JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 9, 153-161 (1964)

A Combinatorial Method in the Theory of Markov Chains

LAJOS TAKÁCS

Department of Mathematical Statistics, Columbia University, New York, New York

Submitted by Richard Bellman

I. INTRODUCTION

In this paper a simple combinatorial method is given for the determination of the n-step transition probabilities of homogeneous Markov chains with transition probability matrices

$$\pi = \left\| \begin{matrix} 1 - q_0 & q_0 & 0 & 0 & \cdots \\ 1 - q_0 - q_1 & q_1 & q_0 & 0 & \cdots \\ 1 - q_0 - q_1 - q_2 & q_2 & q_1 & q_0 & \cdots \\ 1 - q_0 - q_1 - q_2 - q_3 & q_3 & q_2 & q_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{matrix} \right\| \tag{1}$$

and

$$\pi = \left\| \begin{matrix} q_0 & q_1 & q_2 & q_3 & \cdots \\ q_0 & q_1 & q_2 & q_3 & \cdots \\ 0 & q_0 & q_1 & q_2 & \cdots \\ 0 & 0 & q_0 & q_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \cdots \end{matrix} \right\|. \tag{2}$$

Particular cases of these Markov chains play an important role in the theory of queues, storage, dams and elsewhere. In finding the higher transition probabilities we are making use of an elementary combinatorial theorem which is a generalization of the classical ballot theorem. It is very surprising that the classical ballot theorem which was discovered in 1887 by J. Bertrand [1], D. André [2] and É. Barbier [3] has also some importance in the theory of Markov chains. The difficulties usually involved in finding the powers of complicated transition probability matrices are well known. Earlier, when I considered particular cases of both (1) and (2), I was able to express only the double generating function of the higher transition probabilities with the aid of a root of a transcendental equation. (Cf. [4, pp. 70, 105].) Now the more general problem will be solved explicitly in an elementary way.

II. An Auxiliary Theorem

The following theorem is a generalization of the classical ballot theorem.

THEOREM 1. *If* $\nu_1, \nu_2, \cdots, \nu_n$ *are interchangeable random variables that assume nonnegative integer values, then*

$$\mathbf{P}\{\nu_1 + \cdots + \nu_r < r \text{ for } r = 1, \cdots, n \mid \nu_1 + \cdots + \nu_n = k\}$$
$$= \begin{cases} \left(1 - \dfrac{k}{n}\right) & \textit{if} \quad 0 \leqslant k \leqslant n, \\ 0 & \textit{otherwise}, \end{cases}$$

whenever the left hand side is defined.

For the proof of (3) we refer to [5] and [6]. In [6] the connection between Theorem 1 and the classical ballot theorem is also discussed.

III. The Higher Transition Probabilities

Let $\{\nu_n\}$ be a sequence of mutually independent random variables with distribution

$$\mathbf{P}\{\nu_n = j\} = q_j \qquad (j = 0, 1, \cdots). \tag{4}$$

Define two sequences of random variables $\{\xi_n\}$ and $\{\zeta_n\}$ as follows:

$$\xi_n = [\xi_{n-1} + 1 - \nu_n]^+ \qquad (n = 1, 2, \cdots) \tag{5}$$

where ξ_0 is a random variable that assumes nonnegative integer values and independent of $\{\nu_n\}$;

$$\zeta_n = [\zeta_{n-1} - 1]^+ + \nu_n \qquad (n = 1, 2, \cdots) \tag{6}$$

where ζ_0 is a random variable that assumes nonnegative integer values and independent of $\{\nu_n\}$. $[a]^+ = a$ if $a \geqslant 0$ and $[a]^+ = 0$ if $a \leqslant 0$. It can easily be seen that $\{\xi_n\}$ and $\{\zeta_n\}$ are homogeneous Markov chains with transition probability matrices (1) and (2) respectively. The n-step transition probabilities of $\{\xi_n\}$ and $\{\zeta_n\}$ are given by Theorem 2 and Theorem 3.

THEOREM 2. *For the Markov chain* $\{\xi_n\}$ *we have*

$$\mathbf{P}\{\xi_n \geqslant k \mid \xi_0 = i\} = \mathbf{P}\{\nu_1 + \cdots + \nu_n \leqslant n + i - k\}$$

$$+ \sum_{j=k}^{n-1} \frac{k}{j} \mathbf{P}\{\nu_1 + \cdots + \nu_j = j - k\} \mathbf{P}\{\nu_{j+1} + \cdots + \nu_n > n + i - j\} \tag{7}$$

if $k = 1, 2, \cdots; \; i = 0, 1, \cdots$. *In particular,*

$$\mathbf{P}\{\xi_n \geqslant k \mid \xi_0 = 0\} = \sum_{j=k}^{n} \frac{k}{j} \mathbf{P}\{\nu_1 + \cdots + \nu_j = j - k\} \tag{8}$$

if $k = 1, 2, \cdots$.

PROOF. From (5) it follows by induction, that

$$\xi_n = \max (0, 1 - \nu_n, 2 - \nu_{n-1} - \nu_n, \cdots, n - \nu_1 - \cdots - \nu_n + \xi_0). \tag{9}$$

If in (9) we replace $\nu_1, \nu_2, \cdots, \nu_n$ by $\nu_n, \nu_{n-1}, \cdots, \nu_1$ respectively, then we obtain a new random variable

$$\tilde{\xi}_n = \max (0, 1 - \nu_1, 2 - \nu_1 - \nu_2, \cdots, n - \nu_1 - \cdots - \nu_n + \xi_0) \tag{10}$$

which has exactly the same distribution as (9). Thus

$$\mathbf{P}\{\xi_n \geqslant k \mid \xi_0 = i\} = \mathbf{P}\{\tilde{\xi}_n \geqslant k \mid \xi_0 = i\}.$$

Let $\xi_0 = i \geqslant 0$. The event $\{\tilde{\xi}_n \geqslant k\}$ can occur in the following mutually exclusive ways: either

$$\nu_1 + \cdots + \nu_n \leqslant n + i - k,$$

or

$$\nu_1 + \cdots + \nu_n > n + i - k \quad \text{and} \quad \nu_1 + \cdots + \nu_r \leqslant r - k$$

for some $r = 1, \cdots, n - 1$. If $r = j$ is the smallest index for which

$$\nu_1 + \cdots + \nu_r \leqslant r - k$$

then necessarily

$$\nu_1 + \cdots + \nu_j = j - k \quad \text{and} \quad \nu_1 + \cdots + \nu_r > r - k$$

for $r = 1, \cdots, j - 1$. Accordingly,

$$\mathbf{P}\{\tilde{\xi}_n \geqslant k \mid \xi_0 = i\} = \mathbf{P}\{\nu_1 + \cdots + \nu_n \leqslant n + i - k\}$$

$$+ \sum_{j=k}^{n-1} \mathbf{P}\{\nu_1 + \cdots + \nu_n > n + i - k, \nu_1 + \cdots + \nu_j = j - k \text{ and}$$
$$\nu_1 + \cdots + \nu_r > r - k \text{ for } r = 1, \cdots, j - 1\}, \tag{11}$$

whence

$$\mathbf{P}\{\xi_n \geqslant k \mid \xi_0 = i\} = \mathbf{P}\{\nu_1 + \cdots + \nu_n \leqslant n + i - k\}$$

$$+ \sum_{j=k}^{n-1} \mathbf{P}\{\nu_{j+1} + \cdots + \nu_n > n + i - j\}. \qquad (12)$$

$\mathbf{P}\{\nu_{r+1} + \cdots + \nu_j < j - r \text{ for } r = 1, \cdots, j-1 \text{ and } \nu_1 + \cdots + \nu_j = j - k\}$ because $\nu_1, \nu_2, \cdots, \nu_n$ are mutually independent random variables. By Theorem 1

$$\mathbf{P}\{\nu_{r+1} + \cdots + \nu_j < j - r \text{ for } r = 1, \cdots, j-1 \mid \nu_1 + \cdots + \nu_j = j - k\}$$

$$= \begin{cases} k/j & \text{for} \quad 0 \leqslant k \leqslant j, \\ 0 & \text{otherwise}, \end{cases} \qquad (13)$$

and thus (12) and (13) yield (7).

Now let $\xi_0 = 0$. By (10) the event $\{\xi_n \geqslant k\}$ occurs if and only if there is an index r such that $\nu_1 + \cdots + \nu_r = r - k$. If $r = j$ $(j = k, \cdots, n)$ is the smallest index with this property, then $\nu_1 + \cdots + \nu_r > r - k$ $(r = 1, \cdots, j - 1)$ must hold. Accordingly,

$$\mathbf{P}\{\xi_n \geqslant k \mid \xi_0 = 0\}$$

$$= \sum_{j=k}^{n} \mathbf{P}\{\nu_1 + \cdots + \nu_j = j - k \text{ and } \nu_1 + \cdots + \nu_r > r - k \text{ for } r = 1, \cdots, j-1\}$$

$$= \sum_{j=k}^{n} \mathbf{P}\{\nu_1 + \cdots + \nu_j = j - k \text{ and } \nu_{r+1} + \cdots + \nu_j < j - r \text{ for } r = 1, \cdots, j-1\}. \qquad (14)$$

By Theorem 1

$$\mathbf{P}\{\nu_{r+1} + \cdots + \nu_j < j - r \text{ for } r = 1, \cdots, j-1 \mid \nu_1 + \cdots + \nu_j = j - k\}$$

$$= \begin{cases} k/j & \text{if} \quad 0 \leqslant k \leqslant j, \\ 0 & \text{otherwise}. \end{cases} \qquad (15)$$

Equation (14) and (15) prove (8).

THEOREM 3. *For the Markov chain $\{\zeta_n\}$ we have*

$$\mathbf{P}\{\zeta_n \leqslant k \mid \zeta_0 = i\} = \mathbf{P}\{\nu_1 + \cdots + \nu_n \leqslant n + k - i\}$$

$$- \sum_{j=i}^{n-1} \sum_{l=0}^{j-1} \left(1 - \frac{l}{j}\right) \mathbf{P}\{\nu_1 + \cdots + \nu_j = l\} \mathbf{P}\{\nu_{j+1} + \cdots + \nu_n = n + k - j\} \qquad (16)$$

if $k = 0, 1, \cdots$, *and, in particular,*

$$\mathbf{P}\{\zeta_n = 0 \mid \zeta_0 = i\} = \sum_{j=0}^{n-i} \left(1 - \frac{j}{n}\right) \mathbf{P}\{\nu_1 + \cdots + \nu_n = j\}. \tag{17}$$

PROOF. By (6)

$$\zeta_n = \max\{\nu_r + \cdots + \nu_n - (n - r) \text{ for } r = 1, \cdots, n \text{ and } \nu_1 + \cdots + \nu_n - n + \zeta_0\}.$$

If $\zeta_0 = i$, then the event $\{\zeta_n \leqslant k\}$ occurs if and only if

$$\nu_1 + \cdots + \nu_n \leqslant n + k - i$$

and

$$\nu_r + \cdots + \nu_n \leqslant n + k - r \qquad \text{for} \qquad r = i, \cdots, n. \tag{18}$$

Thus

$$\mathbf{P}\{\zeta_n \leqslant k \mid \zeta_0 = i\} = \mathbf{P}\{\nu_1 + \cdots + \nu_n \leqslant n + k - i\}$$
$$- \mathbf{P}\{\nu_1 + \cdots + \nu_n \leqslant n + k - i \text{ and } \nu_r + \cdots + \nu_n > n + k - r \text{ for some } r = i, \cdots, n\}. \tag{19}$$

The event that (18) is violated for some r can occur in several mutually exclusive ways: the smallest r for which (18) does not hold is $r = j + 1$ $(j = i, \cdots, n - 1)$. In this case

$$\nu_{j+1} + \cdots + \nu_n = n + k - j$$

must hold and obviously

$$\nu_r + \cdots + \nu_n \leqslant n + k - r \qquad \text{for} \qquad r = i, \cdots, j.$$

Accordingly,

$$\mathbf{P}\{\zeta_n \leqslant k \mid \zeta_0 = i\}$$

$$= \mathbf{P}\{\nu_1 + \cdots + \nu_n \leqslant n + k - i\} - \sum_{j=i}^{n-1} \mathbf{P}\{\nu_r + \cdots + \nu_n \leqslant n + k - r$$

$$\text{for } r = i, \cdots, j; \nu_{j+1} + \cdots + \nu_n = n + k - j \text{ and } \nu_1 + \cdots + \nu_n \leqslant n + k - i\}, \tag{20}$$

whence

$$
\begin{aligned}
&\mathbf{P}\{\zeta_n \leqslant k \mid \zeta_0 = i\} \\
&= \mathbf{P}\{\nu_1 + \cdots + \nu_n \leqslant n + k - i\} - \sum_{j=1}^{n-1} \mathbf{P}\{\nu_r + \cdots + \nu_j \leqslant j - r \\
&\quad \text{for } r = i, \cdots, j; \nu_1 + \cdots + \nu_j \leqslant j - i \text{ and } \nu_{j+1} + \cdots + \nu_n = n + k - j\} \\
&= \mathbf{P}\{\nu_1 + \cdots + \nu_n \leqslant n + k - i\} - \sum_{j=i}^{n-1} \mathbf{P}\{\nu_{j+1} + \cdots + \nu_n = n + k - j\}.
\end{aligned} \tag{21}
$$

$$
\mathbf{P}\{\nu_r + \cdots + \nu_j \leqslant j - r \text{ for } r = i, \cdots, j \text{ and } \nu_1 + \cdots + \nu_1 \leqslant j - i\},
$$

because $\nu_1, \cdots, \nu_n$ are mutually independent random variables. By Theorem 1

$$
\mathbf{P}\{\nu_r + \cdots + \nu_j \leqslant j - r \text{ for } r = i, \cdots, j \mid \nu_1 + \cdots + \nu_j = l\}
= \begin{cases} \left(1 - \dfrac{l}{j}\right) & \text{if} \quad 0 \leqslant l \leqslant j, \\ 0 & \text{otherwise,} \end{cases} \tag{22}
$$

whence

$$
\mathbf{P}\{\nu_r + \cdots + \nu_j \leqslant j - r \text{ for } r = i, \cdots, j \text{ and } \nu_1 + \cdots + \nu_j \leqslant j - i\}
= \sum_{l=0}^{j-i} \left(1 - \frac{l}{j}\right) \mathbf{P}\{\nu_1 + \cdots + \nu_j = l\}. \tag{23}
$$

Formulas (21) and (23) prove (16). If, in particular, $k = 0$, then

$$
\begin{aligned}
&\mathbf{P}\{\zeta_n = 0 \mid \zeta_0 = i\} \\
&= \mathbf{P}\{\nu_1 + \cdots + \nu_n \leqslant n - i \text{ and } \nu_r + \cdots + \nu_n \leqslant n - r \text{ for } r = i, \cdots, n\} \\
&= \sum_{j=0}^{n-i} \mathbf{P}\{\nu_1 + \cdots + \nu_n = j\} \mathbf{P}\{\nu_r + \cdots + \nu_n \leqslant n - r \text{ for} \\
&\qquad\qquad r = i, \cdots, n \mid \nu_1 + \cdots + \nu_n = j\}.
\end{aligned} \tag{24}
$$

By Theorem 1

$$
\mathbf{P}\{\nu_r + \cdots + \nu_n \leqslant n - r \text{ for } r = i, \cdots, n \mid \nu_1 + \cdots + \nu_n = j\}
= \begin{cases} \left(1 - \dfrac{j}{n}\right) & \text{if} \quad 0 \leqslant j \leqslant n, \\ 0 & \text{otherwise.} \end{cases} \tag{25}
$$

Putting (25) into (24) we get (17).

IV. The Limiting Distributions

If we suppose that $q_0 > 0$, and $q_0 + q_1 < 1$, then both $\{\xi_n\}$ and $\{\zeta_n\}$ are irreducible and aperiodic Markov chains. Consequently the limits

$$\lim_{n\to\infty} \mathbf{P}\{\xi_n = k \mid \xi_0 = i\} = P_k \qquad \text{and} \qquad \lim_{n\to\infty} \mathbf{P}\{\zeta_n = k \mid \zeta_0 = i\} = Q_k$$

exist irrespective of i. There are two possibilities: either $\{P_k\}$ ($\{Q_k\}$) is a probability distribution with positive elements and the Markov chain $\{\xi_n\}$ ($\{\zeta_n\}$) has a unique stationary distribution which agrees with the limiting distribution, or every $P_k = 0$ ($Q_k = 0$) and the Markov chain $\{\xi_n\}$ ($\{\zeta_n\}$) has no stationary distribution.

We can easily guess that if $\{\xi_n\}$ has a stationary distribution, then it is a geometric distribution. However, we shall provide a constructive proof based on Rouché's theorem and the following theorem of complex variables: If $f(z)$ is regular for all finite values of z and $\lim_{|z|\to\infty} f(z)/|z| = 0$, then $f(z)$ is a constant. The stationary distribution of $\{\zeta_n\}$ can easily be obtained by using generating functions.

In what follows we suppose that $q_0 > 0$, and $q_0 + q_1 < 1$. Let us introduce the generating function

$$g(z) = \sum_{j=0}^{\infty} q_j z^j, \tag{26}$$

which is convergent if $|z| \leqslant 1$, and let $\rho = g'(1-0)$.

THEOREM 4. *If $\rho > 1$, then the limiting distribution*

$$\lim_{n\to\infty} \mathbf{P}\{\xi_n = k \mid \xi_0 = i\} = P_k \qquad (k = 0, 1, \cdots)$$

exists and we have

$$P_k = (1-\omega)\,\omega^k \tag{27}$$

where $z = \omega$ is the only root of $g(z) = z$ in the unit circle $|z| < 1$. If $\rho \leqslant 1$, then

$$\lim_{n\to\infty} \mathbf{P}\{\xi_n = k \mid \xi_0 = i\} = 0$$

for every $k = 0, 1, \cdots$.

PROOF. First we prove that the equation $g(z) = z$ has only one root in the unit circle $|z| < 1$ if $\rho > 1$, and has no root if $\rho \leqslant 1$. If $\rho > 1$, then it follows from Rouché's theorem that $g(z) = z$ has one and only one root in the circle $|z| < 1 - \epsilon$ where $\epsilon > 0$ is small enough. For,

$$|g(z)| \leqslant g(1-\epsilon) < 1 - \epsilon$$

if $|z| = 1 - \epsilon$ and ϵ is a sufficiently small positive number. If $\rho \leqslant 1$, then $|g'(z)| < 1$ for $|z| < 1$, whence

$$|1 - g(z)| = \left| \int_z^1 g'(\zeta)\, d\zeta \right| < |1 - z|$$

for $|z| < 1$. Thus $g(z) = z$ is impossible if $|z| < 1$.

Now if we suppose that $\{\xi_n\}$ has a stationary distribution $\{P_k\}$ and introduce the generating function

$$P(z) = \sum_{k=0}^{\infty} P_k z^k, \tag{28}$$

then by (5) we obtain that for $|z| = 1$

$$P(z) = P(z)\, zg\left(\frac{1}{z}\right) + \sum_{j=0}^{\infty} C_j \left(1 - \frac{1}{z^j}\right) \tag{29}$$

where

$$C_j \geqslant 0 \quad (j = 0, 1, \cdots) \qquad \text{and} \qquad C_0 + C_1 + \cdots + C_j + \cdots < 1.$$

Hence for $|z| = 1$ and $z \neq 1$

$$P(z) = \frac{\sum_{j=0}^{\infty} C_j[1 - (1/z^j)]}{1 - zg(1/z)}. \tag{30}$$

Now let us define $P(z)$ also for $|z| > 1$ by (30). By definition $P(z)$ has no singularities for $|z| \leqslant 1$. Thus $P(z)$ has singularities only at the zeros of the denominator of (30) outside the unit circle. These zeros evidently agree with the reciprocal values of the roots of $g(z) = z$ inside the unit circle. If $\rho > 1$, then there is one root $z = \omega$. If $\rho \leqslant 1$, then there is no such root.

If we suppose that $\rho > 1$, then $P(z)\,(z - 1/\omega)$ will be a regular function of z on the whole complex plane. Since obviously,

$$\lim_{|z| \to \infty} P(z) \frac{z - (1/\omega)}{|z|} = 0,$$

therefore $P(z)\,[z - (1/\omega)]$ is a constant. Since $P(1) = 1$, we get finally

$$P(z) = \frac{1 - \omega}{1 - \omega z}, \tag{31}$$

whence (27) follows.

If $\rho \leqslant 1$, then $P(z)$ is a regular function of z on the whole complex plane. Since $\lim_{|z| \to \infty} P(z) = 0$, therefore $P(z) = 0$. Thus in this case a stationary distribution does not exist. This completes the proof of the theorem.

THEOREM 5. *If $\rho < 1$, then the limiting distribution*

$$\lim_{n\to\infty} \mathbf{P}\{\zeta_n = k \mid \zeta_0 = i\} = Q_k \qquad (k = 0, 1, \cdots)$$

exists and we have

$$Q(z) = \sum_{k=0}^{\infty} Q_k z^k = \frac{(1-\rho)(1-z)g(z)}{g(z)-z}. \tag{32}$$

If $\rho \geqslant 1$, then

$$\lim_{n\to\infty} \mathbf{P}\{\zeta_n = k \mid \zeta_0 = i\} = 0$$

for every $k = 0, 1, \cdots$.

PROOF. Suppose that a stationary distribution $\{Q_k\}$ exists. If $Q(z)$ denotes the generating function of the stationary distribution, then by (6)

$$Q(z) = \left(Q_0 + \frac{Q(z) - Q_0}{z}\right) g(z),$$

whence

$$Q(z) = Q_0 \frac{(1-z)g(z)}{g(z)-z}. \tag{33}$$

Since $Q(1) = 1$, we get from (33) that $Q_0 = 1 - \rho$. If $\rho < 1$, then there is a stationary distribution and its generating function is given by (32). If $\rho \geqslant 1$, then the assumption that a stationary distribution exists leads to a contradiction. Thus a stationary distribution cannot exist if $\rho \geqslant 1$. This completes the proof of the theorem.

REFERENCES

1. BERTRAND, J. Solution d'un problème. *C.R. Acad. Sci. Paris* **105**, 369 (1887).
2. ANDRÉ, D. Solution directe du problème résolu par M. Bertrand. *C.R. Acad. Sci. Paris* **105**, 436-437 (1887).
3. BARBIER, É. Généralisation du problème résolu par M. J. Bertrand. *C.R. Acad. Sci. Paris* **105**, 407 (1887).
4. TAKÁCS, L. "Introduction to the Theory of Queues." Oxford Univ. Press, New York, 1962.
5. TAKÁCS, L. The probability law of the busy period for two types of queuing processes. *Operations Res.* **9**, 402-407 (1961).
6. TAKÁCS, L. A generalization of the ballot problem and its application in the theory of queues. *J. Am. Statist. Assoc.* **57**, 327-337 (1962).