

GEOMETRIC ERGODICITY IN A CLASS  
OF DENUMERABLE MARKOV CHAINS

by

H. D. Miller

University of North Carolina, Chapel Hill

and

Birkbeck College, London

Institute of Statistics Mimeo Series No. 431

June 1965

We study the question of geometric ergodicity in a class of Markov chains on the state space of non-negative integers for which, apart from a finite number of boundary rows and columns, the elements  $p_{jk}$  of the one-step transition matrix are of the form  $c_{k-j}$  where  $\{c_k\}$  is a probability distribution on the set of integers. Such a process may be described as a general random walk on the non-negative integers with boundary conditions affecting transition probabilities into and out of a finite set of boundary states. The imbedded Markov chains of several non-Markovian queueing processes are special cases of this form. It is shown that there is an intimate connection between geometric ergodicity and geometric bounds on one of the tails of the distribution  $\{c_k\}$ . This connection is explored fully.

This research was supported by the office of Naval Research Contract No. Nonr-855(09).

DEPARTMENT OF STATISTICS  
UNIVERSITY OF NORTH CAROLINA  
Chapel Hill, N. C.

GEOMETRIC ERGODICTY IN A CLASS  
OF DENUMERABLE MARKOV CHAINS

by

H. D. Miller

University of North Carolina, Chapel Hill  
and  
Birkbeck College, London

1. Introduction

Consider a homogeneous, irreducible, aperiodic Markov chain with a countable number of states identified by the non-negative integers. We denote the transition probability matrix by  $P = (p_{jk})$ , where  $p_{jk}$  ( $j, k = 0, 1, 2, \dots$ ) is the one-step transition probability from state  $j$  to state  $k$ . Let  $P^n = (p_{jk}^{(n)})$  be the matrix of  $n$ -fold transition probabilities. It is well known (see, eg., Chung (3)) that for each  $j, k$  the limit

$$\lim_{n \rightarrow \infty} p_{jk}^{(n)} = \pi_k$$

exists; this limit is positive for all pairs  $j$  and  $k$  if the chain is ergodic and zero if the chain is null-recurrent or transient. The chain is said to be geometrically ergodic (Kendall (7)) if for each pair of states  $j, k$  the rate of approach of  $p_{jk}^{(n)}$  to its limit is geometrically fast. More precisely, the chain is geometrically ergodic when numbers  $M_{jk}$  and  $\rho_{jk}$  exist such that

$$0 \leq M_{jk} < \infty, \quad 0 \leq \rho_{jk} < 1, \quad (1.1)$$
$$|p_{jk}^{(n)} - \pi_k| \leq M_{jk} \rho_{jk}^n \quad (n = 0, 1, 2, \dots),$$

for all pairs of states  $j$  and  $k$ . Kendall showed that the property of geometric ergodicity is a class property of an irreducible set of states in the sense that the geometric rate of approach for one state implies that for all pairs of states. More precisely again, an irreducible aperiodic Markov chain will

be geometrically ergodic if and only if

$$|p_{\infty}^{(n)} - \pi_0| < M \rho^n$$

for some finite non-negative  $M$  and some  $\rho$  satisfying  $0 \leq \rho < 1$ . State 0 is here meant to represent any given state, the choice of 0 being a matter of labelling only.

Vere-Jones (13) went further and showed that the rate parameters  $\rho_{jk}$  in (1.1) may all be replaced by a single parameter  $\rho$  ( $0 \leq \rho < 1$ ) uniform for all pairs of states.

Kendall (8) and Vere-Jones (14) examine the question of geometric ergodicity for some particular Markov chains, namely the imbedded Markov chains of certain queueing process. For example, Kendall considers the queueing system  $M/G/1$ . In this system there is a single server; customers, arriving in a Poisson process of rate  $\beta$ , are served in order of arrival. The service times of successive customers are independent, identically distributed random variables with distribution function  $S(\cdot)$  which is assumed to have a finite non-zero mean, conveniently taken to be the unit of time. It is further assumed that  $S(0+) = 0$ . The Poisson rate  $\beta$  is in fact the traffic intensity, i.e., the ratio of the mean service time to the mean inter-arrival time. If we consider the number of customers present (waiting or being served) immediately after each successive departure then this number forms a Markov chain for which the one-step transition matrix has the form

$$P = \begin{bmatrix} a_0 & a_1 & a_2 & a_3 & \dots \\ a_0 & a_1 & a_2 & a_3 & \dots \\ 0 & a_0 & a_1 & a_2 & \dots \\ 0 & 0 & a_0 & a_1 & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix}, \quad (1.2)$$

where

$$a_n = \int_0^{\infty} e^{-\beta x} \frac{(\beta x)^n}{n!} dS(x) \quad (n = 0, 1, 2, \dots).$$

(For a full derivation of this result see Kendall (5)).

All the imbedded Markov chains considered by Kendall and Vere-Jones have a property in common. They may all be described as being of the random walk type, by which we mean that the transition probabilities  $p_{jk}$  are, apart from a finite number of boundary rows and columns, functions of  $k-j$  only. That is, the  $p_{jk}$  are, apart from a finite number of boundary rows and columns, constant along any one diagonal of  $P$ . The matrix (1.2) in the above example clearly has this property. In general these Markov chains are random walks on the non-negative integers subject to certain boundary conditions.

The aim of the present paper is to consider the geometric ergodicity of a random walk on the non-negative integers whose increments are governed by a general distribution  $\{c_j; j = 0, \pm 1, \pm 2, \dots\}$ . The walk is subject to boundary conditions affecting one-step transition probabilities into and out of the finite set of boundary states  $(0, 1, \dots, \alpha)$ . The one-step transition matrix is of the form

$$P = \begin{bmatrix} p_{00} & p_{01} & \dots & p_{0\alpha} & p_{0,\alpha+1} & p_{0,\alpha+2} & \dots \\ p_{10} & p_{11} & \dots & p_{1\alpha} & p_{1,\alpha+1} & p_{1,\alpha+2} & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ p_{\alpha 0} & p_{\alpha 1} & \dots & p_{\alpha\alpha} & p_{\alpha,\alpha+1} & p_{\alpha,\alpha+2} & \dots \\ p_{\alpha+1,0} & \dots & \dots & p_{\alpha+1,\alpha} & c_0 & c_1 & c_2 \dots \\ p_{\alpha+2,0} & \dots & \dots & p_{\alpha+2,\alpha} & c_{-1} & c_0 & c_1 \dots \\ \dots & \dots & \dots & \dots & c_{-2} & c_{-1} & c_0 \dots \end{bmatrix} \quad (1.3)$$

Here  $p_{jk} = c_{k-j}$  for  $j > \alpha$  and  $k > \alpha$ , while otherwise the  $p_{jk}$  are arbitrary, but given, and are subject of course to the conditions

$$\sum_{k=0}^{\infty} p_{jk} = 1 \quad (j = 0, 1, \dots, \alpha) \quad (1.4)$$

$$p_{j0} + p_{j1} + \dots + p_{j\alpha} = \sum_{i=-\infty}^{\alpha-j} c_i \quad (j > \alpha) \quad (1.5)$$

The imbedded Markov chains considered by Kendall and Vere-Jones are particular cases of the above form (1.3). In addition, the form (1.3) includes the imbedded Markov chain of certain many-server queueing processes discussed by Kendall (6). As an example we consider one such process in Section 7.

As a starting point we discuss in Section 3 a particular case of (1.3), namely the random walk on the non-negative integers in which the origin acts as a natural reflecting barrier. Here the one-step transition matrix is of the form (3.5) below. The main result for a Markov chain represented by (3.5) is roughly that geometric ergodicity occurs if and only if there is some suitable kind of spatial geometric rate of decrease i.e. a geometric bound on one of the tails of the distribution  $\{c_j\}$ . This kind of result, in which, roughly speaking, spatial and temporal geometric bounds imply one another, has been given for the case of the strong law of large numbers of Baum, Katz and Read (1). These authors consider, amongst other questions, the partial sums  $S_1, S_2, \dots$ , of a sequence of independent, identically distributed random variables with common distribution function  $F(x)$  and mean  $\mu$ . They show that for given  $\epsilon$ , the probabilities

$$p_n = \Pr\left(\left|\frac{S_n}{n} - \mu\right| < \epsilon, \left|\frac{S_{n+1}}{n+1} - \mu\right| < \epsilon, \dots\right)$$

have a geometric upper bound, i.e. satisfy  $p_n \leq A\rho^n$  for some  $A \geq 0$ ,

$0 \leq \rho < 1$ , if and only if the tails of the distribution function  $F(x)$  are exponentially bounded, i.e. if and only if

$$F(-x) + 1 - F(x) \leq B e^{-\lambda x} \quad (x > 0)$$

for some  $B > 0$ ,  $\lambda > 0$ .

In Section 5 we examine the connection between spatial and temporal geometric bounds for the more general process represented by (1.3) and again we show that this connection is an intimate one. The question also occurs as to whether there is any necessary such connection for general Markov chains. A simple sufficient condition for geometric ergodicity in a general Markov chain, given in Section 8, shows that the answer to this question is in the negative.

## 2. Some preliminary definitions and results.

Let  $X$  be a one-dimensional real random variable. Define

$$X^+ = \max(0, X), \tag{2.1}$$

$$X^- = \min(0, X),$$

Then  $X = X^+ + X^-$ . Clearly, if  $E(|X|) < \infty$ , then  $E(X^+) < \infty$  and  $E(X^-) > -\infty$ .

We say that

$$E(X) = +\infty \tag{2.2}$$

if  $E(X^+) = +\infty$  and  $E(X^-) > -\infty$ ; correspondingly we say that

$$E(X) = -\infty \tag{2.3}$$

if  $E(X^-) = -\infty$  and  $E(X^+) < \infty$ .

We now define a class  $\mathcal{P}$  of regular functions by saying that a regular function belongs to  $\mathcal{P}$  if and only if its power series has non-negative coefficients and radius of convergence greater than unity.

We require the following lemma.

Lemma A Let  $X_1, X_2, \dots$  be identically distributed, independent random variables satisfying  $\Pr(X_1 > 0) > 0, \Pr(X_1 < 0) > 0$ .

Then  $\Pr(X_1 + \dots + X_n \geq 0) \leq A \rho^n$  ( $n = 1, 2, \dots$ ) for some constants  $A > 0$  and  $\rho$  ( $0 < \rho < 1$ ) if and only if  $-\infty \leq E(X_1) < 0$  and  $\Pr(X_1 \geq x) \leq B e^{-\eta x}$  ( $x > 0$ ) for some constants  $B > 0$  and  $\eta > 0$ .

Proof Suppose first that for  $A > 0$  and  $0 < \rho < 1$

$$\Pr(X_1 + \dots + X_n \geq 0) \leq A \rho^n \quad (n = 1, 2, \dots). \quad (2.4)$$

For any real numbers  $k_1, \dots, k_{n+1}$  satisfying the condition  $k_1 + k_2 + \dots + k_{n+1} = 0$ , we have that

$$\Pr(X_1 + \dots + X_{n+1} \geq 0) \geq \Pr(X_1 \geq k_1) \Pr(X_2 \geq k_2) \dots \Pr(X_{n+1} \geq k_{n+1}).$$

Choose  $\lambda > 0$  such that  $\Pr(X_1 \geq -\lambda) > \rho$  and take  $k_1 = n\lambda$ ,

$k_2 = \dots = k_{n+1} = -\lambda$ . Then

$$A \rho^{n+1} \geq \Pr(X_1 + \dots + X_{n+1} \geq 0) \geq \Pr(X_1 \geq n\lambda) (\Pr(X_1 \geq -\lambda))^n,$$

and, defining the number  $\eta > 0$  by

$$\left\{ \frac{\rho}{\Pr(X_1 \geq -\lambda)} \right\}^{1/\lambda} = e^{-\eta},$$

we have that for  $n = 1, 2, \dots$

$$\Pr(X_1 \geq n\lambda) \leq A \rho \left\{ \frac{\rho}{\Pr(X_1 \geq -\lambda)} \right\}^n = A \rho e^{-\eta n \lambda}.$$

Now for  $n\lambda \leq x < (n+1)\lambda$ ,

$$\begin{aligned} \Pr(X_1 \geq x) &\leq \Pr(X_1 \geq n\lambda) \\ &\leq A \rho e^{-\eta n \lambda} \\ &= A \rho e^{\eta(x-n\lambda) - \eta x} \\ &\leq A \rho e^{\eta \lambda - \eta x} \end{aligned}$$

since  $x - n\lambda < \lambda$ . Define  $B = A \rho e^{\eta \lambda}$  and then we have independently of  $n$  and

for all  $x > 0$ ,

$$\Pr(X_1 \geq x) \leq B e^{-\eta x} \quad (2.5)$$

It follows now that  $E(X_1^+) < \infty$ . We cannot have  $E(X_1) > 0$  for by the weak law of large numbers this would violate (2.4).

Nor can we have  $E(X_1) = 0$  for then, according to a result of Spitzer ((11), Theorem 4.1), the series  $\sum \frac{1}{n} \Pr(X_1 + \dots + X_n \geq 0)$  is divergent and this also violates (2.4). Hence we must have  $-\infty \leq E(X_1) < 0$ , and the necessity part of the lemma is therefore proved.

To prove the sufficiency part, suppose that

$$-\infty \leq E(X_1) < 0$$

and that for some constants  $B > 0$ ,  $\eta > 0$ ,

$$\Pr(X_1 \geq x) \leq B e^{-\eta x} \quad (x > 0).$$

Let  $H(x)$  be the distribution function of  $X_1$ . Then the Laplace integral

$$\int_0^{\infty} e^{tx} dH(x)$$

exists for  $t < \eta$ . Let  $\tau_+(\eta \leq \tau_+ \leq \infty)$  be the abscissa of convergence of this integral so that the integral is convergent for  $t < \tau_+$  and divergent for  $t > \tau_+$  (Widder (15)). Define  $\tau_-(\leq 0)$  to be the abscissa of convergence of the integral

$$\int_{-\infty}^0 e^{tx} dH(x) \quad .$$

It now follows that the moment generating function

$$M(t) = \int_{-\infty}^{\infty} e^{tx} dH(x) \quad (2.6)$$

exists for  $\tau_- < t < \tau_+$  and so do all its derivatives. We have that

$$M''(t) = \int_{-\infty}^{\infty} e^{tx} x^2 dH(x) > 0 \quad (\tau_- < t < \tau_+)$$

so that  $M(t)$  is a strictly convex function of  $t$ .

Since  $-\infty \leq E(X_1) < 0$  it follows that

$$-\infty \leq \lim_{t \rightarrow 0^+} M'(t) < 0$$

and so  $M(t)$  is decreasing in an interval immediately to the right of  $t = 0$ .

Hence by convexity there exists a number  $t_0$  such that

$$\begin{aligned} 0 < t_0 < \tau_+ \\ 0 < M(t_0) < 1 \\ M(t) > M(t_0) \quad (0 \leq t < \tau_+, t \neq t_0). \end{aligned}$$

Actually, it is clear from convexity that  $t_0$  is the value of  $t$  at which  $M(t)$

attains its unique minimum. Either  $M(t)$  is decreasing for  $0 < t < t_0$  and

increasing for  $t_0 < t < \tau_+$  or  $M(t)$  is strictly decreasing for

$0 < t < \tau_+ < \infty$  and  $t_0 = \tau_+$ .

Let  $H^{(n)}(x)$  be the distribution function of  $X_1 + \dots + X_n$ , so that

$$\{M(t)\}^n = \int_{-\infty}^{\infty} e^{tx} dH^{(n)}(x) \quad (\tau_- < t < \tau_+; n = 1, 2, \dots).$$

Then, since  $t_0 > 0$ , we have

$$\begin{aligned} \Pr(X_1 + \dots + X_n \geq 0) &= \int_{0^-}^{\infty} dH^{(n)}(x) \\ &< \int_{0^-}^{\infty} \exp(t_0 x) dH^{(n)}(x) \\ &< \{M(t_0)\}^n. \end{aligned}$$

Since  $0 < M(t_0) < 1$ , this completes the proof of the lemma.

Corollary. Lemma A remains true if we substitute  $\Pr(X_1 + \dots + X_n > 0)$   
for  $\Pr(X_1 + \dots + X_n \geq 0)$  and  $\Pr(X_1 > x)$  for  $\Pr(X_1 \geq x)$ .

The proof is exactly the same as that of the lemma except for the

replacement of 'greater than or equals' by 'greater than' at the appropriate places.

We shall have occasion to use taboo probabilities (Chung (3)). Let  $H$  be a given set of states in a homogeneous Markov chain  $\{Y_n; n = 0, 1, \dots\}$  whose state space may be taken to be the non-negative integers. The  $n$ -fold transition probability of reaching state  $k$  from state  $j$  under the taboo  $H$  is defined as

$$\begin{aligned} {}_H P_{jk}^{(n)} &= \Pr(Y_1 \notin H, \dots, Y_{n-1} \notin H, Y_n = k \mid Y_0 = j) \\ &\quad (n = 2, 3, \dots) \\ {}_H P_{jk}^{(1)} &= P_{jk} \end{aligned}$$

where  $P_{jk}$  is the one-step transition probability for the chain.

### 3. Geometric ergodicity in the random walk with a natural reflecting barrier (process A).

Let  $X_1, X_2, \dots$  be a sequence of independent, identically distributed, integer-valued random variables with common probability distribution  $\{c_k; k = 0, \pm 1, \pm 2, \dots\}$ . Throughout this paper we make the following assumptions about the distribution  $\{c_k\}$ .

A1 The distribution  $\{c_k\}$  is strictly two-sided i.e.

$$\Pr(X_1 < 0) > 0, \quad \Pr(X_1 > 0) > 0. \quad (3.1)$$

A2 The set of values of  $k$  for which  $c_k > 0$  does not belong to the set of multiples of a fixed integer greater than unity.

Let

$$S_n = X_1 + \dots + X_n \quad (n = 1, 2, \dots) \quad (3.2)$$

denote the partial sums of the sequence  $X_1, X_2, \dots$ . Defining  $S_0 = 0$ , we

may regard the process  $\{S_n; n = 0, 1, 2, \dots\}$  as a free random walk on the one-dimensional lattice of integers.

In this section we consider in detail, from the point of view of geometric ergodicity, a new process  $\{T_n; n = 0, 1, \dots\}$  defined as follows

$$\begin{aligned} T_0 &= j \geq 0, \\ T_n &= \max[0, T_{n-1} + X_n] \quad (n = 1, 2, \dots), \end{aligned}$$

where  $j$  is a given non-negative integer representing the point at which the process starts. The process  $T_n$  is a random walk on the non-negative integers in which the origin acts as a reflecting barrier. The origin is a natural barrier in the sense that movement off the barrier is governed essentially by the distribution  $\{c_k\}$  and not by some other given boundary condition. For the sake of brevity we refer to the process  $\{T_n\}$  as process A. Process A is clearly a Markov chain with transition probabilities

$$\begin{aligned} p_{j0} &= q_{-j} \quad (j = 0, 1, 2, \dots), \\ p_{jk} &= c_{k-j} \quad (j \geq 0, k \geq 1), \end{aligned} \tag{3.3}$$

where

$$q_k = \sum_{i=-\infty}^k c_i \tag{3.4}$$

The matrix of one step transition probabilities is thus of the form

$$P = \begin{bmatrix} q_0 & c_1 & c_2 & c_3 & \dots \\ q_{-1} & c_0 & c_1 & c_2 & \dots \\ q_{-2} & c_{-1} & c_0 & c_1 & \dots \\ q_{-3} & c_{-2} & c_{-1} & c_0 & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix}, \tag{3.5}$$

a particular case of the more general form (1.3). The assumption A1 and A2 ensure that the process is aperiodic and irreducible.

Let  $F_{\infty}(s)$  denote the generating function of first return probabilities for the origin, that is

$$F_{\infty}(s) = \sum_{n=1}^{\infty} f_{\infty}^{(n)} s^n, \quad (3.6)$$

where

$$f_{\infty}^{(n)} = \text{prob} (T_1 \neq 0, \dots, T_{n-1} \neq 0, T_n = 0 | T_0 = 0). \quad (3.7)$$

It was shown by Kendall (7) that an irreducible aperiodic Markov chain with a countable number of states is geometrically ergodic if and only if the power series (3.6) has radius of convergence greater than unity or alternatively if and only if the function

$$\frac{1-s}{1-F_{\infty}(s)} \quad (3.8)$$

is analytic for  $|s| < 1 + \delta$ , for some  $\delta > 0$ . Thus according to Kendall's first condition mentioned above, the process  $\{T_n\}$  is geometrically ergodic if and only if  $F_{\infty}(s) \in \mathcal{P}$ .

Now it follows from a theorem due to Baxter (Spitzer (12), Theorem 3.1) that  $F_{\infty}(s)$  has the following form:

$$F_{\infty}(s) = 1 - \exp \left\{ - \sum_{n=1}^{\infty} \frac{s^n}{n} \text{Pr}(S_n \leq 0) \right\} \quad (|s| < 1) \quad (3.9)$$

where  $S_n$  is the free partial sum defined at (3.2). In Spitzer's statement of Baxter's theorem there is a condition on the first moment of the distribution  $\{c_k\}$ . However, by a suitable interpretation of the generating functions involved, this condition can be seen to be unnecessary. Thus we can assume (3.9) to hold for a process governed by a quite arbitrary distribution  $\{c_k\}$ . Regarding the nature of process A we have the following result. Process A is

(i) ergodic if and only if

$$\sum_{n=1}^{\infty} \frac{1}{n} \Pr(S_n > 0) < \infty ; \quad (3.10)$$

(ii) transient if and only if

$$\sum_{n=1}^{\infty} \frac{1}{n} \Pr(S_n \leq 0) < \infty ; \quad (3.11)$$

(iii) null-recurrent if and only if the above series (3.10) and (3.11) are both divergent.

To prove this we let  $s \rightarrow 1^-$  in (3.9) and use Abel's theorem and its converse for power series with non-negative coefficients. If (3.11) holds then  $F_{\infty}(1) < 1$  which gives transience. If the series (3.11) is divergent then  $F_{\infty}(1) = 1$  and we have that the process is recurrent. Since for all  $n$ ,  $\Pr(S_n \leq 0) = 1 - \Pr(S_n > 0)$  and since

$$\exp \left\{ - \sum_{n=1}^{\infty} \frac{s^n}{n} \right\} = 1 - s \quad (|s| < 1),$$

it follows that

$$F_{\infty}(s) = 1 - (1-s) \exp \left\{ \sum_{n=1}^{\infty} \frac{s^n}{n} \Pr(S_n > 0) \right\} \quad (3.12)$$

and hence that

$$\frac{1 - F_{\infty}(s)}{1 - s} = \exp \left\{ \sum_{n=1}^{\infty} \frac{s^n}{n} \Pr(S_n > 0) \right\} \quad (3.13)$$

Assuming that  $F_{\infty}(1) = 1$ , and letting  $s \rightarrow 1^-$  in (3.13) we have that the mean recurrence time is given by

$$F'_{\infty}(1) = \exp \left\{ \sum_{n=1}^{\infty} \frac{1}{n} \Pr(S_n > 0) \right\} ,$$

a result due to Spitzer ((12), p. 158). The results (i), (ii), and (iii) therefore follow.

It was shown by Spitzer ((11), Theorem 4.1) that a sufficient condition for (3.10) to hold is that  $E(|X_1|) < \infty$  and  $E(X_1) < 0$ . Using the definitions (2.2) and (2.3) of infinite expectation, it follows that a slightly more general condition for (3.10) and hence for process A to be ergodic is a fortiori that

$$0 > E(X_1) \geq -\infty \quad . \quad (3.14)$$

Similarly a sufficient condition for the transience of process A is that

$$0 < E(X_1) \leq \infty \quad . \quad (3.15)$$

We now turn to the question of geometric ergodicity. We need only deal with the cases where process A is either ergodic or transient since a null-recurrent Markov chain, not possessing finite moments of recurrence times, can clearly not be geometrically ergodic. Note that in our terminology geometric ergodicity does not imply that the chain is ergodic; it implies that the n-fold transition probabilities converge geometrically fast to their limits. The limits will be zero in the case of a transient chain.

We now prove the following result connecting geometric ergodicity with the tails of the distribution  $\{c_k\}$ .

(i) Necessary and sufficient conditions for process A to be ergodic and geometrically ergodic are that  $-\infty \leq E(X_1) < 0$  and that  $c_k \leq C \lambda^k$  ( $k = 1, 2, \dots$ ) for some  $C > 0$  and some  $\lambda$  ( $0 < \lambda < 1$ ).

(ii) Necessary and sufficient conditions for process A to be transient and geometrically ergodic are that  $0 < E(X_1) \leq \infty$  and that  $c_{-k} \leq D \mu^k$  ( $k = 1, 2, \dots$ ) for some  $D > 0$  and some  $\mu$  ( $0 < \mu < 1$ ).

We may express this result alternatively as follows.

(i)  $F_{\infty}(s) \in \mathcal{P}$  and  $F_{\infty}(1) = 1$  if and only if  $-\infty \leq E(X_1) < 0$  and  $\sum_{k>0} c_k z^k \in \mathcal{P}$ .

(ii)  $F_{\infty}(s) \in \mathcal{P}$  and  $F_{\infty}(1) < 1$  if and only if  $0 < E(X_1) \leq \infty$  and  
 $\sum_k > 0 c_{-1} z^k \in \mathcal{P}$ .

To prove (i) suppose first that for given  $C, \lambda$  ( $C > 0, 0 < \lambda < 1$ ),

$$-\infty \leq E(X_1) < 0 \quad (3.16)$$

$$c_k \leq C \lambda^k \quad (k = 1, 2, \dots). \quad (3.17)$$

It follows from (3.17) that for some constants  $B > 0$  and  $\eta > 0$

$$\Pr(X_1 > x) \leq B e^{-\eta x}. \quad (3.18)$$

We then have from the Corollary to Lemma A that there exist constants  $A > 0$  and  $\rho$  ( $0 < \rho < 1$ ) such that

$$\Pr(S_n > 0) \leq A \rho^n$$

In fact we may take  $A = 1$  and  $\rho = M(t_0)$  where  $t_0 > 0$  is the unique value of  $t$  at which the moment generating function,

$$M(t) = \sum_{k=-\infty}^{\infty} c_k e^{kt}, \quad (3.19)$$

assumes its minimum value. It follows that the series  $\sum (1/n) \Pr(S_n > 0)$  is convergent and so the process is ergodic. Further, the power series  $\sum (s^n/n) \Pr(S_n > 0)$  has radius of convergence at least  $1/M(t_0)$  and hence the function

$$\frac{1-s}{1-F_{\infty}(s)} = \exp \left\{ - \sum_{n=1}^{\infty} \frac{s^n}{n} \Pr(S_n > 0) \right\} \quad (3.20)$$

is regular for  $0 \leq |s| < 1/M(t_0)$ . Geometric ergodicity now follows from Kendall's condition (3.8). It follows also from (3.12) that the radius of convergence of  $F_{\infty}(s)$  is at least  $1/M(t_0)$ .

Now suppose that process A is ergodic and also geometrically ergodic.

Then the function

$$\frac{1-s}{1-F_{\infty}(s)}$$

is analytic in the circle  $|s| = 1 + \delta$  for some  $\delta > 0$  and is, clearly, also free of zeros in this circle. Hence the function

$$\log\left\{\frac{1-s}{1-F_{\infty}(s)}\right\} = -\sum_{n=1}^{\infty} \frac{s^n}{n} \Pr(s_n > 0) \quad (3.21)$$

is also regular for  $|s| < 1 + \delta$ . Hence

$$\Pr(S_n > 0) < A\rho^n$$

for some  $A > 0$  and  $\rho$  ( $0 < \rho < 1$ ). The conclusions that  $-\infty \leq E(X_1) < 0$  and  $c_k \leq C\lambda^k$  now follow from the Corollary to Lemma A and the proof of (i) is complete.

The proof of (ii) is similar: suppose that

$$0 < E(X_1) \leq \infty$$

and

$$c_{-k} \leq D\mu^k \quad (k = 1, 2, \dots)$$

for some  $D > 0$  and some  $\mu$  ( $0 < \mu < 1$ ). By applying Lemma A to the sequence  $-X_1, -X_2, \dots$  we find that

$$\Pr(S_n \leq 0) \leq \{M(t_0)\}^n \quad (3.22)$$

where  $0 < M(t_0) < 1$  and  $t_0 < 0$  is the unique value of  $t$  at which  $M(t)$ , defined at (3.19), assumes its minimum value. The convergence of the series  $\sum (1/n)\Pr(S_n \leq 0)$  and hence the transience of the process now follow. From (3.9),  $F_{\infty}(s)$  is regular for at least  $|s| < 1/M(t_0)$  and so we have geometric ergodicity. Further, the radius of convergence of (3.6) is at least  $1/M(t_0)$ . This completes the proof of (ii)

4. Some further results for process A.

The results of this section are required for the following section.

As usual, let

$$F_{jk}(s) = \sum_{n=1}^{\infty} f_{jk}^{(n)} s^n \quad (j \neq k, j, k = 0, 1, 2, \dots)$$

be the generating function of the first passage probabilities  $f_{jk}^{(n)}$  ( $n=1, 2, \dots$ ) from state  $j$  to state  $k$ . Then for process A, the theorem of Baxter which gives the expression (3.9) also gives the result

$$F_{10}(s) = 1 - \exp \left\{ - \sum_{n=1}^{\infty} \frac{s^n}{n} \Pr(S_n < 0) \right\} . \quad (4.2)$$

Note that on the right hand side of (4.2) we have strict inequalities  $S_n < 0$ , whereas in (3.9) we have  $S_n \leq 0$ .

We wish to prove some inequalities concerning the generating functions  $F_{j0}(s)$  ( $j = 1, 2, \dots$ ). Suppose that

$$-\infty < \sum_{k=-\infty}^{\infty} kc_k < 0 , \quad (4.3)$$

$$\sum_{k>0} c_k z^k \in \mathcal{P}$$

so that process A is ergodic and geometrically ergodic. It then follows, as in Section 2, that the moment generating function

$$M(t) = \sum_{k=-\infty}^{\infty} c_k e^{kt} \quad (4.4)$$

exists for  $0 \leq t < \tau_+$  ( $\tau_+ \leq \infty$ ) and is strictly decreasing for  $0 < t < t_0$ , where  $t_0 (\leq \tau_+)$  is the unique value of  $t$  for which  $M(t)$  attains its minimum value. Consider the equation

$$sM(t) = 1 \quad (1 \leq s < \{M(t_0)\}^{-1}, 0 \leq t < t_0) . \quad (4.5)$$

This equation has a unique real root  $t = t_1(s)$  which satisfies

$$0 \leq t_1(s) \leq t_0 \quad (1 \leq s < \{M(t_0)\}^{-1}) \quad (4.6)$$

and  $t_1(s)$  is a strictly increasing, continuous function of  $s$  with  $t_1(1) = 0$ .

We have the following result. Suppose that the distribution  $\{c_k\}$  satisfies (4.3), i.e. that process A is ergodic and geometrically ergodic.

Then

$$F_{j_0}(s) \geq e^{jt_1(s)} \quad (1 \leq s < \{M(t_0)\}^{-1}), \quad (4.7a)$$

$$F_{j_0}(s) \leq s + \frac{s(s-1)M(t)}{1-s M(t)} e^{jt} \quad (0 \leq t < t_0, 1 \leq s < \{M(t)\}^{-1}). \quad (4.7b)$$

Consider first the proof of the inequality (4.7a). We modify the free random walk defined by (3.2) in such a way that it starts at  $j > 0$  and that the states  $0, -1, -2, \dots$  are all made absorbing states. If  $N$  is the time to absorption and if  $F_{j_0}(s)$  is a generating function for process A, then

$$F_{j_0}(s) = E(s^N) \quad (4.8)$$

An extension of the argument by Miller (10) easily shows that Wald's identity,

$$E[\{M(t)\}^{-N} \exp(tS_N)] = e^{jt}, \quad (4.9)$$

holds for  $0 \leq t < t_0$ . Here  $S_N (\leq 0)$  is the state reached at the time of absorption. We now make the substitution

$$s = \{M(t)\}^{-1} \quad (0 \leq t < t_0)$$

and obtain

$$E[s^N \exp\{t_1(s)S_N\}] = e^{jt_1(s)} \quad (1 \leq s < \{M(t_0)\}^{-1}). \quad (4.10)$$

Since  $S_N \leq 0$  and  $t_1(s) \geq 0$  we have

$$E(s^N) \geq E[s^N \exp\{t_1(s)S_N\}] = e^{jt_1(s)} \quad (1 \leq s < \{M(t_0)\}^{-1}) \quad (4.11)$$

which, by virtue of (4.8), gives the result (4.7a).

To prove (4.7b), let  $o^p_{jk}^{(n)}$  denote the taboo probabilities for process A. These are clearly identical with the transition probabilities of the modified random walk with absorbing states defined in the previous paragraph.

We clearly have that

$$\Pr(N > n) = \sum_{k=1}^{\infty} o^p_{jk}^{(n)} \quad (4.12)$$

and also that

$$F_{j0}(s) = s + (s-1) \sum_{n=1}^{\infty} s^n \Pr(N > n) \quad (|s| < 1) \quad (4.13)$$

Now if  $\{g_k^{(n)}; k = 0, \pm 1, \pm 2, \dots\}$  is the probability distribution of the unrestricted sum  $X_1 + \dots + X_n$ , we must have that

$$o^p_{jk}^{(n)} \leq g_{k-j}^{(n)} \quad (n = 1, 2, \dots; j, k > 0) .$$

Hence

$$\begin{aligned} \Pr(N > n) &= \sum_{k=1}^{\infty} o^p_{jk}^{(n)} \leq \sum_{k=1}^{\infty} g_{k-j}^{(n)} \\ &\leq \sum_{k=1}^{\infty} e^{kt} g_{k-j}^{(n)} \quad (t \geq 0) \\ &= e^{jt} \sum_{k=1}^{\infty} e^{(k-j)t} g_{k-j}^{(n)} \\ &\leq e^{jt} \sum_{k=-\infty}^{\infty} e^{(k-j)t} g_{k-j}^{(n)} = e^{jt} [M(t)]^n . \quad (4.14) \end{aligned}$$

In particular we may set  $t = t_0 > 0$  and obtain

$$\Pr(N > n) \leq e^{jt_0} [M(t_0)]^n , \quad (4.15)$$

from which it follows that the series on the right hand side of (4.13) is convergent for  $|s| < [M(t_0)]^{-1}$ . This also follows from the results of Vere-Jones (13). Hence by analytic continuation (4.13) holds for

$|s| < \{M(t_0)\}^{-1}$  and in addition we have from (4.14) and (4.15) the desired inequality (4.7b),

$$\begin{aligned} F_{j_0}(s) &\leq s + (s-1) e^{jt} \sum_{n=1}^{\infty} s^n \{M(t)\}^n \\ &= s + \frac{s(s-1)M(t)}{1-s M(t)} e^{jt} \quad (0 \leq t < t_0, 1 \leq s < \{M(t)\}^{-1}) . \end{aligned}$$

We now turn to the case where process A is transient and geometrically ergodic. In this case we have

$$\begin{aligned} 0 < \sum_{k=-\infty}^{\infty} kc_k &\leq \infty \\ \sum_{k > 0} c_{-k} z^k &\in \mathcal{P} \end{aligned} \tag{4.16}$$

and the moment generating function (4.4) now exists for  $\tau_- < t \leq 0$  ( $\tau_- \geq -\infty$ ) and is strictly increasing for  $t_0 < t < 0$  where  $t_0$  ( $\geq \tau_-$ ) is again the unique value of  $t$  for which  $M(t)$  attains its minimum value. We observe that  $t_0$  is now negative.

We have the following result. Suppose that the distribution  $\{c_k\}$  satisfies (4.16) i.e. that process A is transient and geometrically ergodic.

Then

$$F_{j_0}(s) \leq \frac{s M(t)}{1-s M(t)} e^{jt} \quad (t_0 < t \leq 0, 0 \leq s < \{M(t)\}^{-1}). \tag{4.17}$$

To prove this we observe that

$$f_{j_0}^{(n)} \leq \Pr(S_n \leq -j) \quad (j > 0),$$

where  $S_n$  is defined at (3.2). Hence

$$\begin{aligned} f_{j_0}^{(n)} &\leq \sum_{k=-\infty}^{-j} g_k^{(n)} \\ &\leq \sum_{k=-\infty}^{-j} e^{(j+k)t} g_k^{(n)} \quad (t \leq 0) \\ &\leq e^{jt} \sum_{k=-\infty}^{\infty} e^{kt} g_k^{(n)} = e^{jt} \{M(t)\}^n . \end{aligned}$$

On multiplying by  $s^n$  ( $0 < s < \{M(t)\}^{-1}$ ) and summing over  $n$ , we obtain (4.17).

5. Geometric ergodicity of random walk with imposed boundary conditions.

We now consider the question of geometric ergodicity in the more general Markov chain with transition matrix (1.3). We assume that the  $p_{jk}$  are such that the process is irreducible. For the sake of brevity we call this Markov chain process B. Logically, it would be more appropriate to consider first the relation between processes A and B in respect of transience, recurrence, null-recurrence, etc. We shall see in this section that processes A and B are recurrent together or transient together, but we postpone until the next section the question of ergodicity and null-recurrence.

To distinguish from process A we use Greek letters for the transition probabilities, etc., of process B. Thus  $\pi_{jk}^{(n)}$ ,  $\varphi_{jk}^{(n)}$ ,  $H_{jk}^{\pi^{(n)}}$  denote respectively transition, first passage and taboo probabilities for process B while

$$\Pi_{jk}(s) = \sum_{n=1}^{\infty} \pi_{jk}^{(n)} s^n, \quad \Phi_{jk}(s) = \sum_{n=1}^{\infty} \varphi_{jk}^{(n)} s^n, \quad H_{jk}^{\Pi}(s) = \sum_{n=1}^{\infty} H_{jk}^{\pi^{(n)}} s^n$$

denote the corresponding generating functions.

Throughout this section  $H$  will denote the set of states  $\{0, 1, \dots, \alpha\}$ . Our first task is to obtain some relations between the generating functions of process A and process B.

We have the following decomposition of the first passage probabilities  $\varphi_{jo}^{(n)}$  based on first returns to states in the set  $H$ . For  $j = 0, 1, \dots$ ,

$$\begin{aligned} \varphi_{jo}^{(1)} &= p_{jo} \\ \varphi_{jo}^{(n)} &= H_{jo}^{\pi^{(n)}} + \sum_{k=1}^{\alpha} \sum_{r=1}^{n-1} H_{jk}^{\pi^{(r)}} \varphi_{ko}^{(n-r)} \quad (n = 2, 3, \dots). \end{aligned} \quad (5.1)$$

By taking generating functions we obtain the following set of equations

$$\phi_{00}(s) = H_{00}^{\Pi}(s) + \sum_{k=1}^{\alpha} H_{0k}^{\Pi}(s) \phi_{k0}(s), \quad (|s| \leq 1) \quad (5.2)$$

$$\phi_{j0}(s) = H_{j0}^{\Pi}(s) + \sum_{k=1}^{\alpha} H_{jk}^{\Pi}(s) \phi_{k0}(s) \quad (j = 1, \dots, \alpha; |s| \leq 1) \quad (5.3)$$

Consider the row sums of the coefficients  $H_{jk}^{\Pi}(s)$  on the right hand side of (5.2) and (5.3). For  $j \in H$ ,

$$\begin{aligned} \sum_{k=0}^{\alpha} H_{jk}^{\Pi}(s) &= H_{jH}^{\Pi}(s) \\ &= (p_{j0} + p_{j1} + \dots + p_{j\alpha})s + s \sum_{i=\alpha+1}^{\infty} p_{ji} H_{iH}^{\Pi}(s) \end{aligned} \quad (5.4)$$

Now for  $i = \alpha+1, \alpha+2, \dots$

$$H_{iH}^{\Pi}(s) = (p_{i0} + p_{i1} + \dots + p_{i\alpha})s + s \sum_{k=\alpha+1}^{\infty} H_{ik}^P(s) (p_{k0} + \dots + p_{k\alpha}) \quad (5.5)$$

and the generating functions  $H_{ik}^P(s)$  on the right hand side are now those of process A since they are based on transitions among the states  $\alpha+1, \alpha+2, \dots$  only. Using (1.5) and (3.4) we obtain from (5.5) that

$$H_{iH}^{\Pi}(s) = q_{\alpha-i} s + s \sum_{k=\alpha+1}^{\infty} H_{ik}^P(s) q_{\alpha-k} \quad (i > \alpha). \quad (5.6)$$

The right hand side of (5.6) is easily seen to be the generating function  $F_{i-\alpha,0}(s)$  of process A. Hence it follows from (5.4) that

$$\sum_{k=0}^{\alpha} H_{jk}^{\Pi}(s) = (p_{j0} + \dots + p_{j\alpha})s + s \sum_{i=\alpha+1}^{\infty} p_{ji} F_{i-\alpha,0}(s) \quad (j \in H) \quad (5.7)$$

The relation (5.7) will be our main tool in relating the behavior of process B to that of process A.

We have the following result. Process B is ergodic and geometrically ergodic if and only if the following conditions hold:

(i) the corresponding process A is ergodic and geometrically ergodic,  
i.e. (4.3) holds;

(ii) the series

$$\sum_{k=0}^{\infty} p_{0k} z^k, \quad \sum_{k=0}^{\infty} p_{1k} z^k, \quad \dots, \quad \sum_{k=0}^{\infty} p_{\alpha k} z^k \quad (5.8)$$

all belong to the class P.

To prove this result suppose first that (i) and (ii) hold. With a view to writing (5.3) in matrix form we define the matrices

$$\Pi(s) = \begin{bmatrix} H_{11}^{\Pi}(s) & \dots & H_{1\alpha}^{\Pi}(s) \\ \vdots & & \vdots \\ H_{\alpha 1}^{\Pi}(s) & \dots & H_{\alpha\alpha}^{\Pi}(s) \end{bmatrix},$$

$$\pi(s) = \begin{bmatrix} H_{10}^{\Pi}(s) \\ \vdots \\ H_{\alpha 0}^{\Pi}(s) \end{bmatrix}, \quad \varphi(s) = \begin{bmatrix} \phi_{10}(s) \\ \vdots \\ \phi_{\alpha 0}(s) \end{bmatrix}.$$

Equations (5.3) then become

$$\varphi(s) = \pi(s) + \Pi(s) \varphi(s), \quad (5.9)$$

$$\text{i.e. } \{I - \Pi(s)\} \varphi(s) = \pi(s), \quad (5.10)$$

where I is the unit  $\alpha \times \alpha$  matrix. It follows from the irreducibility of process B that for  $0 < s \leq 1$

$$H_{jk}^{\Pi}(s) > 0 \quad (j, k \in H) \quad (5.11)$$

and so for  $0 < s \leq 1$  all the elements of the matrices  $\pi(s)$  and  $\Pi(s)$  are positive. Since we are assuming process A to be ergodic we have that  $F_{j0}(1) = 1$  ( $j = 0, 1, \dots$ ). It follows from (5.7) that

$$\sum_{k=0}^{\alpha} H^{\Pi}_{jk}(1) = 1 \quad (j, k \in H) \quad (5.12)$$

Hence the row sums of  $\Pi(1)$  are all strictly less than unity. Now the maximal positive eigenvalue of a positive matrix does not exceed the maximal row sum of the matrix (see, e.g., Debreu and Herstein (4)). Thus  $I - \Pi(1)$  is non-singular and, a fortiori, so is  $I - \Pi(s)$  ( $0 < s \leq 1$ ). Thus we may write (5.10) as

$$\varphi(s) = \{I - \Pi(s)\}^{-1} \pi(s) = \left\{ \sum_{n=0}^{\infty} \Pi^n(s) \right\} \pi(s), \quad (0 < s \leq 1). \quad (5.13)$$

where  $\Pi^0(s)$  is defined to be  $I$ .

Now (5.9) may also be written in the form

$$\begin{bmatrix} 1 \\ \varphi(s) \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ -\frac{1}{\pi(s)} & \frac{0}{\Pi(s)} & \dots & \frac{0}{\dots} \end{bmatrix} \begin{bmatrix} 1 \\ \varphi(s) \end{bmatrix}. \quad (5.14)$$

The matrix

$$\begin{bmatrix} 1 & 0 & \dots & 0 \\ -\frac{1}{\pi(1)} & \dots & \dots & \frac{0}{\Pi(1)} \end{bmatrix} \quad (5.15)$$

is a stochastic matrix all of whose elements, apart from those in the first row, are positive. It follows that the matrix (5.15) has a simple eigenvalue 1 corresponding to which there is a column eigenvector all of whose elements are equal and this eigenvector is unique up to a multiplicative constant. On setting  $s = 1$  in (5.14) we must have that

$$\begin{bmatrix} 1 \\ \varphi(1) \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \quad \text{i.e. } \phi_{j0}(1) = 1 \quad (j = 1, \dots, \alpha). \quad (5.16)$$

Thus if we set  $s = 1$  in (5.2) and (5.7) it follows from positivity and (5.16) that  $\phi_{00}(1) = 1$ . Hence process B is recurrent. To prove that process B

is ergodic and geometrically ergodic we observe that we can, by assumption, choose  $t$  ( $0 < t < t_0$ ) so that for all  $j \in H$

$$\sum_{k=0}^{\infty} p_{jk} e^{kt} < \infty. \quad (5.17)$$

Now for  $j = 0, 1, \dots$ ,  $F_{j_0}(s)$  is regular in the circle  $|s| < \{M(t_0)\}^{-1}$ . By applying the inequality (4.7b) and taking the value of  $t$  for which (5.17) holds we see that the right hand side of (5.7) is regular for  $1 \leq |s| < \{M(t)\}^{-1}$  and a fortiori for  $0 \leq |s| < \{M(t)\}^{-1}$ . Since the <sup>power series</sup> coefficients of all the functions in (5.7) are non-negative it follows that each of the functions  $H^{\Pi}_{jk}(s)$  ( $j, k \in H$ ) on the left hand side of (5.7) is regular for  $0 \leq |s| < \{M(t)\}^{-1}$ . Therefore each element of the matrix  $\Pi(s)$  is regular for  $0 \leq |s| < \{M(t)\}^{-1}$ . Since  $I - \Pi(s)$  is non-singular for  $0 \leq s \leq 1$  it follows by continuity that  $I - \Pi(s)$  is non-singular for  $0 \leq s < 1 + \delta$ , for some  $\delta > 0$ . Thus by (5.13) and non-negativity each element of  $\varphi(s)$  belongs to the class  $\mathcal{P}$ . Finally, by (5.2),  $\Phi_{00}(s) \in \mathcal{P}$  and so process  $B$  is ergodic and geometrically ergodic.

Conversely, suppose process  $B$  is ergodic and geometrically ergodic. Then  $\Phi_{j_0}(1) = 1$  for  $j = 0, 1, 2, \dots$ . Hence from (5.2) and (5.3)

$$\sum_{k=0}^{\alpha} H^{\Pi}_{jk}(1) = 1 \quad (j \in H), \quad (5.18)$$

and so from (5.7)

$$p_{j_0} + \dots + p_{j\alpha} + \sum_{i=\alpha+1}^{\infty} p_{ji} F_{i-\alpha, 0}(1) = 1 \quad (j \in H). \quad (5.19)$$

Since  $0 < F_{j_0}(1) \leq 1$  and  $\sum_{k=0}^{\infty} p_{jk} = 1$  ( $j \in H$ ) it follows from the non-negativity of the  $p_{jk}$  that  $F_{i-\alpha, 0}(1) = 1$  for all  $i$  and  $j$  for which  $p_{ji} > 0$  ( $i = \alpha+1, \alpha+2, \dots, j \in H$ ). That some such  $i$  and  $j$  exist follows from the

irreducibility of process B. Thus process A is recurrent. Since process B is geometrically ergodic the left hand side of (5.7) belongs to the class  $\mathcal{P}$ . Choose  $i > \alpha$  and  $j \in H$  such that  $p_{ji} > 0$ . From (5.7) it follows that  $F_{i-\alpha,0}(s) \in \mathcal{P}$ . Now from (4.12) and (4.13) we have

$$F_{j0}(s) = s + (s-1) \sum_{n=1}^{\infty} s^n \left( \sum_{k=1}^{\infty} o_{jk}^{(n)} \right) \quad (j = 0,1,2,\dots). \quad (5.20)$$

Since process A with zero as a taboo state has independent increments we have that

$$\sum_{k=1}^{\infty} o_{ok}^{(n)} \leq \sum_{k=1}^{\infty} o_{jk}^{(n)} \quad (j = 1,2,\dots) \quad (5.21)$$

Thus since  $F_{i-\alpha,0}(s) \in \mathcal{P}$  for some  $i > \alpha$  it follows from (5.20) and (5.21) that  $F_{00}(s) \in \mathcal{P}$ . Hence process A is geometrically ergodic and since it is recurrent it is also ergodic.

Further, from the inequality (4.7a), it follows that unless all the functions (5.8) belong to the class  $\mathcal{P}$ , the right hand side of (5.7) would, for some  $j \in H$ , be divergent for  $s > 1$ , contradicting the fact that the left hand side of (5.7) belongs to the class  $\mathcal{P}$ . The proof of our result is now complete.

Next, we deal with the question of transience and geometric ergodicity and we have the following result. Process B is transient and geometrically ergodic if and only if process A is.

The proof of the previous result shows that process B is recurrent if and only if process A is recurrent. Thus process B is transient if and only if process A is transient.

Now suppose process A is transient and geometrically ergodic. We apply inequality (4.17) to the right hand side of (5.7). Since for  $j = 0,1,\dots$   $F_{j0}(s)$  is regular in the circle  $|s| = \{M(t_0)\}^{-1}$  and since  $t < 0$  in (4.17)

it follows that the right hand side of (5.7) belongs to  $\mathcal{P}$ . So therefore does the left hand side and the same argument used for the previous result shows that  $\Phi_{00}(s) \in \mathcal{P}$ . Hence process A is transient and geometrically ergodic.

Conversely, suppose that process B is transient and geometrically ergodic. Then the left hand side of (5.7) belongs to  $\mathcal{P}$  and so for some  $i > \alpha$ ,  $F_{i-\alpha, 0}(s) \in \mathcal{P}$ . For this  $i$ , write  $k = i - \alpha$ . Choose  $m$  so that  $p_{kl}^{(m)} > 0$ . Now for  $n > m$

$$f_{ko}^{(n)} \geq p_{kl}^{(m)} f_{10}^{(n-m)},$$

from which it follows that  $F_{10}(s) \in \mathcal{P}$ . Hence from (4.2) we have that

$$\sum_{n=1}^{\infty} \frac{s^n}{n} \Pr(S_n < 0) \in \mathcal{P},$$

and so from Lemma A and from the results of Section 3 it follows that process A is transient and geometrically ergodic.

## 6. Null-recurrence and ergodicity in process B

We have seen in Section 5 that processes A and B are recurrent together or transient together. However, the relation between these processes in respect of ergodicity and null-recurrence is a more delicate question. If we assume that

$$\sum_{k=-\infty}^{\infty} |k| c_k < \infty \quad (6.1)$$

then it is possible to explore fully the relation between these processes.

If however we assume, say, that process A is ergodic with

$$E(X_1) = \sum_{k=-\infty}^{\infty} k c_k = -\infty \quad (6.2)$$

in the sense of (2.3), then it seems that the ergodicity of process B will depend on that rates at which  $c_k \rightarrow 0$  as  $k \rightarrow \infty$  and  $p_{jk} \rightarrow 0$  as  $k \rightarrow \infty$

for  $j \in H$ . We shall not explore any further the implication of (6.2) for ergodicity in process B in the absence of further assumptions.

Suppose (6.1) holds. Then it is necessary and sufficient for the ergodicity of process A that

$$E(X_1) = \sum_{-\infty}^{\infty} k c_k < 0 \quad (6.3)$$

This follows from the results of Section 3. We investigate now the ergodicity of process B when (6.1) and (6.3) hold.

If (6.1) and (6.3) hold and if we let  $E(X_1) = -\beta$  ( $\beta > 0$ ) then it follows from a renewal-type theorem of Chow and Robbins ((2), Theorem 2), that

$$F'_{i_0}(1) \sim \frac{1}{\beta} \quad (i \rightarrow \infty) \quad (6.4)$$

Differentiating (5.7) and setting  $s = 1$  we have

$$\sum_{k \in H} \Pi'_{jk}(1) = 1 + \sum_{i=\alpha+1}^{\infty} p_{ji} F'_{i-\alpha,0}(1) \quad (j \in H) \quad (6.5)$$

Thus provided that

$$\sum_{i=0}^{\infty} i p_{ji} < \infty \quad (j \in H) \quad , \quad (6.6)$$

we have from (6.4) and (6.5) that

$$\sum_{k \in H} \Pi'_{jk}(1) < \infty \quad (j \in H) \quad (6.7)$$

while if for some  $j \in H$

$$\sum_{i=0}^{\infty} i p_{ji} = \infty \quad , \quad (6.8)$$

then from some  $j \in H$

$$\sum_{k \in H} \Pi'_{jk}(1) = \infty \quad . \quad (6.9)$$

Hence if (6.6) holds for all  $j \in H$  then for all  $j, k \in H$

$$\sum_{H} \Pi_{jk}^j(1) < \infty, \quad (6.10)$$

while if (6.8) holds for some  $j \in H$ , then for some  $j, k \in H$ ,

$$\sum_{H} \Pi_{jk}^j(1) = \infty. \quad (6.11)$$

Thus if we apply these results to (5.2) and (5.3) after differentiating and setting  $s = 1$ , then we have that process B is ergodic if (6.6) holds for all  $j \in H$  and null-recurrent if (6.6) does not hold for all  $j \in H$ .

Next suppose that (6.1) holds and that

$$E(X_1) = \sum_{-\infty}^{\infty} k c_k = 0, \quad (6.12)$$

so that process A is null-recurrent. Then  $F'_{j_0}(1) = \infty$  for  $j = 0, 1, \dots$ .

It follows from (6.5) that (6.11) is true for some  $j, k \in H$  and again using (5.2) and (5.3) we see that process B is null-recurrent.

Again let (6.1) hold and suppose conversely that process B is ergodic. Then process A must be ergodic and (6.6) must hold since the ergodicity of process B precludes the transience or null-recurrence of process A. If (6.1) holds and process B is null-recurrent then clearly either process A is ergodic and (6.8) holds for some  $j \in H$  or process A is null-recurrent.

Summarizing, we have the following. If (6.1) holds then process B is ergodic if and only if process A is ergodic and (6.6) holds for all  $j \in H$ ; if process A is null recurrent process B is null-recurrent/or if process A is ergodic and (6.6) does not hold for all  $j \in H$ .

## 7. An Example

We consider the imbedded Markov chain of the queueing system GI/M/m. (general, independent, identically distributed inter-arrival times, exponential service times, m (> 1) servers, natural queue discipline). Let A(u) be the distribution function of inter-arrival times and let b be the mean service time. Kendall (5) showed that the number of persons ahead (waiting or being served) of a newly arrived customer at the instant of his arrival forms a Markov chain with one-step transition matrix of the form

$$P = \begin{bmatrix} Q_1 & Q_2 \\ Q_3 & Q_4 \end{bmatrix} .$$

Here  $Q_1$  is an  $m \times m$  matrix, and  $Q_4$  is a matrix of the form

$$Q_4 = \begin{bmatrix} b_1 & b_0 & 0 & 0 & 0 & \dots \\ b_2 & b_1 & b_0 & 0 & 0 & \dots \\ b_3 & b_2 & b_1 & b_0 & 0 & \dots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \dots \end{bmatrix}$$

Apart from the element in the lower left hand corner,  $Q_2$  is composed entirely of zeros. From the point of view of geometric ergodicity only the form of  $Q_4$  concerns us. The non-zero elements of  $Q_4$  are given by

$$b_k = \int_0^\infty \frac{e^{-(mu/b)} (mu/b)^k}{k!} dA(u) .$$

It follows from our results that the process is

(i) ergodic and geometrically ergodic if

$$-\infty \leq \sum_{k=0}^{\infty} (-k+1)b_k < 0 ,$$

i.e.  $1 < \sum_{k=0}^{\infty} kb_k \leq \infty ;$

(ii) transient and geometrically ergodic if

$$\sum_{k=0}^{\infty} k b_k < 1 \quad \text{and} \quad \sum_{k=0}^{\infty} b_k z^k \in \mathcal{P} .$$

For  $0 \leq z < 1$  we have

$$B(z) = \sum_{k=0}^{\infty} b_k z^k = \int_0^{\infty} e^{-\mu(z-1)u/b} dA(u) . \quad (7.1)$$

Hence

$$B'(1) = \frac{m}{b} \int_0^{\infty} u dA(u) = \frac{1}{\rho}$$

where  $\rho$  is the relative traffic intensity, and  $\rho = 0$  if  $\int u dA(u) = \infty$ .

Hence the process is ergodic and geometrically ergodic if  $0 \leq \rho < 1$ .

From the theory of moment generating functions the integral (7.1) defines a regular function of  $z$  in at least the half-plane  $\Re z < 1$  (Lukacs, 1960, Chapter 7). For  $B(z)$  to belong to  $\mathcal{P}$  the integral (7.1) must converge for some  $z > 1$ , i.e.  $A(u)$  must satisfy

$$1 - A(u) \leq C e^{-cu} \quad (7.2)$$

for some positive constants  $C, c$ . In other words, the process is transient and geometrically ergodic if  $\rho > 1$  and if (7.2) holds.

## 8. General Markov chains

In the special class of Markov chains we have considered so far we have seen that there is an inevitable connection between geometric ergodicity and some form of geometric bound on the one-step transition probabilities. As part of the general and as yet unsolved problem of determining conditions on the one-step transition probabilities of a general Markov chain which will be necessary and sufficient for geometric ergodicity it is natural to ask whether there is any similar connection in a general Markov chain. That such a

connection is not in general necessary is shown by the following result.

A sufficient condition for an irreducible aperiodic Markov chain to be ergodic and geometrically ergodic is that the elements in any one column of the one-step transition matrix, apart from the diagonal element, be bounded away from zero.

To prove this we label the state corresponding to the given column as state 0 and assume that the state space is the set of non-negative integers. With the usual notation for transition probabilities etc., we have the relation

$$(f_{00}^{(1)} + f_{00}^{(2)} + \dots + f_{00}^{(n)}) + (p_{01}^{(n)} + p_{02}^{(n)} + \dots) = 1$$

$$(n = 1, 2, \dots). \quad (8.1)$$

Now

$$\sum_{k=1}^{\infty} p_{ok}^{(n)} = \sum_{k=1}^{\infty} p_{ok}^{(n-1)} (1 - p_{ko}) \quad (n = 2, 3, \dots)$$

By assumption there exists a number  $\rho$  ( $0 < \rho < 1$ ) such that for  $k = 1, 2, \dots$ ,

$$p_{ko} \geq 1 - \rho \quad \text{i.e.} \quad 1 - p_{ko} \leq \rho.$$

Hence

$$\sum_{k=1}^{\infty} p_{ok}^{(n)} \leq \rho \sum_{k=1}^{\infty} p_{ok}^{(n-1)} \quad (n = 2, 3, \dots)$$

and from repeated applications of this inequality we have that

$$\sum_{k=1}^{\infty} p_{ok}^{(n)} \leq \rho^n \quad (n = 2, 3, \dots).$$

Hence from (8.1),

$$\sum_{r=1}^n f_{00}^{(r)} \geq 1 - \rho^n \quad (n = 2, 3, \dots)$$

and ergodicity and geometric ergodicity follow immediately.

#### REFERENCES

- (1) Baum, L. E., Katz, Melvin and Read, R. R. Exponential convergence rates for the law of large numbers. Trans. Amer. Math. Soc., 102 (1962), 187-199.
- (2) Chow, Y. S. and Robbins, Herbert. A renewal theorem for random variables which are dependent or non-identically distributed. Ann. Math. Statist., 34 (1963), 390-395.
- (3) Chung, Kai Lai. Markov chains with stationary transition probabilities. (Springer-Verlag; Berlin, 1960).
- (4) Debreu, Gerard and Herstein, I. N. Non-negative square matrices. Econometrica, 21 (1953), 597-607.
- (5) Kendall, David G. Some problems in the theory of queues. J. Roy. Statist. Soc. Ser. B, 13 (1951), 151-173.
- (6) Kendall, David G. Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded Markov chain. Ann. Math. Statist., 24 (1953), 338-354.
- (7) Kendall, David G. Unitary dilations of Markov transition operators and the corresponding integral representations for transition probability matrices. In U. Grenander, ed., Probability and Statistics (Wiley; New York, 1959).
- (8) Kendall, David G. Geometric ergodicity and the theory of queues. In Arrow, Kenneth J., Karlin, Samuel and Suppes, Patrick, ed., Mathematical methods in the social sciences, 1959 (Stanford University Press, 1960).
- (9) Lukacs, Eugene. Characteristic Functions (Griffin; London, 1960).
- (10) Miller, H. D. A generalization of Wald's identity with applications to random walks. Ann. Math. Statist., 24 (1961), 549-560.
- (11) Spitzer, Frank. A combinatorial lemma and its applications to probability theory. Trans. Amer. Math. Soc., 82 (1956), 323-339.
- (12) Spitzer, Frank. A Tauberian theorem and its probability interpretation. Trans. Amer. Math. Soc., 94 (1960), 150-169.
- (13) Vere-Jones, D. Geometric ergodicity in denumerable Markov chains. Quart. J. Math. Oxford 2nd Ser., 13 (1962), 7-28.

References (cont.)

- (14) Vere-Jones, D. A rate of convergence problem in the theory of queues. Teor. Veroyat. i ee Prim., 9 (1964), 104-112.
- (15) Widder, D. V. The Laplace Transform (Princeton University Press, 1941).