbb{N}$-valued random variables defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$, and suppose that the initial random variable $Y_0$ has distribution $(a_n)$ and that the later random variables $Y_1, Y_2, \ldots$ have identical distribution $(f_n)$ with $f_0=0$. Setting

$$S_m=Y_0+\cdots+Y_m,\qquad m\in\mathbb{N}, \tag{6.1}$$

we say that $(S_m, m\in\mathbb{N})$ is a *delayed renewal process* with *delay distribution* $(a_n)$ and *recurrence distribution* $(f_n)$, and putting

$$v_n=\mathbb{P}\{S_m=n \text{ for some } m\in\mathbb{N}\},\qquad n\in\mathbb{N}, \tag{6.2}$$

we say that $(v_n)$ is the *delayed renewal sequence* associated with the delay distribution $(a_n)$ and the recurrence distribution $(f_n)$. The sequence $(v_n)$ is the obviously unique solution to the renewal equation

$$v_n=a_n+\sum_{k=0}^{n} v_k f_{n-k},\qquad n\in\mathbb{N}, \tag{6.3}$$

and $(v_n)$ is also determined by

$$v_n=\sum_{k=0}^{n} a_k u_{n-k},$$

where $(u_n)$ is the zero delay renewal sequence associated with $(f_n)$ through (6.3) with $a_0=1$, $a_n=0$, $n=1, 2, \ldots$. For background on this set up the reader is referred to Feller [6], XIII.10. We consider here various limit theorems for delayed renewal sequences $(v_n)$ which extend those to be found in Feller [5] and [6] and Karlin [9]. These results can either be interpreted probabilistically through (6.2), or alternatively they can be viewed as purely analytic theorems concerning the limit behaviour of the convolution equation (6.3), where it is assumed that $f_n \geqq 0$, $a_n \geqq 0$, and

(6.4) $$\sum_{n=1}^{\infty} f_n = 1, \quad \sum_{n=0}^{\infty} a_n = 1.$$

It may be noted that by arguing as in Feller [6] it is a simple matter to extend results to the more general situation where only the sequence $(f_n)$ is required to be positive and the assumptions (6.4) are dropped. We shall also assume from now on that $(f_n)$ is such that

$$\text{g.c.d.}\{n: f_n > 0\} = 1,$$

but again it is well known how to reduce questions for periodic renewal sequences to this aperiodic case (see [6]).

The connection between Markov chains and renewal theory is extremely close. Indeed if $(X_n)$ is an irreducible and recurrent Markov chain with initial distribution $\lambda$ and transition matrix $P$, then the sequence of times that $(X_n)$ visits a particular state $i$ forms a delayed renewal process with delay distribution the distribution of the first passage time to state $i$, recurrence distribution the recurrence time distribution of the state $i$, and associated delayed renewal sequence the sequence of transition probabilities $(p_{\lambda i}^{(n)})_{n \in \mathbb{N}}$. Conversely we see below that every delayed renewal process can be viewed as the sequence of times that its own forward recurrence time chain visits state zero, and there is thus a one to one correspondence between theorems concerning sequences $(p_{\lambda i}^{(n)})_{n \in \mathbb{N}}$ of Markov chain transition probabilities to a fixed state $i$ and theorems concerning delayed renewal sequences $(v_n)_{n \in \mathbb{N}}$ (cf. Kingman [13], Theorems 1.1 and 1.6, (xiv)).

Consider now our delayed renewal process $(S_m)$ defined by (6.1). Let $V_n$ be the indicator function of the event $\{S_m = n \text{ for some } m \in \mathbb{N}\}$, so that from (6.2)

$$v_n = \mathbb{P}(V_n = 1), \quad n \in \mathbb{N}.$$

In the terminology of Kingman [13] the process $(V_n)$ of zeros and ones is a *delayed regenerative phenomenon.* The forward recurrence time random variables $F_n$ are defined by

$$F_n = \inf\{k: k > 0,\ V_{n+k} = 1\}, \quad n \in \mathbb{N}.$$

If $V_m = 1$ we say there is a *renewal at time m,* and thus $F_n$ is the time which elapses between time $n$ and the time of the first renewal at or after time $n$. The sequence $(F_n)$ is a Markov chain with state space $\mathbb{N}$ and transition matrix $P$ given by

$$\begin{aligned} p_{0j} &= f_{j+1}, \quad j \in \mathbb{N}, \\ p_{ij} &= 1 \quad \text{if } i \geqq 1,\ j = i-1 \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

We clearly have that

(6.5) $$\{F_n = 0\} = \{V_n = 1\}, \quad n \in \mathbb{N},$$

and $(S_m)$ is just the sequence of times that the Markov chain $(F_n)$ visits state zero. Furthermore, $F_0 = S_0$ so that the delay distribution $(a_n)$ is not only the distribution of the first passage time to zero of the forward recurrence time chain $(F_n)$, but also its initial distribution. This fact is particularly useful since in view of (6.5) we have simply

(6.6) $$v_n = p_{a0}^{(n)},$$

and thus we are immediately in a position to apply the Markov chain theorems of the previous sections to obtain limit theorems for the most general delayed renewal sequence $(v_n)$. We observe that if $M = \sup\{n: f_n > 0\} = \infty$ then the forward recurrence time chain is irreducible, aperiodic and recurrent so that our Markov chain theorems can be applied directly. If on the other hand $M$ is finite the state space of the forward recurrence time chain decomposes into an irreducible, aperiodic and positive recurrent class $\{0, 1, \ldots, M\}$ and a transient class $\{M+1, M+2, \ldots\}$, with the probability of absorption into the recurrent class being one regardless of the initial distribution. This creates no real difficulties however since the theorems of the previous sections are easily adapted to cover this situation, but details are omitted.

The first result is of course the renewal theorem (see Feller [6], XIII.10):

(6.7) **Theorem.** *Whatever the delay distribution* $(a_n)$,

$$\lim_{n\to\infty} v_n = 1/\mu,$$

*where* $\mu = \sum_{n=1}^{\infty} n f_n$ *is the mean of the recurrence distribution* $(f_n)$ *and* $1/\mu$ *is taken to be zero if* $\mu = \infty$.

*Proof.* Just as for (2.14), one simply applies Kolmogorov's theorem to the forward recurrence time chain with initial distribution $(a_n)$, exploiting (6.6).

If we go back through the proof of the Kolmogorov theorem given in Section 2 we find that the argument being used here to establish the renewal theorem is essentially this: On a suitable probability space $(\Omega, \mathscr{F}, \mathbb{P})$ we define two independent delayed renewal processes, both with recurrence distributions $(f_n)$ but with delay distributions $(a_n)$ and $(a'_n)$ respectively, and we let $(v_n)$ and $(v'_n)$ be the associated delayed renewal sequences. Let $T$ be the time of the first simultaneous renewal. Then it is easy to see directly that just as for (2.6) we have

(6.8) $$|v_n - v'_n| \leqq 2\mathbb{P}(T > n).$$

If the mean recurrence time $\mu$ is finite then the second delay distribution $(a'_n)$ can be chosen so as to make the sequence $(v'_n)$ identically equal to a constant (necessarily $1/\mu$), by taking

(6.9) $$a'_n = (1/\mu) \sum_{k=n+1}^{\infty} f_k, \qquad n \in \mathbb{N}$$

(this is of course the invariant probability for the forward recurrence time chain). The renewal theorem (6.7) then follows from (6.8) since $T$ must be finite with probability one by virtue of part (i) of the following proposition:

(6.10) **Proposition.** *Suppose that on a probability space there are defined two independent delayed renewal processes with the same aperiodic recurrence distribution* $(f_n)$, *but with possibly different delay distributions. Let* $T$ *be the time of the first simultaneous renewal. Then*

(i) *if the recurrence distribution has finite mean then* $T$ *is finite with probability one.*

(ii) *For positive integers* $r$, $T$ *has finite moment of order* $r$ *if and only if both the delay distributions and the recurrence distribution have finite moment of order* $r$.

*Proof.* These results follow immediately from (2.6) and (4.15) applied to the forward recurrence time chain.

Another way to prove (6.10)(i) is to observe that if $(S_m)$ and $(S'_m)$ denote the two independent delayed renewal processes, then the time $T$ is certainly earlier than $S_N$, where $N$ is the first $m$ such that $S_m = S'_m$. But $N$ is just the time of the first visit to zero of the aperiodic random walk on the integers $(S_m - S'_m, m \in \mathbb{N})$ which has mean zero increments if $(f_n)$ has finite mean, whence $N$ is then a.s. finite by a well known theorem of Chung and Fuchs (see Chung [3], 8.3). This furnishes a direct proof of the renewal Theorem (6.7) for $\mu < \infty$ which avoids the use of the forward recurrence time chain, but unfortunately there does not seem to be any such simple argument when $\mu = \infty$. Part (ii) of (6.10) leads of course through (6.8) to rates of convergence in the renewal theorem, and indeed (6.10)(ii) is equivalent to the case of Proposition (4.15) when $\eta$ is of product form $\eta_1 \times \eta_2$, the result which is so important for our proofs of Theorems 1 and 2. It would be most desirable to prove (6.10)(ii) directly without recourse to the rather devious Markov chain methods employed in the proof of (4.15), but I have been unable to do this.

Let us now consider rates of convergence in the renewal theorem (6.7) when the aperiodic recurrence distribution $(f_n)$ has finite mean $\mu$. For the rest of the section let $(a_n)$ and $(a'_n)$ be two different delay distributions with finite means $\mu_a$ and $\mu'_a$, and let $(v_n)$ and $(v'_n)$ be the delayed renewal sequences associated with $(f_n)$ through $(a_n)$ and $(a'_n)$. The central result is the general Theorem (6.11) below which compares the asymptotic behaviour of the differently delayed renewal sequences $(v_n)$ and $(v'_n)$. Note that the quantity $\sum_{k=1}^{n} v_k$ is the expected number of renewals after time zero and up to and including time $n$ for delay distribution $(a_n)$.

(6.11) **Theorem.** *If* $\mu_a$, $\mu'_a$ *and* $\mu$ *are all finite, then*

$$\lim_{n\to\infty} n\,|v_n - v'_n| = 0, \tag{6.12}$$

$$\lim_{n\to\infty} \left[\sum_{k=1}^{n} v_k - \sum_{k=1}^{n} v'_k\right] = \sum_{n=1}^{\infty} (v_n - v'_n) = (\mu'_a - \mu_a)/\mu, \tag{6.13}$$

*and the central series is absolutely convergent. Let* $r$ *be a positive integer. Generalising the above results for* $r=1$ *we have that if* $(a_n)$, $(a'_n)$ *and* $(f_n)$ *all have finite moment of order* $r$, *then*

$$\lim_{n\to\infty} n^r\,|v_n - v'_n| = 0, \tag{6.14}$$

$$\sum_{n=1}^{\infty} n^{r-1}\,|v_n - v'_n| < \infty, \tag{6.15}$$

*and*

(6.16) $$\lim_{n\to\infty} n^{r-1}\left[\sum_{k=1}^{n} v_k-\sum_{k=1}^{n} v'_k-(\mu'_a-\mu_a)/\mu\right]=0.$$

*Proof* is a straightforward application of Theorem 2 to the forward recurrence time chain, using (6.6). Alternatively the results can be derived directly from (6.8) and (6.10)(ii).

The following corollary to Theorem (6.11) corresponds to Theorem (5.2) applied to the forward recurrence time chain:

(6.17) **Corollary.** *If $\mu$ is finite and $(b_n)$ is a sequence of real numbers with*

$$\sum_{n=0}^{\infty} b_n=0,\qquad \sum_{n=1}^{\infty} n\,|b_n|<\infty,\qquad \sum_{n=1}^{\infty} n\,b_n=\mu_b,$$

*then the unique solution $(w_n)$ to the renewal equation*

$$w_n=b_n+\sum_{k=0}^{n} w_k f_{n-k},\qquad n=0,1,\ldots,$$

*is such that*

$$\lim_{n\to\infty} n\,w_n=0,$$

$$\sum_{n=1}^{\infty} w_n=-\mu_b/\mu,$$

*the series is absolutely convergent, and the obvious analogues of* (6.14), (6.15) *and* (6.16) *hold if $(f_n)$ has finite r-th moment and* $\sum_{n=1}^{\infty} n^r\,|b_n|<\infty$.

*Proof.* Let $B=\sum_{n=0}^{\infty} b_n^+=\sum_{n=0}^{\infty} b_n^-$, and then apply Theorem (6.11) with $a_n=b_n^+/B$, $a'_n=b_n^-/B$.

Taking $b_n=a_n-a'_n$ for an arbitrary pair of delay distributions $(a_n)$ and $(a'_n)$, Corollary (6.17) makes it clear that it is unnecessary to assume for (6.12) and (6.13) that both $(a_n)$ and $(a'_n)$ have finite means $\mu_a$ and $\mu'_a$; rather it need only be assumed that $\sum_{n=1}^{\infty} n\,|a_n-a'_n|$ is convergent, provided $\mu'_a-\mu_a$ is replaced by $\sum_{n=1}^{\infty} n(a'_n-a_n)$, and similar remarks apply to (6.14)–(6.16).

Given an arbitrary delay distribution $(a_n)$ with finite mean we may take the second delay distribution $(a'_n)$ in Theorem (6.11) to be given by $a'_0=0$, $a'_n=a_{n-1}$, $n\geqq 1$, when we get $v'_0=0$, $v'_n=v_{n-1}$, $n\geqq 1$, and we obtain the following corollary:

(6.18) **Corollary.** *Let r be a positive integer. If $(a_n)$ and $(f_n)$ both have finite moment of order r then*

(6.19) $$\lim_{n\to\infty} n^r\,|v_n-v_{n-1}|=0,$$

*and*

(6.20) $$\sum_{n=1}^{\infty} n^{r-1}\,|v_n-v_{n-1}|<\infty.$$

*Proof.* Immediate.

This corresponds of course to the Markov chain result expressed as Corollary 2 in the Introduction. Perhaps surprisingly, the only known case of (6.18) seems to be the result (6.20) for the zero delay renewal sequence $(u_n)$ with $r=1$ (see (3.14)). Even the corresponding case of (6.19) does not seem to be known: if $(u_n)$ is the renewal sequence associated with an aperiodic distribution $(f_n)$ with finite mean, then

$$\lim_{n\to\infty} n\,|u_n-u_{n-1}|=0.$$

Finally, taking $(a'_n)$ in Theorem (6.11) to be the stationary delay distribution (6.9) which makes $v'_n \equiv 1/\mu$, we have the following result corresponding to Corollary 1 in the Introduction.

(6.21) **Corollary.** *Let $r \geqq 2$ be a positive integer. If $(a_n)$ has finite moment of order $r-1$ and $(f_n)$ has finite moment of order $r$, then*

$$\lim_{n\to\infty} n^{r-1}\,|v_n-1/\mu|=0,$$

$$\sum_{n=1}^{\infty} n^{r-2}\,|v_n-1/\mu|<\infty,$$

*and*

$$\lim_{n\to\infty} n^{r-2}\left|\sum_{k=1}^{n} v_k-n/\mu-[(\mu^{(2)}+\mu)/2\mu-\mu_a]/\mu\right|=0,$$

*where $\mu^{(2)}$ is the second moment of $(f_n)$.*

*Proof.* Immediate.

The results of Corollary (6.21) are due to Feller [5] ($r=2$) and Karlin [9] ($r \geqq 3$), both these authors basing their proofs on power series arguments involving Wiener's theorem on the reciprocal of an absolutely convergent Fourier series. Sharper results for the zero delay renewal sequence $(u_n)$ were obtained by Stone in [19] using different Fourier analytic techniques, but it does not seem to be possible to achieve Stone's results with the present methods. In fact both Stone's and Karlin's results apply to distributions $(f_n)$ on the whole set of integers and these authors also have analogues of Corollary (6.21) for renewal theory on the line, but unfortunately these generalisations seem to be quite beyond the scope of the present probabilistic methods.

The emphasis here on Theorem (6.11) being the basic result rather than the known Corollary (6.21) is quite important: neither Theorem (6.11) nor either of its Corollaries (6.17) and (6.18) could possibly be derived from Corollary (6.21), since for example the former all give results when $\mu<\infty$ but $\mu^{(2)}=\infty$, whereas the latter does not. As far as I know both Theorem (6.11) and its Corollaries (6.17) and (6.18) are new, though Theorem (6.11) could no doubt also be established using the techniques of Karlin and Stone, and this would lead to generalisations to distributions on the whole set of integers and to analogues for renewal theory on the line.

*Acknowledgement.* I would like to thank Dr. T. P. Speed for his helpful advice during the preparation of this paper.

## References


1. Breiman, L.: Probability and Stochastic Processes. Boston: Houghton Mifflin 1969
2. Chung, K. L.: Markov Chains with Stationary Transition Probabilities. 2nd ed. Heidelberg-Berlin-New York: Springer 1967
3. Chung, K. L.: A Course in Probability Theory. New York: Harcourt, Brace and World 1960
4. Erdös, P., Feller, W., Pollard, H.: A theorem on power series. Bull. Amer. Math. Soc. **55**, 201–204 (1949)
5. Feller, W.: Fluctuation theory of recurrent events, Trans. Amer. Math. Soc. **67**, 98–119 (1949)
6. Feller, W.: An Introduction to Probability Theory and its Applications, Vol. 1, 3rd ed. New York: Wiley 1968
7. Ferrar, W. L.: A Textbook of Convergence. Oxford: Clarendon 1938
8. Freedman, D.: Markov Chains. San Francisco: Holden-Day 1971
9. Karlin, S.: On the renewal equation. Pacific J. Math. **5**, 229–257 (1955)
10. Kemeny, J. G., Snell, J., Knapp, A. W.: Denumerable Markov Chains. Princeton: Van Nostrand 1966
11. Kendall, D. G.: Geometric ergodicity and the theory of queues. Math. Methods in Social Science 176–195 (1959)
12. Kingman, J. F. C.: The exponential decay of Markov chain transition probabilities. Proc. Lond. Math. Soc. **13**, 337–358 (1963)
13. Kingman, J.F.C.: Regenerative Phenoma. New York: Wiley 1972
14. Kolmogorov, A. N.: Anfangsgründe der Theorie der Markoffschen Ketten mit unendlich vielen möglichen Zuständen. Mat. Sb. 607–610 (1936); Bull. Univ. Moscow 1 (1937)
15. Orey, S.: An ergodic theorem for Markov chains. Z. Wahrscheinlichkeitstheorie verw. Gebiete **1**, 174–176 (1962)
16. Pitman, J. W.: Some identities for stopping times of a Markov chain. Univ. of Sheffield, Dept. Probability and Statistics Research Report 122 (1972)
17. Pitman, J. W.: Ph. D. Thesis, Univ. of Sheffield (1974)
18. Scheffé, H.: A useful convergence theorem for probability distributions. Ann. Math. Statist. **18**, 434–438 (1947)
19. Stone, C.: On characteristic functions and renewal theory. Trans. Amer. Math. Soc. **120**, 327–342 (1965)
20. Vere-Jones, D.: Geometric ergodicity in denumberable Markov chains. Quart. J. Math. Oxford Ser. **13**, 7–28 (1962)


James W. Pitman
Department of Statistics
University of California
Berkeley, California 94720/USA

*(Received January 4, 1974)*