Journal of Multivariate Analysis **70**, 157–176 (1999)
Article ID jmva.1999.1814, available online at http://www.idealibrary.com on IDEAL®

Joint Distributions of the Numbers of Visits for Finite-State Markov Chains

Wolfgang Stadje

University of Osnabrück, Osnabrück, Germany

Received February 13, 1997

For a discrete-time Markov chain with finite state space $\{1, ..., r\}$ we consider the joint distribution of the numbers of visits in states $1, ..., r-1$ during the first N steps or before the Nth visit to r. From the explicit expressions for the corresponding generating functions we obtain the limiting multivariate distributions as $N \to \infty$ when state r becomes asymptotically absorbing and for $j = 1, ..., r-1$ the probability of a transition from r to j is of order $1/N$. © 1999 Academic Press

AMS 1991 subject classifications: 60J10, 60F05.

Key words and phrases: Markov chain; finite state-space; numbers of visits; joint distribution; generating function; asymptotic distribution.

1. INTRODUCTION

Consider a sequence of N independent replications of an experiment with the $r \geqslant 2$ different outcomes $1, ..., r$. Let $p_j(N)$ be the probability to get outcome j and define $V_j^{(N)}$ to be the number of trials resulting in j. Assume that $N \to \infty$, $p_r(N) \to 1$ and $Np_j(N) \to \sigma_j$ for $j = 1, ..., r-1$. It is well-known (and easy to prove) that $V_1^{(N)}, ..., V_{r-1}^{(N)}$ are asymptotically Poisson distributed and independent, i.e.,

$$P((V_1^{(N)}, ..., V_{r-1}^{(N)}) = (m_1, ..., m_{r-1})) \to \prod_{j=1}^{r-1} e^{-\sigma_j} \frac{\sigma_j^{m_j}}{m_j!}, \qquad (m_1, ..., m_{r-1}) \in \mathbb{Z}_+^{r-1}, \tag{1.1}$$

where $\mathbb{Z}_+ = \{0, 1, 2, ...\}$. The purpose of this paper is to generalize (1.1) to the case of Markov dependent trials. We study a sequence of Markov chains having the same finite state space $\{1, ..., r\}$, with r becoming an "almost absorbing" state, and consider the joint distribution of the numbers of visits in states $1, ..., r-1$ during the first N steps or before the Nth visit to r. Specifically, let the Nth Markov chain be given by the initial distribution $\pi = (\pi_j)_{1 \leqslant j \leqslant r}$ and the transition matrix $P(N) = (p_{i,j}(N))_{1 \leqslant i, j \leqslant r}$.

0047-259X/99 $30.00

For $i=1, ..., r-1$ we denote by $V_i^{(N)}$ and $U_i^{(N)}$ the number of visits in state i before time N and before the Nth visit in state r, respectively. We suppose that $p_{r,r}(N) \to 1$ and the other transition probabilities $p_{r,j}(N)$ for leaving r are of the form $p_{r,j}(N) = (\sigma_j + \varepsilon_j(N))/N$ for certain constants $\sigma_1, ..., \sigma_{r-1} \geqslant 0$, where $\varepsilon_j(N) \to 0$ for $j=1, ..., r-1$. The first $r-1$ rows of $P(N)$ converge componentwise or are simply constant. Summarizing, there are a matrix $P = (p_{i,j})_{1 \leqslant i,j \leqslant r}$ and constants $\sigma_1, ..., \sigma_{r-1} \geqslant 0$, at least one of them positive, such that

$$\lim_{N \to \infty} P(N) = P \qquad \text{componentwise} \tag{1.2}$$

$$\lim_{N \to \infty} N p_{r,j}(N) = \sigma_j, \qquad j = 1, ..., r-1. \tag{1.3}$$

(Note that (1.3) entails $p_{r,r} = \lim_{N \to \infty} p_{r,r}(N) = 1$, so that r is an absorbing state for the limiting Markov chain.) For the transition matrix P we assume that $\{1, ..., r-1\}$ does not contain a closed set.

Our main result is that the random vectors $(V_1^{(N)}, ..., V_{r-1}^{(N)})$ and $(U_1^{(N)}, ..., U_{r-1}^{(N)})$ converge weakly to the same probability distribution. In the course of the proof we will derive an explicit formula for the multivariate generating function l of this asymptotic distribution,

$$l(x) = \left(1 - \frac{\alpha_\pi(x)}{\beta(x)}\right) \exp\{-|\sigma|_1 \, \alpha_{\tilde{\sigma}}(x)/\beta(x)\}, \; x = (x_1, ..., x_{r-1}) \in [-1, 1]^{r-1}, \tag{1.4}$$

where $\sigma = (\sigma_1, ..., \sigma_{r-1})$, $|\sigma|_1 = \sum_{i=1}^{r-1} \sigma_i$ and $\tilde{\sigma} = (\sigma_1/|\sigma|_1, ..., \sigma_{r-1}/|\sigma|_1, 0)$. The functions $\alpha_\pi(x)$, $\alpha_{\tilde{\sigma}}(x)$ and $\beta(x)$ in (1.4) are certain determinants and are linear in each variable x_i separately; their precise definitions are given in Section 2 after Lemma 1. While $\alpha_\pi(x)$ and $\alpha_{\tilde{\sigma}}(x)$ depend on P and π or σ, respectively, $\beta(x)$ depends only on P. Starting from (1.4) and the formulas for $\beta(x)$, $\alpha_\pi(x)$ and $\alpha_{\tilde{\sigma}}(x)$ given below, $l(x)$ can, at least in principle, be expanded into a power series around the origin, leading to explicit expression for the limiting distribution.

For example, if P has $(1, ..., 1)^\top$ as its last column and $\pi_r = 1$, we find that

$$l(x) = \exp\left\{-\sum_{j=1}^{r-1} \sigma_j (1 - x_j)\right\}.$$

This yields a slight generalization of the limit Theorem (1.1) for the multinomial distribution: In the case of independent replications, assumed in

(1.1), the probabilities $p_{i,j}(N)$ are independent of i, while our result shows that (1.1) also holds for Markov-dependent trials satisfying $\lim_{N\to\infty} p_{i,j}(N)=\delta_{jr}$, $i=1, ..., r-1$, and $\lim_{N\to\infty} jp_{r,j}(N)=\sigma_j$, $j=1, ..., r-1$.

As a second example, consider the case of a two-state Markov chain, i.e., $r=2$. We obtain

$$l(x_1)=\left(1-\frac{\pi_1(1-x_1)}{1-p_{1,1}x_1}\right)\exp\left\{-\frac{\sigma_1(1-x_1)}{1-p_{1,1}x_1}\right\}.$$

This generating function can be expanded in closed form, yielding the asymptotic distribution in terms of Laguerre polynomials; see (3.12)–(3.13).

The subject of Poisson approximation has attracted a lot of attention in recent years due to the Chen-Stein method for deriving bounds for the total variation distance between the distribution of sums of dependent indicator variables and the appropriate Poisson distribution (see, e.g., Arratia, Goldstein, and Gordon, 1990, and the monograph by Barbour, Holst, and Janson, 1992). This technique is most useful for situations in which the summands satisfy some mixing conditions and does not seem to be applicable in our setting in which a new class of limiting multidimensional distributions arises. Our approach mainly uses classical matrix analysis (see, e.g., Seneta, 1981, and Berman and Plemmons, 1979) and is based on conditions which ensure the existence of certain inverse matrices. Similar methods have been applied to compute the distributions of sojourn times of absorption times (in discrete or continuous time); see, e.g., the paper on sojourn times by Rubino and Sericola (1989) and the work on reduced systems for Markov chains by Lal and Bhat (1987).

A related Poisson limit theorem for stationary and m-dependent indicator variables has been given by Kobus (1995). Wang (1981) and Gani (1982) considered the asymptotic behavior of the Markov binomial distribution which arises from a two-state stationary Markov chain and was proposed by Edwards (1960) as a generalization of the binomial distribution to allow for correlation between trials. In this special case they derived the limiting distributed for $V_1^{(N)}$ (but not for $U_1^{(N)}$). Their result was extended to nonhomogenous two-state chains by Pawlowski (1989).

The paper is organized as follows. In Section 2 we develop the required basic analysis of joint sojourn time distributions. In Section 3 the announced asymptotic results are proved and some examples are presented. Finally, Section 4 deals with the case that the limiting Markov chain has, besides the absorbing state r, a closed subset of $\{1, ..., r-1\}$. In this case the common weak limit of $(V_1^{(N)}, ..., V_{r-1}^{(N)})$ and $(U_1^{(N)}, ..., U_{r-1}^{(N)})$ is a subprobability measure.

The following notation is used. We write x for $(x_1, ..., x_{r-1}) \in \mathbb{R}^{r-1}$ and $\mathbf{m} = (m_1, ..., m_{r-1}) \in \mathbb{Z}_+^{r-1}$, $\mathbf{0} = (0, ..., 0) \in \mathbb{Z}_+^{r-1}$, and $x^{\mathbf{m}} = \prod_{i=1}^{r-1} x_i^{m_i}$. We use the norms $|x|_1 = \sum_{i=1}^{r-1} |x_i|$, $|x|_\infty = \max |x_i|$ and the cubes $U_\varepsilon = \{x \in \mathbb{R}^{r-1} \mid |x|_\infty \leqslant \varepsilon\}$, $\varepsilon > 0$, and $U = U_1$. If A is an $(r \times r)$-matrix and b a row vector with r components, $[A]_b$ is the matrix obtained by replacing the rth row of A by b. For $s = 1, ..., r$, E_s is the $(s \times s)$-identity matrix, $E = E_r$, and e_s is the unit row vector of length r given by $e_s = (\delta_{1s}, ..., \delta_{rs})$. For $u = (u_1, ..., u_s) \in \mathbb{R}^s$ we denote by $\Delta(u)$ the $(s \times s)$-diagonal matrix with diagonal entries $u_1, ..., u_s$; in particular, for $x \in \mathbb{R}^{r-1}$ and $y \in \mathbb{R}$ the $(r \times r)$-diagonal matrix $\Delta(x, y)$ has the diagonal elements $x_1, ..., x_{r-1}, y$.

2. JOINT SOJOURN TIME DISTRIBUTION FOR A MARKOV CHAIN

We start by considering a single Markov chain $X_1, X_2, ...$ with state space $\{1, ..., r\}$, transition matrix $P = (p_{i,j})_{1 \leqslant i, j \leqslant r}$ and initial distribution $\pi = (\pi_j)_{1 \leqslant j \leqslant r}$. Let $V_j^{(n)} = \#\{l \in \{1, ..., n\} \mid X_l = j\}$ and $V^{(n)} = (V_1^{(n)}, ..., V_r^{(n)})$. We are interested in the probabilities

$$\rho_j(\mathbf{m}, n) = P(V^{(|\mathbf{m}|_1 + n)} = (\mathbf{m}, m), X_{|\mathbf{m}|_1 + n} = j), \qquad j = 1, ..., r,$$

where $\mathbf{m} = (m_1, ..., m_{r-1}) \in \mathbb{Z}_+^{r-1}$, $n \in \mathbb{Z}_+$, $(\mathbf{m}, n) \neq (\mathbf{0}, 0)$. For related sojourn time distributions for a partitioned state-space see Csenki (1992) and the references given there.

We set $\rho_j(\mathbf{0}, 0) = 0$ and introduce the row vector $\rho(\mathbf{m}, n) = (\rho_1(\mathbf{m}, n), ..., \rho_r(\mathbf{m}, n))$. Let

$$\tilde{\rho}(x, y) = \sum_{\mathbf{m}, n} \rho(\mathbf{m}, n)\, x^{\mathbf{m}} y^n$$

be the corresponding row vector of multivariate generating functions of the sequences $\rho_j(\cdot, \cdot)$.

LEMMA 1.

$$\tilde{\rho}(x, y) = \tilde{\rho}(x, y)\, P\Delta(x, y) + \pi\Delta(x, y). \tag{2.1}$$

Proof. Let $\mathbf{l} = (\mathbf{m}, n)$, $|\mathbf{l}|_1 > 0$. Equation (2.1) follows immediately from the obvious recursion

$$\rho_j(\mathbf{l} + e_j) = \sum_{i=1}^{r} p_{i,j}\, \rho_i(\mathbf{l}), \qquad j = 1, ..., r$$

together with the boundary conditions

$$\rho_j(\mathbf{l}) = 0, \qquad \text{if} \quad l_j = 0$$
$$\rho_j(e_j) = \pi_j$$

for $j = 1, ..., r$. ∎

Equation (2.1) can be used to compute $\tilde{\rho}(x, y)$ provided that $\mathrm{Det}(E - P\Delta(x, y)) \neq 0$. In the sequel we need this determinant and various related ones:

$$\begin{aligned} \alpha(x, y) &= \mathrm{Det}(E - P\Delta(x, y)) \\ \alpha(x) &= \alpha(x, 1) \\ \alpha_\pi(x, y) &= \mathrm{Det}(E - [P]_\pi\, \Delta(x, y)) \\ \alpha_\pi(x) &= \alpha_\pi(x, 1) \\ \beta(x) &= \mathrm{Det}([E - P\Delta(x, y)]_{e_r}). \end{aligned}$$

For $i = 1, ..., r-1$ let $\delta_i(x, y)$ $(\bar{\delta}_i(x))$ be the determinant of the matrix obtained from $E - P\Delta(x, y)$ by replacing the ith row by $\pi\Delta(x, y)$ (and additionally the rth row by e_r). Note that $\beta(x)$ and $\bar{\delta}_i(x)$ are independent of y.

Lemma 2. *Let $A = (a_{ij})$ be a substochastic $(s \times s)$-matrix. Then*

$$g(u) = \mathrm{Det}(E_s - A\Delta(u)) > 0 \qquad \textit{for every} \quad u = (u_1, ..., u_s) \in (-1, 1)^s.$$

Proof. Suppose that $g(u) = 0$ for some $u \in (-1, 1)^s$. Then there is a non-vanishing row vector $z = (z_1, ..., z_s)$ (say $z_{j_0} \neq 0$) satisfying $zA\Delta(u) = z$, i.e.,

$$z_j = u_j \sum_{i=1}^{s} z_i a_{ij}, \qquad j = 1, ..., s.$$

Thus,

$$0 < |z_{j_0}| \leqslant |u_{j_0}| \left| \sum_{i=1}^{s} z_i a_{ij_0} \right| < \sum_{i=1}^{s} |z_i|\, a_{ij_0}$$

(note that $0 < |u_{j_0}| < 1$ and $|\sum_{i=1}^{s} z_i a_{ij_0}| > 0$) and

$$|z_j| \leqslant \sum_{i=1}^{s} |z_i|\, a_{ij} \qquad \text{for} \quad j \in \{1, ..., s\} \setminus \{j_0\}.$$

Adding these s inequalities yields $\sum_{j=1}^{s} |z_j| < \sum_{i=1}^{s} |z_i|$, which is impossible.

Hence, $g(u) \neq 0$ for all $u \in (-1, 1)^s$. As $g(\mathbf{0}) = 1$ and g is continuous, the result follows. ∎

Lemma 2 implies that $\alpha(x, y) > 0$ and $\beta(x) > 0$ for $x \in (-1, 1)^{r-1}$ and $y \in (-1, 1)$.

Remark. One could reduce Lemma 2 to the Neumann lemma for non-negative matrices (see, e.g., Berman and Plemmons (1979, p. 133)); the short proof above makes no use of infinite series or nonnegativity (note that the u_i can be negative).

Below we will express the generating functions $\tilde{p}_i$ in terms of $\delta_i(x, 1)$, $\bar{\delta}_i(x)$ and the functions

$$f(x) = 1 - \frac{\alpha(x)}{\beta(x)}$$

$$f_\pi(x) = 1 - \frac{\alpha_\pi(x)}{\beta(x)}.$$

Clearly, $f(x)$ and $f_\pi(x)$ are rational functions without poles in $(-1, 1)^{r-1}$. The following lemma provides a method to determine the coefficients of their Taylor expansions around $\mathbf{0}$.

LEMMA 3. *Let $\varepsilon \in [0, 1)$ and assume that $\beta(x) > 0$ for all $x \in [-1+\varepsilon, 1-\varepsilon]^{r-1} = U_{1-\varepsilon}$. Write $P = P_1 + P_2$, where P_1 has the first $r-1$ columns equal to those of P and the rth column equal to zero. Then*

$$(E - \Delta(x, 1)(P_1)^\top)^{-1} = \sum_{n=0}^{\infty} (\Delta(x, 1)(P_1)^\top)^n \tag{2.2}$$

for all $x \in U_{1-\varepsilon}$ and the series in (2.2) converges uniformly. Furthermore, $1 - (\alpha(x, y)/\beta(x))$ is equal to the entry in the rth row and rth column of the matrix

$$yP_2^\top \sum_{n=0}^{\infty} (\Delta(x, 1)\, P_1^\top)^n.$$

Proof. The matrix $E - \Delta(x, 1)\, P_1^\top$ is invertible for every $x \in U_{1-\varepsilon}$, because

$$\begin{aligned} \operatorname{Det}(E - \Delta(x, 1)\, P_1^\top) &= \operatorname{Det}(E - \Delta(x, y)\, P_1^\top) \\ &= \operatorname{Det}(E - P_1\, \Delta(x, y)) \\ &= \operatorname{Det}([E - P\, \Delta(x, y)]_{e_r}) \\ &= \beta(x) > 0. \end{aligned} \tag{2.3}$$

For a matrix A let $\|A\|_1$ be the sum of the absolute values of its entries. By (2.3) and the assumption on $\beta(x)$, the standard algorithm for computing the inverse of a matrix from its subdeterminants yields

$$\begin{aligned} c &= \sup\{\|(E-\varDelta(x,1)\,P_1^{\top})^{-1}\|_1 \mid x \in [-1+\varepsilon, 1-\varepsilon]^{r-1}\} \\ &\leqslant (r+1)!\,2^{r-1} \sup_{x\in U_{1-\varepsilon}} \beta(x)^{-1} < \infty. \end{aligned}$$

First consider the case $\varepsilon \in (0,1)$. Since $(\varDelta(x,1)\,P_1^{\top})^n$ is a polynomial in $x_1, ..., x_{r-1}$ with nonnegative coefficients and the substochactic matrix P_1 satisfies $\|(P_1^{\top})^n\|_1 \leqslant r$, it follows that

$$\begin{aligned} &\left\| E-(E-\varDelta(x,1)\,P_1^{\top}) \sum_{n=0}^{m} (\varDelta(x,1)\,P_1^{\top})^n \right\|_1 \\ &\quad = \|(\varDelta(x,1)\,P_1^{\top})^{m+1}\|_1 \\ &\quad \leqslant \|(\varDelta(1-\varepsilon, 1-\varepsilon, ..., 1-\varepsilon, 1)\,P_1^{\top})^{m+1}\|_1 \\ &\quad \leqslant (1-\varepsilon)^{m+1}\,\|(P_1^{\top})^{m+1}\|_1 \leqslant (1-\varepsilon)^{m+1}\,r \end{aligned}$$

for every $m \in \mathbb{N}$. Thus

$$\left\|(E-\varDelta(x,1)\,P_1^{\top})^{-1} - \sum_{n=0}^{m} (\varDelta(x,1)\,P_1^{\top})^n\right\|_1 \leqslant c(1-\varepsilon)^{m+1}\,r.$$

As $m \to \infty$, the finite sum on the left-hand side converges to the series $\sum_{n=0}^{\infty} (\varDelta(x,1)\,P_1^{\top})^n$, while the right-hand side tends to zero. This proves (2.2) for $\varepsilon > 0$.

Now let $\varepsilon = 0$. The above reasoning shows that (2.2) holds for all $x \in (-1,1)^{r-1}$. The left-hand side of (2.2) is a continuous function of $x \in [-1,1]^{r-1}$ and the right-hand series depends monotonically on $x_1, ..., x_{r-1} \in [0,1]$ (as all coefficients are nonnegative). Thus, taking the limit as $x_1, ..., x_{r-1} \nearrow 1$ shows that $\sum_{n=0}^{\infty} (P_1^{\top})^n$ converges. It follows that the series on the right-hand side of (2.2) converges absolutely and uniformly on $[-1,1]^{r-1}$ and that (2.2) holds for all $x \in [-1,1]^{r-1}$.

To prove the last assertion, not that

$$\begin{aligned} \alpha(x,y)/\beta(x) &= \operatorname{Det}(E-P\,\varDelta(x,y))/\operatorname{Det}(E-\varDelta(x,y)\,P_1^{\top}) \\ &= \operatorname{Det}((E-\varDelta(x,y)\,P^{\top})(E-\varDelta(x,y)\,P_1^{\top})^{-1}) \\ &= \operatorname{Det}((E-\varDelta(x,y)\,P_1^{\top}-\varDelta(x,y)\,P_2^{\top})(E-\varDelta(x,y)\,P_1^{\top})^{-1}) \\ &= \operatorname{Det}(E-\varDelta(x,y)\,P_2^{\top}(E-\varDelta(x,y)\,P_1^{\top})^{-1} \end{aligned}$$

$$=\operatorname{Det}\left(E-\Delta(x,y)\,P_2^\top\sum_{n=0}^{\infty}(\Delta(x,1)\,P_1^\top)^n\right)$$

$$=\operatorname{Det}\left(E-yP_2^\top\sum_{n=0}^{\infty}(\Delta(x,1)\,P_1^\top)^n\right). \tag{2.4}$$

For any $(r\times r)$-matrix $A=(a_{ij})$ the product $P_2^\top A$ is a matrix whose first $r-1$ rows consist of zeros, while the entry in the rth row and rth column is $\sum_{j=1}^r p_{j,r}a_{j,r}$. Applying this observation to the right-hand side of (2.4) concludes the proof of the lemma. ▮

Remarks. (1) It follows from Lemma 3 that

$$f(x)=\left(P_2^\top\sum_{n=0}^{\infty}(\Delta(x,1)\,P_1^\top)^n\right)_{r,r}, \tag{2.5}$$

where $(A)_{r,r}=a_{r,r}$. In particular, all coefficients of the Taylor expansion of $f(x)$ around $\mathbf{0}$ are nonnegative.

(2) Lemma 3 of course also holds for $f_\pi(x)$ instead of $f(x)$ with P replaced by $[P]_\pi$.

(3) As we have already shown that $\beta(\cdot)>0$ on $U_{1-\varepsilon}$ for any $\varepsilon>0$, the assumption of Lemma 3 is satisfied for every $\varepsilon>0$. Later we will also have to use the case $\varepsilon=0$ (i.e., the assumption $\beta(\cdot)>0$ on U).

LEMMA 4. *$f(x)\in[-1,1]$ for all $x\in(-1,1)^{r-1}$.*

Proof. Since $\alpha(x,y)>0$ for $x\in(-1,1)^{r-1}$ and $y\in(-1,1)$, it follows that $\alpha(x,1)\geqslant 0$ for $x\in(-1,1)^{r-1}$. Hence $f(x)=1-(\alpha(x,1)/\beta(x))\leqslant 1$. Having a Taylor expansion with nonnegative coefficients, f further satifies

$$|f(x)|\leqslant f(|x_1|,\ldots,|x_{r-1}|)\leqslant 1. \quad ▮$$

THEOREM 1. *The generating functions $\tilde{p}_1,\ldots,\tilde{p}_r$ are given by*

$$\tilde{p}_j(x,y)=\frac{y\delta_j(x,1)+(1-y)\,\bar{\delta}_j(x)}{\beta(x)(1-yf(x))},\qquad j=1,\ldots,r-1 \tag{2.6}$$

$$\tilde{p}_r(x,y)=\frac{yf_\pi(x)}{1-yf(x)} \tag{2.7}$$

for $x\in(-1,1)^{r-1}$, $y\in(-1,1)$.

Proof. It follows from Lemmas 1 and 2 that $\tilde{\rho}_j$ can be represented as

$$\tilde{\rho}_j(x, y) = \delta_j(x, y)/\alpha(x, y), \qquad j = 1, ..., r \tag{2.8}$$

by Cramer's rule. The functions δ_j and α satisfy

$$\alpha(x, y) = (1 - y)\,\beta(x) + y\alpha(x, 1), \tag{2.9}$$

$$\alpha_\pi(x, y) = (1 - y)\,\beta(x) + y\alpha_\pi(x, 1) \tag{2.10}$$

$$\delta_j(x, y) = (1 - y)\,\bar{\delta}_j(x) + y\delta_j(x, 1), \qquad j = 1, ..., r - 1. \tag{2.11}$$

To see (2.9), note that

$$y\alpha(x, 1) = \begin{vmatrix} 1 - p_{1,1}x_1, & -p_{1,2}x_2, & ..., & -p_{1,r-1}x_{r-1}, & -p_{1,r}y \\ \vdots & \vdots & & \vdots & \vdots \\ -p_{r-1,1}x_1, & -p_{r-1,2}x_2, & ..., & (1 - p_{r-1,r-1})\,x_{r-1}, & -p_{r-1,r}y \\ -p_{r,1}x_1, & -p_{r,2}x_2, & ..., & -p_{r,r-1}x_{r-1}, & (1 - p_{r,r})\,y \end{vmatrix}. \tag{2.12}$$

If we multiply the rth row of $\beta(x)$ by $1 - y$, we find that $(1 - y)\,\beta(x) = \operatorname{Det}(A)$, where A has the same first $r - 1$ rows as $y\alpha(x, 1)$ in (2.12) and the rth row $(0, ..., 0, 1 - y)$. Now we can use the linearity of the determinant in the rth row to derive (2.9). Equations (2.10) and (2.11) are proved in the same way.

Further we need the relation

$$\delta_r(x, y) = \beta(x) - \alpha_\pi(x, y). \tag{2.13}$$

Indeed, the first $r - 1$ rows of each of the three determinants in (2.13) are equal to those of $E - P\,\Delta(x, y)$, and the rth rows are $\pi\,\Delta(x, y)$, $(0, 0, ..., 0, 1)$ and $(-\pi_1 x_1, -\pi_2 x_2, ..., -\pi_{r-1}x_{r-1}, 1 - \pi_r y)$, respectively, so that (2.13) follows from the linearity of Det. Inserting (2.9)–(2.11) and (2.13) in (2.8) we arrive at (2.6)–(2.7). ∎

Now we define

$$h_j(x \mid n) = \sum_{\mathbf{m} \in \mathbb{Z}_+^{r-1}} \rho_j(\mathbf{m}, n)\, x^{\mathbf{m}}, \qquad j = 1, ..., r$$

$$h(x \mid n) = (h_1(x \mid n), ..., h_r(x \mid n))$$

for $x \in (-1, 1)^{r-1}$, $n \in \mathbb{Z}_+$. In particular, $h_r(x \mid n)$ is the joint generating function of the counting variables $U_1, ..., U_{r-1}$, where U_j denotes the number of visits in state j before the nth visit to state r. Indeed, the coefficients of $h_r(x \mid n)$ are given by $\rho_r(\mathbf{m}, n) = P(V^{(|\mathbf{m}|_1 + n)} = (\mathbf{m}, n),\ X_{|\mathbf{m}|_1 + n} = r)$.

THEOREM 2. *Define* $g(x) = (g_1(x), ..., g_r(x))$ *by*

$$g_j(x) = [\delta_j(x, 1) - (1 - f(x))\, \bar{\delta}_j(x)]/\beta(x), \qquad j = 1, ..., r-1$$

$$g_r(x) = f_\pi(x).$$

Then we have

$$\begin{aligned} h(x \mid 0) &= \beta(x)^{-1}\, (\bar{\delta}_1(x), ..., \bar{\delta}_{r-1}(x), 0) \\ h(x \mid n) &= g(x)\, f(x)^{n-1}, \qquad n \in \mathbb{N}. \end{aligned} \tag{2.14}$$

Proof. One has to use the relation

$$\tilde{\rho}(x, y) = \sum_{\mathbf{m}, n} \rho(\mathbf{m}, n)\, x^{\mathbf{m}} y^n = \sum_{n=0}^{\infty} h(x \mid n)\, y^n,$$

expand $\tilde{\rho}_j(x, y)$, as given by (2.6) and (2.7), in a power series in y and compare coefficients. ▮

In order to exclude the possibility that state r is only visited a finite number of times, we assume in the sequel that there is no closed set of states contained in $\{1, ..., r-1\}$. (If there were such a set and if this set were reachable from r, it would follow that r is almost surely visited only finitely often.)

We proceed with a series of auxiliary results. Lemma 5 is related to the well-known fact that a real square matrix is a nonsingular M-matrix if and only if all of its principal minors are positive (Berman and Plemmons, 1979, p. 134).

LEMMA 5. *Let* $A = (a_{ij})_{1 \leqslant i, j \leqslant s}$ *be a substochastic* $(s \times s)$*-matrix. Assume that*

(a) *A is not stochastic*;

(b) *every principal minor of order* $s-1$ *of* $E_s - A\Delta(u_1, ..., u_s)$ *is positive for all* $(u_1, ..., u_s) \in [-1, 1]^s$.

Then $D(u) = \mathrm{Det}(E_s - A\Delta(u_1, ..., u_s)) > 0$ *for all* $u = (u_1, ..., u_s) \in [-1, 1]^s$.

Proof. Assume on the contrary that $D(u)=0$ for some $u \in [-1, 1]^s$. Then there are $z_1, ..., z_s \in \mathbb{C}$, not all zero, satifying

$$z_j = u_j \sum_{i=1}^{s} z_i a_{ij}, \qquad j = 1, ..., s. \tag{2.15}$$

As $|u|_\infty \leqslant 1$, this implies

$$|z_j| \leqslant \sum_{i=1}^{s} |z_i| \, a_{ij}, \qquad j = 1, ..., s. \tag{2.16}$$

By (a), there is an $i_0 \in \{1, ..., s\}$ for which $\sum_{j=1}^{s} a_{i_0 j} = \delta < 1$. Summing the s inequalities (2.16) we thus obtain

$$\sum_{j=1}^{s} |z_j| \leqslant \sum_{i=1}^{s} |z_i| - (1-\delta) \, |z_{i_0}|, \tag{2.17}$$

and (2.17) yields $z_{i_0} = 0$. Therefore, (2.15) entails

$$z_j = u_j \sum_{\substack{i=1 \\ i \neq i_0}}^{s} z_i a_{ij}, \qquad j \in \{1, ..., s\} \setminus \{i_0\}.$$

The coefficient matrix of this homogeneous system of linear equations has as its determinant one of the principal minors of $E_s - A\Delta(u)$ of order $s-1$, which is not zero by (b). Thus $(z_1, ..., z_s) = (0, ..., 0)$, contradicting our initial assumption. Hence $D(u) \neq 0$ for $|u|_\infty \leqslant 1$, and $D(u) > 0$ on $[-1, 1]^s$ now follows from $D(\mathbf{0}) = 1$. ∎

LEMMA 6. *$\{1, ..., r-1\}$ contains no closed set if and only if $\beta(x) > 0$ for all $x \in U$.*

Proof. Let $\{1, ..., r-1\}$ contain no closed set. We show by induction on $s \in \{1, ..., r\}$ that every principal minor of the determinant $\beta(x)$ of order s is positive for all $x \in U$. For $s = r$ we get $\beta(x) > 0$ on U.

Since $\{i\}$ is not closed for $i \in \{1, ..., r-1\}$, we have $p_{i,i} < 1$, so that the principal minors of order 1 are all positive on U, because they have the form $1 - p_{i,i} x_i$. This proves the assertion for $s = 1$.

Let $i_0, j_0 \in \{1, ..., r-1\}$, $i_0 \neq j_0$. The matrix $Q = (p_{k,l})_{k,l \in \{i_0, j_0\}}$ is not stochastic. By the above reasoning, the two principal minors of order 1 of $\operatorname{Det}(E_2 - Q\Delta(x_{i_0}, x_{j_0}))$ are positive for $|x_{i_0}| \leqslant 1$, $|x_{j_0}| \leqslant 1$. Thus Lemma 5 yields $\operatorname{Det}(E_2 - Q\Delta(x_{i_0}, x_{j_0})) > 0$ for $|x_{i_0}|$, $|x_{j_0}| \leqslant 1$. This proves the assertion for $k = 2$, and the general induction step is carried out in the same way.

Now assume that there is a closed subset C of $\{1, ..., r-1\}$. If $C = \{1, ..., r-1\}$, the matrix $P' = (p_{i,j})_{1 \leqslant i, j \leqslant r-1}$ has the eigenvalue 1 so that

$$\beta(1, ..., 1) = \mathrm{Det}(E_{r-1} - P') = 0.$$

If C has $s < r-1$ elements, we may assume that $C = \{1, ..., s\}$. Then

$$P' = \begin{pmatrix} P_1 & 0 \\ P_2 & P_3 \end{pmatrix},$$

where P_1 is a stochastic $(s \times s)$-matrix. Hence, $\mathrm{Det}(E_s - P_1) = 0$ and

$$\beta(1, ..., 1) = \mathrm{Det}(E_{r-1} - P') = \mathrm{Det}(E_s - P_1)\, \mathrm{Det}(E_{r-1-s} - P_3) = 0. \quad \blacksquare$$

LEMMA 7. *If $\{1, ..., r-1\}$ does not contain a closed subset, the functions $f(x)$, $f_\pi(x)$, $h_r(x \mid n)$ are analytic on U (i.e., on an open set containing U). They are generating functions of probability distributions on $\mathbb{Z}_+^{r-1}$.*

Proof. f, f_π and $h_r(\cdot \mid n)$ are rational functions with the denominator $\beta(x)$ which, by Lemma 6, is positive on U. Thus they are analytic on U. The power series expansions of f and f_π around $\mathbf{0}$ have only nonnegative coefficients (by Lemma 3), and by Theorem 2 this holds for $h_r(\cdot \mid n)$, too.

Using Lemma 3 for P and for $[P]_\pi$ we see that the power series expansions of $\alpha(x, 1)/\beta(x)$, and $\alpha_\pi(x)/\beta(x)$ converge uniformly on U. Thus the same holds for $f(x)$, $f_\pi(x)$ and $h_r(x \mid n)$. Therefore, the sums of the coefficients of these series are given by $f(1, ..., 1)$, $f_\pi(1, ..., 1)$ and $h_r(1, ..., 1 \mid n)$, respectively. Since $\alpha(1, ..., 1) = \mathrm{Det}(E - P) = 0$ (and $\beta(1, ..., 1) > 0$), we have $f(1, ..., 1) = 1$, and $f_\pi(1, ..., 1) = 1$ follows in the same way. $\blacksquare$

3. THE LIMIT THEOREMS

In this section we consider a sequence $P(N) = (p_{i,j}(N))_{1 \leqslant i, j \leqslant r}$ of transition matrices satisfying

$$p_{i,j}(N) \to p_{i,j} \quad \text{for} \quad i = 1, ..., r-1 \quad \text{and} \quad j = 1, ..., r \tag{3.1}$$

$$Np_{r,j}(N) \to \sigma_j \quad \text{for} \quad j = 1, ..., r-1. \tag{3.2}$$

We assume that $\sigma_j > 0$ for some $j \in \{1, ..., r-1\}$. The initial distribution $\pi = (\pi_j)$ is fixed (although it would pose no additional difficulty to consider a convergent sequence $\pi(N)$). We use the notation $\alpha^{(N)}(x)$, $\beta^{(N)}(x)$, $f^{(N)}(x)$, $h^{(N)}(x \mid n)$ etc. for the functions introduced in Section 2 when they refer to the Markov chain corresponding to $P(N)$ and π. Note that $p_{r,j} = \lim_{N\to\infty} p_{r,j}(N) = \delta_{rj}$ for $j = 1, ..., r$ by (3.2). Thus, in this section the matrix $P = (p_{i,j})_{1\leqslant i,j\leqslant r}$ has the last row $(0, ..., 0, 1)$.

Furthermore, we assume that $\{1, ..., r-1\}$ does not contain a closed subset with respect to the stochastic matrix P. Then the same holds for $P(N)$ if N is sufficiently large; for if $\sum_{j\in C(N)} p_{ij}(N) = 1$ for certain sets $C(N) \subset \{1, ..., r-1\}$, for all $i \in C(N)$ and infinitely many N, then infinitely many of the $C(N)$ are equal, so that without loss in generality $C(N) = C$ for all N, and thus $\sum_{j\in C} p_{ij} = \lim_{N\to\infty} \sum_{j\in C} p_{ij}(N) = 1$ for all $i \in C$. The case when there is a closed subset of $\{1, ..., r-1\}$ is dealt with in Section 4.

In the subsequent limit theorems the function $U \ni x \to \mathrm{Det}([E - P\Delta(x, 1)]_{z(x)})$ occurs, where the row vector forming the last row of the determinant is given by

$$z(x) = (-\sigma_1 x_1, -\sigma_2 x_2, ..., -\sigma_{r-1} x_{r-1}, |\sigma|_1).$$

Let $\tilde{\sigma}$ be the row probability vector

$$\tilde{\sigma} = (\sigma_1/|\sigma|_1, ..., \sigma_{r-1}/|\sigma|_1, 0).$$

Then it is easily seen that

$$\mathrm{Det}([E - P\Delta(x, 1)]_{z(x)}) = |\sigma|_1 \,\mathrm{Det}(E - [P]_{\tilde{\sigma}}\, \Delta(x, 1)) = |\sigma|_1\, \alpha_{\tilde{\sigma}}(x),$$

where $\alpha_{\tilde{\sigma}}(x)$ is defined in the same way as $\alpha_\pi(x)$. Now let

$$l(x) = \left(1 - \frac{\alpha_\pi(x)}{\beta(x)}\right) \exp\{-|\sigma|_1\, \alpha_{\tilde{\sigma}}(x)/\beta(x)\}. \tag{3.3}$$

As in the proof of Lemma 7 it follows that $l(x)$ is analytic on U. We will show in the proof of Theorem 3 that $l(x)$ is the uniform limit of analytic functions $l^{(N)}(x)$ having a Taylor series expansion around $\mathbf{0}$ with non-negative coefficients. Actually this proof shows that $l(z)$, considered as a function of n complex variables $(z_1, ..., z_n) = z$ satisfying $|z_i| \leqslant 1$, is the uniform limit of a sequence of functions $l^{(N)}(z)$ which are holomorphic for $|z_i| \leqslant 1$, $i = 1, ..., r-1$. Thus if

$$l(z) = \sum_{m_1, ..., m_{r-1}=0}^{\infty} a_{\mathbf{m}}\, z_1^{m_1} \cdots z_{r-1}^{m_{r-1}}, \qquad \mathbf{m} = (m_1, ..., m_{r-1}).$$

Cauchy's integral formula for several complex variables yields that the coefficients $a_{\mathbf{m}}$ are given by

$$a_{\mathbf{m}} = (2\pi i)^{1-r} \int_{|z_1|=1/2} \cdots \int_{|z_{r-1}|=1/2} \frac{l(z)}{z_1^{m_1+1} \cdots z_{r-1}^{m_{r-1}+1}} \, dz_1 \cdots dz_{r-1}$$

$$= \lim_{N \to \infty} (2\pi i)^{1-r} \int_{|z_1|=1/2} \cdots \int_{|z_{r-1}|=1/2} \frac{l^{(N)}}{z_1^{m_1+1} \cdots z_{r-1}^{m_{r-1}+1}} \, dz_1 \cdots dz_{r-1}$$

(see, e.g., Grauert and Fritzsche, 1976, Chap. I.3). Hence $a_{\mathbf{m}}$ is the limit of the corresponding coefficients of the functions $l^{(N)}$ and thus also nonnegative.

From Lemma 3 (for the case $\varepsilon = 0$) we conclude that $\alpha_\pi(x)/\beta(x)$ and $\alpha_{\tilde{\sigma}}(x)/\beta(x)$ have Taylor series converging uniformly on U. Thus, the Taylor series of $l(x)$ converges uniformly on U and the sume of its (nonnegative) co-efficients is equal to $l(1, ..., 1)$. Let us show that $l(1, ..., 1) = 1$. We know already that $\alpha_\pi(1, ..., 1) = 0$ so that, by (3.3), it suffices to prove that $\alpha_{\tilde{\sigma}}(1, ..., 1) = 0$. Note that

$$\operatorname{Det}([E-P]_{z(1, ..., 1)}) = \begin{vmatrix} 1-p_{1,1} & -p_{1,2} & \cdots & -p_{1,r-1} & -p_{1,r} \\ \vdots & & & & \vdots \\ -p_{r-1,1} & -p_{r-1,2} & & 1-p_{r-1,r-1} & -p_{r-1,r} \\ -\sigma_1 & -\sigma_2 & & -\sigma_{r-1} & |\sigma|_1 \end{vmatrix}$$

$$= |\sigma|_1 \, M_{r,r} - \sum_{i=1}^{r-1} \sigma_i M_{r,i}, \tag{3.4}$$

where $M_{j,i}$ is the (j, i) cofactor of $E-P$. Observe that expanding $\operatorname{Det}(E-P)$ along its rth row yields

$$0 = \operatorname{Det}(E-P) = M_{r,r} - \sum_{i=1}^{r} p_{r,i} M_{r,i}. \tag{3.5}$$

Since $M_{r,1}, ..., M_{r,r}$ are independent of the last row of $E-P$, we can replace the vector $(p_{r,i})_{1 \leqslant i \leqslant r}$ in (3.5) by $(\delta_{j,i})_{1 \leqslant i \leqslant r}$ for any $j \in \{1, ..., r\}$. It follows that $M_{r,r} = M_{r,j}$, $j = 1, ..., r$. Using this in (3.4) shows that

$$\alpha_{\tilde{\sigma}}(1, ..., 1) = |\sigma|_1^{-1} \operatorname{Det}([E-P]_{z(1, ..., 1)}) = 0.$$

We conclude that $l(x)$ is generating function of some probability distribution $(q_{\mathbf{m}})_{\mathbf{m} \in \mathbb{Z}_+^{r-1}}$.

THEOREM 3. *If the sequence $n_1, n_2, \ldots$ of positive integers satisfies*

$$\lim_{N \to \infty} n_N/N = 1, \tag{3.6}$$

then

$$h_r^{(N)}(x \mid n_N) \to l(x) \tag{3.7}$$

$$h_j^{(N)}(x \mid n_N) \to 0, \qquad j = 1, \ldots, r-1 \tag{3.8}$$

uniformly for $|x|_\infty \leqslant 1$, as $N \to \infty$. In particular, $(U_1^{(N)}, \ldots, U_{r-1}^{(N)})$ converges in distribution to $(q_{\mathbf{m}})_{\mathbf{m} \in \mathbb{Z}_+^{r-1}}$.

Proof. By (2.14) we have

$$\begin{aligned} h_r^{(N)}(x \mid n_N) &= f_\pi^{(N)}(x)(f^{(N)}(x))^{n_N - 1} \\ &= f_\pi^{(N)}(x)[1 - n_N^{-1}(n_N \alpha^{(N)}(x)/\beta^{(N)}(x))]^{n_N - 1} \end{aligned} \tag{3.9}$$

and

$$\begin{aligned} h_j^{(N)}(x \mid n_N) = (f^{(N)}(x))^{n_N - 1} \, (\delta_j^{(N)}(x, 1) - (1 - f^{(N)}(x)) \, \bar{\delta}_j^{(N)}(x))/\beta^{(N)}(x), \\ j = 1, \ldots, r-1. \end{aligned} \tag{3.10}$$

Consider the determinants defining $\beta^{(N)}(x)$ and $n_N \alpha^{(N)}(x)$ (taking n_N as a factor in the rth row). Obviously, $n_N \alpha^{(N)}(x)$ and $\beta^{(N)}(x)$ are polynomials in $x_1, \ldots, x_{r-1}$ of degree $r-1$ whose coefficients are continuous functions of the entries of $P(N)$. Thus the first $r-1$ rows of $n_N \alpha^{(N)}(x)$ and all r rows of $\beta^{(N)}(x)$ converge to the corresponding rows of $\alpha(x)$ and $\beta(x)$. The rth row of $n_N \alpha^{(N)}(x)$ is given by

$$\left(-n_N p_{r,1}(N) \, x_1, \, -n_N p_{r,2}(N) \, x_2, \ldots, -n_N p_{r,r-1}(N) \, x_{r-1}, \, n_N \sum_{i=1}^{r-1} p_{r,i}(N) \right) \tag{3.11}$$

and consequently, by (3.2) and (3.6), converges to

$$\left(-\sigma_1 x_1, \, -\sigma_2 x_2, \ldots, -\sigma_{r-1} x_{r-1}, \, \sum_{i=1}^{r-1} \sigma_i \right).$$

It follows that

$$[1 - n_N^{-1}(n_N \alpha^{(N)}(x)/\beta^{(N)}(x))]^{n_N - 1} \to \exp\{-|\sigma|_1 \, \alpha_{\tilde{\sigma}}(x)/\beta(x)\}$$

uniformly for $|x|_\infty \leqslant 1$.

As above it is seen that $\alpha_\pi^{(N)}(x)$, $\delta_j^{(N)}(x, 1)$ and $\bar{\delta}_j^{(N)}(x)$ are polynomials of degree at most $r-1$ whose coefficients are continuous functions of $P(N)$. Thus, $f_\pi^{(N)}(x) = 1-(\alpha_\pi^{(N)}(x)/\beta^{(N)}(x))$ converges to $f_\pi(x) = 1-(\alpha_\pi(x)/\beta(x))$ uniformly for $|x|_\infty \leqslant 1$. Relation (3.7) is proved. Furthermore, $\bar{\delta}_j^{(N)}(x)$ and $\beta^{(N)}(x)$ converge to $\bar{\delta}_j(x)$ and $\beta(x)$, again uniformly for $|x|_\infty \leqslant 1$, and $f^{(N)}(x) \to 1$. The rth row of $\delta_j^{(N)}(x, 1)$ is given by (3.11) divided by n_N so that it converges uniformly to $(0, ..., 0)$. Hence $\delta_j^{(N)}(x, 1) \to 0$ uniformly for $|x|_\infty \leqslant 1$. The numerator in (3.10) converges to 0, yielding (3.8). Finally, $(U_1^{(N)}, ..., U_{r-1}^{(N)})$ has the generating function $h_r^{(N)}(x \mid N)$, so that weak convergence to the distribution corresponding to l follows from the continuity theorem for generating functions. ∎

Recall that $\rho_r^{(N)}(\mathbf{m}, n)$ is the probability that for the Nth Markov chain the states $1, ..., r-1$ are visited $m_1, ..., m_{r-1}$ times, respectively, before the nth visit of state r. Thus

$$\tau^{(N)}(\mathbf{m}) = \rho_r^{(N)}(\mathbf{m}, N - |\mathbf{m}|_1)$$

is the probability that in the first N steps of the Nth Markov chain the states $1, ..., r-1$ are visited $m_1, ..., m_{r-1}$ times. Our final theorem shows that the sequence $\tau^{(N)}(\cdot)$, $N = 1, 2, ...,$ of probability distributions converges to $(q_\mathbf{m})_{\mathbf{m} \in \mathbb{Z}_+^{r-1}}$.

THEOREM 4. $(\tau^{(N)}(\mathbf{m}))_{\mathbf{m} \in \mathbb{Z}_+^{r-1}} \xrightarrow{D} (q_\mathbf{m})_{\mathbf{m} \in \mathbb{Z}_+^{r-1}}$, *as* $N \to \infty$.

Proof. It follows from Theorem 3 that

$$\rho_r^{(N)}(\mathbf{m}, n_N) \to q_\mathbf{m} \qquad \text{for all} \quad \mathbf{m} \in \mathbb{Z}_+^{r-1},$$

if the sequence n_N satisfies $\lim_{N \to \infty} n_N/N = 1$. Now fix $\mathbf{m} \in \mathbb{Z}_+^{r-1}$ and define n_N by

$$n_N = \begin{cases} N - |\mathbf{m}|_1, & \text{if} \quad N > |\mathbf{m}|_1 \\ 1, & \text{if} \quad N \leqslant |\mathbf{m}|_1. \end{cases}$$

Then $\lim_{N \to \infty} n_N/N = 1$ and as $N \to \infty$,

$$\begin{aligned} \tau^{(N)}(\mathbf{m}) &= \rho_r^{(N)}(\mathbf{m}, N - |\mathbf{m}|_1) \\ &= \rho_r^{(N)}(\mathbf{m}, n_N) \to q_\mathbf{m}. \end{aligned}$$

But the relation $\lim_{N \to \infty} \tau^{(N)}(\mathbf{m}) = q_\mathbf{m}$ for every $\mathbf{m} \in \mathbb{Z}_+^{r-1}$ entails convergence in distribution. ∎

Finally let us reconsider the two examples mentioned in the Introduction.

(a) Let $P=(e_r^\top, ..., e_r^\top)^\top$, i.e., the entries are all 0 in the first $r-1$ columns and all 1 in the rth column, and $\pi = e_r$. Then $\alpha_\pi(x) \equiv 0$, $\beta(x) \equiv 1$ and

$$|\sigma|_1 \, \alpha_{\tilde{\sigma}}(x) = \begin{vmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ \vdots & \vdots & \vdots & \vdots \\ -\sigma_1 x_1 & -\sigma_2 x_2 & -\sigma_{r-1} x_{r-1} & |\sigma|_1 \end{vmatrix}$$
$$= |\sigma|_1 - \sigma_1 x_1 - \cdots - \sigma_{r-1} x_{r-1}.$$

Hence,

$$l(x) = \exp\left\{ -\sum_{j=1}^{r-1} \sigma_j (1 - x_j) \right\}$$

so that the limiting distribution is multivariate Poisson with independent components:

$$q_{\mathbf{m}} = e^{-|\sigma|_1} \sigma^{\mathbf{m}} / (m_1! \cdots m_{r-1}!).$$

(b) Let $r=2$ so that $\pi = (\pi_1, 1-\pi_1)$ and

$$P = \begin{pmatrix} p_{1,1} & 1-p_{1,1} \\ 0 & 1 \end{pmatrix}.$$

In this case

$$\alpha_\pi(x_1) = \begin{vmatrix} 1 - p_{1,1} x_1 & -(1-p_{1,1}) \\ -\pi_1 x_1 & \pi_1 \end{vmatrix} = \pi_1 (1 - x_1)$$
$$|\sigma|_1 \, \alpha_{\tilde{\sigma}}(x_1) = \sigma_1 (1 - x_1)$$
$$\beta(x) = \begin{vmatrix} 1 - p_{1,1} x_1 & -(1-p_{1,1}) \\ 0 & 1 \end{vmatrix} = 1 - p_{1,1} x_1.$$

Thus,

$$l(x_1) = \left(1 - \frac{\pi_1 (1 - x_1)}{1 - p_{1,1} x_1} \right) \exp\left\{ -\frac{\sigma_1 (1 - x_1)}{1 - p_{1,1} x_1} \right\}.$$

To find a closed formula for the corresponding probability distribution $(q_m)_{m \in \mathbb{Z}_+}$, one can use the Laguerre polynomials $L_m(u) = \sum_{k=0}^m (-1)^k \binom{m}{k} u^k / k!$, which appear as coefficients in the well-known expansion

$$(1-t)^{-1} \exp\{-ut/(1-t)\} = \sum_{n=0}^{\infty} L_n(u)\, t^n, \qquad |t| < 1.$$

Write $l(x_1)=\sum_{m=0}^{\infty} q_m x_1^m$ as

$$
\begin{aligned}
l(x_1) &= \frac{1-\pi_1}{1-p_{1,1}x_1} \exp\left\{\frac{(1-p_{1,1})(\sigma_1/p_{1,1})\, p_{1,1}x_1}{1-p_{1,1}x_1} - \sigma_1\right\} \\
&\quad + \frac{(\pi_1-p_{1,1})\, x_1}{1-p_{1,1}x_1} \exp\left\{\frac{(1-p_{1,1})(\sigma_1/p_{1,1})\, p_{1,1}x_1}{1-p_{1,1}x_1} - \sigma_1\right\} \\
&= e^{-\sigma_1}(1-\pi_1) \sum_{n=0}^{\infty} L_n\left(-\frac{(1-p_{1,1})\,\sigma_1}{p_{1,1}}\right)(p_{1,1}x_1)^n \\
&\quad + e^{-\sigma_1}(\pi_1-p_{1,1})\, x_1 \sum_{n=0}^{\infty} L_n\left(-\frac{(1-p_{1,1})\,\sigma_1}{p_{1,1}}\right)(p_{1,1}x_1)^n
\end{aligned}
$$

and compare the coefficients of x_1^m to obtain

$$
\begin{aligned}
q_m &= e^{-\sigma_1}\Bigg[(1-\pi_1)\, p_{1,1}^m L_m\left(-\frac{(1-p_{1,1})\,\sigma_1}{p_{1,1}}\right) \\
&\quad + (\pi_1-p_{1,1})\, p_{1,1}^{m-1} L_{m-1}\left(-\frac{(1-p_{1,1})\,\sigma_1}{p_{1,1}}\right)\Bigg], \qquad m \geqslant 1 \qquad (3.12)
\end{aligned}
$$

$$
q_0 = e^{-\sigma_1}(1-\pi_1). \qquad (3.13)
$$

4. CLOSED SUBSETS OF THE LIMITING CHAIN

Let us finally turn to the case that $\beta(1, ..., 1)=0$, i.e., $\{1, ..., r-1\}$ contains a closed subset for the limiting Markov chain. Following the proofs of Theorems 3 and 4, it is seen that relations (3.7) and (3.8) hold for $x \in (-1, 1)^{r-1}$, and $(U_1^{(N)}, ..., U_{r-1}^{(N)})$ and $(V_1^{(N)}, ..., V_{r-1}^{(N)})$ converge weakly to some subprobability measure having the generating function

$$
l(x)=\left(1-\frac{\alpha_\pi(x)}{\beta(x)}\right)\exp\{-|\sigma|_1\, \alpha_{\tilde{\sigma}}(x)/\beta(x)\}, \qquad |x|_\infty < 1. \qquad (4.1)
$$

Actually, $l(x)$ can be analytically extended to a region $|x|_\infty < 1+\varepsilon$ for some $\varepsilon > 0$, and if $\{1, ..., s\} \subset \{1, ..., r-1\}$ is a closed set of states, then $l(x)$ is independent of the variables $x_1, ..., x_s$.

To prove these statements, let $\tilde{P}$ be one of the matrices $[P]_\pi$ or $[P]_{\tilde{\sigma}}$. If there is a closed subset of $\{1, ..., r-1\}$, we can write $\tilde{P}$, after a suitable permutation of $\{1, ..., r-1\}$, as

$$
\tilde{P}=\begin{pmatrix} Q_1 & 0 & 0 \\ Q_2 & Q_3 & \mathbf{a} \\ \mathbf{b}_1^\top & \mathbf{b}_2^\top & q_{rr} \end{pmatrix},
$$

where Q_1 is a stochastic $(s \times s)$-matrix, Q_3 is a $((r-1-s) \times (r-1-s))$-matrix for which the corresponding group of states $\{s+1, ..., r-1\}$ does not contain a closed subset, Q_2 is a $((r-1-s) \times s)$-matrix, $\mathbf{a}, \mathbf{b}_2 \in \mathbb{R}^{r-1-s}$, $\mathbf{b}_1 \in \mathbb{R}^s$ and $q_{rr} = \pi_r$ or $q_{rr} = 0$. (In the case $s = r-1$ the matrices Q_2, Q_3 and the vectors $\mathbf{a}$, $\mathbf{b}_2$ disappear.) It follows that

$$E_r - \tilde{P}\Delta(x, 1) = \begin{pmatrix} E_s - Q_1\,\Delta', & 0 & 0 \\ -Q_2\,\Delta', & E_{r-1-s} - Q_3\,\Delta'', & -\mathbf{a} \\ -\mathbf{b}_1^\top\,\Delta', & -\mathbf{b}_2^\top, & 1 - q_{rr} \end{pmatrix}, \tag{4.2}$$

where we have set $\Delta' = \Delta(x_1, ..., x_s)$ and $\Delta'' = \Delta(x_{s+1}, ..., x_{r-1})$. Let

$$C(x_{s+1}, ..., x_{r-1}) = \mathrm{Det} \begin{pmatrix} E_{r-1-s} - Q_3\,\Delta'', & -\mathbf{a} \\ -\mathbf{b}_2^\top\,\Delta'', & 1 - q_{rr} \end{pmatrix}.$$

From (4.2) we obtain that $\alpha_\pi(x)$ and $\alpha_{\tilde{\sigma}}(x)$ are of the form $\mathrm{Det}(E_s - Q_1\,\Delta')$ $\mathrm{Det}\; C(x_{s+1}, ..., x_{r-1})$ and

$$\beta(x) = \mathrm{Det}(E_s - Q_1\,\Delta')\,\mathrm{Det}(E_{r-1-s} - Q_3\,\Delta'').$$

By Lemma 2, it is seen that $\mathrm{Det}(E_s - Q_1\,\Delta') > 0$ for $|x_1|, ..., |x_s| < 1$, while using Lemma 6, applied to Q_3, we find that $\mathrm{Det}(E_{r-1-s} - Q_3\,\Delta'') > 0$ for $|x_{s+1}|, ..., |x_{r-1}| \leqslant 1$.

Hence, in the ratio $\alpha_\pi(x)/\beta(x)$ the common factor $\mathrm{Det}(E_s - Q_1\,\Delta')$ cancels out for $|x|_\infty < 1$ and we obtain

$$f_\pi(x) = 1 - \frac{\alpha_\pi(x)}{\beta(x)} = 1 - \frac{\mathrm{Det}\; C(x_{s+1}, ..., x_{r-1})}{\mathrm{Det}(E_{r-1-s} - Q_3\,\Delta'')}, \qquad |x|_\infty < 1.$$

It follows that $f_\pi(x)$ is a rational function of the variables $x_{s+1}, ..., x_{r-1}$ (i.e., independent of $x_1, ..., x_s$) with no poles in $|x|_\infty \leqslant 1$. Therefore, it has no poles in $|x|_\infty < 1 + \varepsilon$ for some $\varepsilon > 0$. As in Lemma 3, it is shown that $f_\pi(x)$ has a Taylor series expansion around zero with nonnegative coefficients. The same holds for $-\alpha_{\tilde{\sigma}}(x)/\beta(x)$ and thus for $l(x)$ in (4.1).

As a simple example in which $l(x)$ does not represent a probability measure, consider the case that $\{1, ..., r-1\}$ is closed. Then $f_\pi(x) \equiv \pi_r$ and the total mass of the limiting measure is $l(1, ..., 1) = \pi_r$.

REFERENCES

[1] R. Arratia, L. Goldstein, and L. Gordon, Poisson approximation and the Chen–Stein method, *Statist. Sci.* **5** (1990), 403–434.

[2] A. D. Barbour, L. Holst, and S. Janson, "Poisson Approximation," Oxford Univ. Press, Oxford, 1992.

[3] A. Berman and R. J. Plemmons, "Nonnegative Matrices in the Mathematical Sciences," Academic Press, New York, 1979.

[4] A. Csenki, The joint distribution of sojourn times in finite Markov processes, *Adv. Appl. Prob.* **24** (1992), 141–160.

[5] A. W. F. Edwards, The meaning of binomial distribution, *Nature* **186** (1960), 1074.

[6] J. Gani, On the probability generating function of the sum of Markov Bernoulli random variables (Essays in Statistical Science, Papers in Honor of P. A. P. Moran), *J. Appl. Probab.* **19A** (1982), 321–326.

[7] H. Grauert and K. Fritzsche, "Several Complex Variables," Springer-Verlag, New York, 1976.

[8] M. Kobus, Generalized Poisson distributions as limits of sums for arrays of dependent random vectors, *J. Multivariate Anal.* **52** (1995), 199–244.

[9] R. Lal and U. N. Bhat, Reduced systems in Markov chains and their applications in queueing theory, *Queueing Systems* **2** (1987), 147–172.

[10] J. Pawlowski, Poisson theory for non-homogeneous Markov chains, *J. Appl. Probab.* **26** (1989), 637–642.

[11] G. Rubino and B. Sericola, Sojourn times in finite Markov processes, *J. Appl. Probab.* **27** (1989), 744–756.

[12] E. Seneta, "Non-negative Matrices and Markov Chains," 2nd ed., Springer-Verlag, New York, 1980.

[13] Y. H. Wang, On the limit of the Markov binomial distribution, *J. Appl. Probab.* **18** (1981), 937–942.