SOME PROPERTIES OF REGULAR MARKOV CHAINS

By B. R. Bhat[1]

University of California, Berkeley[2]

Summary. In a regular Markov chain with one absorbing state, for sequences starting from a given state and continuing until the absorbing state is reached, the distribution and moment formulae of the total number of transitions is derived in Section 2, and also its probability generating function (p.g.f.). The joint p.g.f. of the transition frequencies is given in Section 3, from which the p.g.f. of one or more transition frequencies is deduced. In Section 4 some moment formulae associated with these transition frequencies are derived. Section 5 is concerned with inference for such Markov chains, when there are a large number of sequences starting from the same given state.

1. Introduction. Consider a time-homogeneous Markov chain with a finite number $s + 1$ of states $E_0, E_1, \cdots, E_s$. Let

$$\mathbf{P} = \{p_{ij}\} \qquad (i, j = 0, 1, \cdots, s) \tag{1}$$

be the matrix of transition probabilities, where $p_{ij} = \Pr(E_j \mid E_i)$. We define regularity and positive regularity of a chain as in [2]. The necessary and sufficient condition that the chain is regular is that the only latent root $\lambda_0 = 1$ of modulus unity of $\mathbf{P}$ is simple. For a positively regular Markov chain $\mathbf{P}$ is irreducible, but for a regular, but not positively regular, chain it can be expressed in the form

$$\mathbf{P} = \begin{pmatrix} \mathbf{Q} & \mathbf{O} \\ \mathbf{R} & \mathbf{S} \end{pmatrix}, \tag{2}$$

where $\mathbf{Q}$ is a $(r + 1) \times (r + 1)$ submatrix of transition probabilities between $E_0, E_1, \cdots, E_r$ $(s \geqq r + 1 > 0)$ and is irreducible (cf., Bartlett [2]). For convenience, the states $E_0, E_1, \cdots, E_r$ will be called absorbing states; $E_{r+1}, \cdots, E_s$ transient states. It is readily seen that the simple latent root $\lambda_0 = 1$ of $\mathbf{P}$ is also a latent root of $\mathbf{Q}$, so that the latent roots λ_i of $\mathbf{S}$ must all have moduli less than unity. Further, $\mathbf{R} \neq \mathbf{O}$, since otherwise the latent root λ_0 will be of multiplicity greater than one and the chain will not be regular.

Sequences from a regular Markov chain may be classified into three categories as follows: (i) those starting and stopping with an absorbing state, (ii) those starting with a transient state and stopping with an absorbing state and (iii) those starting and stopping with a transient state.

Received October 22, 1959; revised April 15, 1960.

[1] On leave from Karnatak University, Dharwar, India.

[2] This work was carried out while the author was at the University of Western Australia, Nedlands, Australia. It was revised at the University of California, Berkeley with the partial support of the Office of Naval Research (Nonr-222-43). This paper in whole or in part may be reproduced for any purpose of the United States Government.

Of these we shall consider here only those belonging to category (ii). But a sequence of this category may be split up into two sections, the first consisting of those transitions until one of the absorbing states is reached, and the second consisting of the remaining transitions. This latter section is a sequence starting and stopping with an absorbing state and belongs to category (i). Since these two sections can be studied independently we shall restrict our study to that of the former section. For such a study, without loss of generality, we may assume that there is only one absorbing state E_0. Thus, in this paper, by a sequence from a regular Markov chain we mean one starting with a given transient state and continuing until the absorbing state is reached. Evidently for such a sequence the total number of transitions is a random variable. For this interesting case we shall derive the distributions and moment formulae of transition frequencies, and also give some related results.

Properties of finite Markov chains have been studied by many authors; among them we may mention Romanovsky [10], Fréchet [7], Feller [6] and Bartlett [2]. The distributions of transition frequencies in sequences of fixed size from a finite Markov chain have been studied by Whittle [11] and Goodman [8]; Anderson and Goodman [1] have derived the variance-covariance matrix of the frequencies. It will be seen from this paper that similar distribution and moment formulae are available when the total number of transitions is a random variable.

It is well known that sequences from a positively regular Markov chain may be considered to be made up of a series of independent sequences, starting with a random initial state and stopping as soon as a given state is reached [6]. Thus, the present study gives an insight into the properties of sequences from a positively regular Markov chain. It will also be useful in solving first emptiness problems connected with dams and queues [9].

2. Distribution of the total number of transitions. If a sequence starts with one of the transient states E_a ($a \neq 0$), let it be absorbed at E_0 after the nth transition. The distribution of n is derived by Feller ([6], Section 16.4) and Bartlett ([2], p. 68), but it is given here for completeness.

When E_0 is the only absorbing state, (2) can be written as

$$\mathbf{P} = \begin{pmatrix} 1 & \mathbf{O} \\ \mathbf{R} & \mathbf{S} \end{pmatrix}, \tag{3}$$

where $\mathbf{R}$ is a column vector. Let $f_a(n)$ be the probability that a sequence starting from E_a will be at E_0 for the first time after the nth transition. Then $f_a(n)$ is the ath element of the column vector

$$\mathbf{S}^{n-1}\mathbf{R}. \tag{4}$$

Since it can be verified that

$$\sum_{n=1}^{\infty} f_a(n) = 1,$$

$\{f_a(n)\}$ is the required probability distribution of n.

Now we shall derive the moments of n. Let $\lambda_1, \lambda_2, \cdots, \lambda_t$ be the t distinct latent roots of $\mathbf{S}$ with multiplicities $m_1, m_2, \cdots, m_t$ respectively. Then

$$\mathbf{S}^r = \sum_{k=1}^{t} \lambda_k^r \mathbf{C}_k(r),$$

where the elements of the spectral matrices $\mathbf{C}_k(r)$ are polynomials of degree not greater than $m_k - 1$ in r. So

$$\mathbf{S}^{n-1}\mathbf{R} = \sum_{k=1}^{t} \lambda_k^{n-1} \mathbf{C}_k(n-1)\mathbf{R}$$

and

$$f_a(n) = \sum_{k=1}^{t} \lambda_k^{n-1} C_{ak}(n-1),$$

where $C_{ak}(n-1)$ is a polynomial in $(n-1)$ of degree at most $m_k - 1$. Let $C_{ak}(n-1) = b_{ak}^0 + b_{ak}^1(n-1)^{(1)} + b_{ak}^2(n-1)^{(2)} + \cdots + b_{ak}^{m_k-1}(n-1)^{(m_k-1)}$ where

$$n^{(r)} = n(n-1) \cdots (n-r+1).$$

Then we see that the expectations of n and n^2 are given by

$$E(n) = \sum_{k=1}^{t} \sum_{r=0}^{m_k-1} \frac{b_{ak}^r \lambda_k^r (r+1)!}{(1-\lambda_k)^{r+2}}$$

and

$$E(n^2) = \sum_k \sum_r b_{ak}^r \lambda_k^r (r+1)! \left[\frac{r+1+\lambda_k}{(1-\lambda_k)^{r+3}}\right],$$

respectively, for sequences starting with the state E_a.

If all the m_k are equal to unity, the formulae for $E(n)$ and $E(n^2)$ reduce to

$$E(n) = \sum_k \frac{b_{ak}^0}{(1-\lambda_k)^2} \qquad E(n^2) = \sum_k \frac{1+\lambda_k}{(1-\lambda_k)^3} b_{ak}^0 .$$

Alternatively, the p.g.f. may be used to evaluate $f_a(n)$. It is particularly useful when $\mathbf{R}$ has only a few nonzero elements. The p.g.f. of n is the ath element of

$$\mathbf{G}(z; n) = \sum_{n=1}^{\infty} z^n \mathbf{S}^{n-1}\mathbf{R} = (\mathbf{I} - z\mathbf{S})^{-1} z\mathbf{R} = \frac{adj(\mathbf{I} - z\mathbf{S})}{|\mathbf{I} - z\mathbf{S}|} z\mathbf{R},$$

where $\mathbf{I}$ is the unit matrix.

This ath element is also equal to

$$G_a(z; n) = D_a(z)/D(z), \tag{5}$$

where $D(z) = |\mathbf{I} - z\mathbf{S}|$, and $D_a(z)$ is the determinant $D(z)$ with its ath column replaced by $z\mathbf{R}$.

Further, it can easily be verified that the distribution of n is geometric with parameter $1 - p$ if all the elements of $\mathbf{R}$ are equal to p.

EXAMPLE. Let the transition probability matrix (3) be

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 \\ 1-p & p & 0 \\ 0 & 1-q & q \end{pmatrix}.$$

The latent roots of $\mathbf{S}$ are p and q, and the corresponding spectral matrices are

$$\begin{pmatrix} 1 & 0 \\ \dfrac{q-1}{q-p} & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 & 0 \\ \dfrac{q-1}{p-q} & 1 \end{pmatrix}.$$

Thus,

$$\mathbf{S}^{n-1}\mathbf{R} = p^{n-1}\begin{pmatrix} 1-p \\ \dfrac{(q-1)(1-p)}{q-p} \end{pmatrix} + q^{n-1}\begin{pmatrix} 0 \\ \dfrac{(q-1)(1-p)}{p-q} \end{pmatrix},$$

and therefore

$$f_1(n) = p^{n-1}(1-p),$$

$$f_2(n) = \frac{(q-1)(1-p)}{q-p}\,[p^{n-1} - q^{n-1}].$$

The p.g.f. of n, for sequences starting from E_1 is

$$G_1(z; n) = z(1-p)/(1-zp)$$

as can be verified otherwise; and for sequences starting from E_2, it is

$$G_2(z; n) = \frac{z^2(1-p)(1-q)}{(1-zp)(1-zq)},$$

$$= \frac{z(1-p)}{1-zp} \cdot \frac{z(1-q)}{1-zq}.$$

Hence the distribution of n in this case is the convolution of two geometric distributions with parameters p and q respectively.

In exactly the same way, if in general the transition probability matrix is

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 1-p_1 & p_1 & 0 & \cdots & 0 & 0 \\ 0 & 1-p_2 & p_2 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & & & 1-p_s & p_s \end{bmatrix}$$

the distribution of n for sequences starting from E_s is found to be the convolution of s geometric distributions with parameters $p_1, p_2, \cdots$, and p_s.

3. Distribution of transition frequencies. In a sequence from the Markov chain described in the previous sections, let n_{ij} be the frequency of transitions from E_i to E_j $(i, j = 0, 1, \cdots, s)$. These n_{ij} may also be viewed as the number of times the pair of states (E_i, E_j) occur in the sequence.

(i) *Joint distribution of* n_{ij}. Let

$$\mathbf{S}(z_{ij}) = \begin{pmatrix} z_{11}p_{11} & z_{12}p_{12} & \cdots & z_{1s}p_{1s} \\ z_{21}p_{21} & z_{22}p_{22} & \cdots & z_{2s}p_{2s} \\ \cdot & \cdot & \cdots & \cdot \\ z_{s1}p_{s1} & z_{s2}p_{s2} & \cdots & z_{ss}p_{ss} \end{pmatrix}$$

and

$$\mathbf{R}(z_{ij}) = \begin{Bmatrix} z_{10}p_{10} \\ z_{20}p_{20} \\ \cdot \\ \cdot \\ \cdot \\ z_{s0}p_{s0} \end{Bmatrix}$$

Then as in (5) the joint p.g.f. of n_{ij} is the ath element of $(\mathbf{I} - \mathbf{S}(z_{ij}))^{-1}\mathbf{R}(z_{ij})$. Let the ath element be

$$G_a(z_{ij}\,;n_{ij}) = D_a(z_{ij})/D(z_{ij}), \tag{6}$$

where $D(z_{ij}) = |\mathbf{I} - \mathbf{S}(z_{ij})|$ and $D_a(z_{ij})$ is the determinant $D(z_{ij})$ with its ath column replaced by $\mathbf{R}(z_{ij})$.

Formula (6) can also be derived as follows. Let the generating function of the probabilities of observing the frequencies n_{ij}, for sequences starting from E_a, such that they satisfy the relation $n_{i\cdot} - n_{\cdot i} = \delta_{ia} - \delta_{ib}$ $(i = 1, 2, \cdots, s)$ (i.e., the sequence stops at E_b), be $G_{ab}(z_{ij}\,;n_{ij})$, where $n_{i\cdot} = \sum_{j=1}^{s} n_{ij}$, $n_{\cdot i} = \sum_{j=1}^{s} n_{ji}$ and δ_{ij} is the Kronecker delta. From Whittle [11], it may be written as

$$G_{ab}(z_{ij}\,;n_{ij}) = \Delta_{ab}(z_{ij})/\Delta(z_{ij}),$$

where $\Delta(z_{ij}) = |\mathbf{I} - \mathbf{P}(z_{ij})|$ and $\Delta_{ab}(z_{ij})$ is the cofactor of the (b, a)th element of the determinant $\Delta(z_{ij})$. Here $\mathbf{P}(z_{ij})$ is defined similarly to $\mathbf{S}(z_{ij})$. But from (3),

$$\Delta(z_{ij}) = (1 - z_{00}) \mid \mathbf{I} - \mathbf{S}(z_{ij}) \mid = (1 - z_{00})D(z_{ij})$$

and for sequences starting from E_a and stopping at E_0, $\Delta_{a0}(z_{ij}) = D_a(z_{ij})$. Since n_{00} vanishes for realizations stopping as soon as E_0 is reached, the joint p.g.f. of n_{ij} is $D_a(z_{ij})/D(z_{ij})$, as given by (6).

It can be verified that $G_a(1; n_{ij}) = 1$.

To obtain the explicit expression for the joint probability distribution of the n_{ij} we need to expand (6) in ascending powers of z_{ij}. It may be noted that since $\mathbf{S}$ has all its latent roots less than unity, $\mathbf{S}(z_{ij})$ has its latent roots less than unity for all values of $0 \leqq z_{ij} \leqq 1$ $(i, j = 1, 2, \cdots, s)$. Hence $(\mathbf{I} - \mathbf{S}(z_{ij}))$ is nonsingular for z_{ij} in the above range.

Let the required probability $\pi_a(n_{ij})$ involve

$$p_{b0} \prod_{i,j=1}^{s} p_{ij}^{n_{ij}}. \tag{7}$$

The numerical coefficient of this term is the same as that of

$$\prod_{i,j=1}^{s} p_{ij}^{n_{ij}}$$

in

$$D_{ab}(z_{ij})/D(z_{ij}) \tag{8}$$

where D_{ab} is the cofactor of the (b, a)th element of $D(z_{ij})$. But the latter combinatorial formula has been evaluated by Whittle [11] as

$$T_{ab}(n_{ij}) \frac{\prod_{i=1}^{s} n_{i}.!}{\prod_{i,j=1}^{s} n_{ij}!} \tag{9}$$

where T_{ab} is the cofactor of the (b, a)th element of the $s \times s$ matrix

$$[\delta_{ij} - (n_{ij}/n_{i\cdot})] \tag{10}$$

if $n_{i\cdot} - n_{\cdot i} = \delta_{ia} - \delta_{ib}$, and zero otherwise. Thus $\pi_a(n_{ij})$ is the product of (7) and (9), where E_b is the last nonabsorbing state of the observed sequence. Since the sum of elements in any row of (10) vanishes, by the lemma in the Appendix (cf. also Goodman [8]), T_{ab} is the same for all values of a for fixed value of b.

An alternate derivation of $\pi_a(n_{ij})$ is due to Goodman (private communication). Let $\phi_a(b, n_{ij})$ be the joint distribution of b and n_{ij} $(i, j = 1, 2, \cdots, s)$ for a given a and n, the total number of frequencies; an explicit expression for $\phi_a(b, n_{ij})$ has been given in [8]. Since $n_{i\cdot} - n_{\cdot i} = \delta_{ia} - \delta_{ib}$ $(i = 1, 2, \cdots, s)$, for a given a, b is uniquely determined by n_{ij} $(i, j = 1, 2, \cdots, s)$ as a function $b(n_{ij})$. Thus the joint distribution of n_{ij} $(i, j = 1, 2, \cdots, s)$ when n is random is $\phi_a[b(n_{ij}), n_{ij}]p_{b0}$, which is the required probability $\pi_a(n_{ij})$.

(ii) *Distribution of* $n_{\alpha\beta}$. From the joint p.g.f. of all the n_{ij}, we can easily get the p.g.f. of a particular n_{ij}, $n_{\alpha\beta}(n_x$, say), by putting

$$z_{ij} \begin{cases} = z_x & (i = \alpha, j = \beta), \\ = 1 & \text{otherwise}, \end{cases} \tag{11}$$

in (6). The required p.g.f. is

$$G_a(z_x\,; n_x) = D_a(z_x)/D(z_x),$$

where $D_a(z_x)$ and $D(z_x)$ are the determinants $D_a(z_{ij})$ and $D(z_{ij})$ subject to (11). Because these are linear functions of z_x, we may write

$$G_a(z_x; n_x) = \frac{Q_{a,x} + z_x\, p_x\, P_{a,x}}{Q_x + z_x\, p_x\, P_x}, \tag{12}$$

where P_x, $P_{a,x}$, Q_x, $Q_{a,x}$ do not involve z_x.

From (12) the probability distribution of n_x, $p_a(n_x)$, may be derived. In fact

$$p_a(n_x) = \begin{cases} \left(-\frac{p_x P_x}{Q_x}\right)^{n_x} \left(\frac{Q_{a,x}}{Q_x} - \frac{P_{a,x}}{P_x}\right), & n_x = 1, 2 \cdots, \infty \\ \frac{Q_{a,x}}{Q_x}, & n_x = 0. \end{cases} \tag{13}$$

It is to be noted that Q_x does not vanish since $|\mathbf{I} - \mathbf{S}(z_{ij})|$ is nonsingular for all $0 \leqq z_{ij} \leqq 1$. Thus, the distribution of n_x is geometric, with a modified first term; it will be geometric for $\beta = a$, since $P_{a,x}$ vanishes in this case. These latter results can also be proved by using the theory of recurrent events as given by Feller ([6], Chapter 13).

Since $-P_x$ is the cofactor of the (α, β)th element of $D(z_x)$, it will vanish if transition from E_β to E_α is impossible. In this case $n_x = 0$ or 1, as can also be verified from (12). It is interesting to note that if n_x is a fixed number, its value is zero or one.

(iii) *Joint distribution of n_x and n_y*. By putting $z_{ij} = 1$ for all values of i and j, except $z_{\alpha\beta}(= z_x)$ and $z_{\gamma\delta}(= z_y)$ in (6) we get the joint p.g.f. of n_x and n_y. It may be written as

$$G_a(z_x, z_y; n_x, n_y) = \frac{D_a(z_x, z_y)}{D(z_x, z_y)} = \frac{R_a + z_x S_a + z_y T_a + z_x z_y U_a}{R + z_x S + z_y T + z_x z_y U} \tag{14}$$

where R, R_a, etc., do not involve z_x or z_y.

4. Moment formulae. In this section we shall derive some moment formulae associated with the n_{ij}.

(a) From (12) we have

$$E(n_x) = \frac{P_x(Q_x P_{a,x} - P_x Q_{a,x})}{(Q_x + p_x P_x)^2}. \tag{15}$$

Since

$$G_a(1; n_x) = 1,$$
$$D(1) = Q_x + p_x P_x = Q_{a,x} + p_x P_{a,x} = D_a(1)$$

and hence it readily follows that

$$\begin{aligned} Q_x P_{a,x} - P_x Q_{a,x} &= P_{a,x}\{D(1) - p_x P_x\} - P_x\{D_a(1) - p_x P_{a,x}\} \\ &= D(1)(P_{a,x} - P_x). \end{aligned}$$

But $P_{a,x} - P_x$ is the coefficient of $-p_x$ in the expansion of $M = D(1) - D_a(1)$. Since D_a and D differ only in their ath column, M is a determinant for which the sum of elements in any row equals zero. Further, since $E(n_{ij}) > 0$ for at least one i, j, all the minors of order $s - 1$ of M do not vanish. Hence from the lemma in the Appendix, the cofactors of elements from any one row of M are equal. To determine their actual value we see that

$$E(n_{\alpha a}) = -p_{\alpha a}P_{\alpha a}/D(1),$$

where $P_{\alpha a}$ is the coefficient of $p_{\alpha a}$ in $D(1)$. Hence

(16) $$P_{a,x} - P_x = -P_{\alpha a}.$$

Substituting,

(17) $$E(n_x) = -p_x P_{\alpha a}/D(1).$$

It may be noted that (17) holds true for $\beta = 0, 1, 2, \cdots, s$ as can be verified from (6) or otherwise. Hence we get

(18) $$E\left(\sum_{\beta=0}^{s} n_{\alpha\beta}\right) = \frac{-P_{\alpha a}}{D(1)} = E(n_{\alpha .})$$

from which we can see that

(19) $$E(n_{\alpha\beta})/E(n_{\alpha\cdot}) = p_{\alpha\beta}.$$

But, in general,

$$E(n_{\alpha\beta}/n_{\alpha\cdot}) \neq p_{\alpha\beta}.$$

To evaluate $\mathrm{Var}(n_x)$, we note that

$$E[n_x(n_x - 1)] = 2p_x^2 P_x[P_x Q_{a,x} - Q_x P_{a,x}]/[D(1)]^3.$$

Hence from (16) and (17)

(20) $$\mathrm{Var}(n_x) = \frac{-p_x P_{\alpha a}}{[D(1)]^2}\{D(1) - 2p_x P_x\} - \left[\frac{p_x P_{\alpha a}}{D(1)}\right]^2.$$

(b) From (19)

(21) $$E(n_x - p_x n_{\alpha\cdot}) = 0 \qquad (\alpha, \beta = 1, 2, \cdots, s).$$

Now we derive the variance-covariance matrix of $n_{ij} - p_{ij}n_{i\cdot}$.

Differentiating (14) with respect to z_x and z_y and putting $z_x = z_y = 1$, we have

(22) $$E(n_x n_y) = [D(1)D''_{a,xy} - \{D'_{a,x}D'_y + D'_{a,y}D'_x + D_a(1)D''_{xy}\} + 2D'_xD'_y]/[D(1)]^2,$$

where

$$D_a(1) = R_a + S_a + T_a + U_a = D(1) = R + S + T + U,$$

$$D'_{a,x} = S_a + U_a, \qquad D'_x = S + U,$$

$$D'_{a,y} = T_a + U_a, \qquad D'_y = T + U,$$

$$D''_{a,xy} = U_a, \qquad D''_{xy} = U,$$

and

$$S_a + U_a = \text{coefficient of } z_x \text{ in } D_a(z) = p_x P_{a,x}\,,$$
$$(23) \qquad T_a + U_a = \text{coefficient of } z_y \text{ in } D_a(z) = p_y P_{a,y}\,,$$
$$U_a = \text{coefficient of } z_x z_y \text{ in } D_a(z).$$

Notice that S, T, U also satisfy relations similar to (23). Thus from (16) we see that

$$S_a + U_a - S - U = -p_x P_{\alpha a}\,,$$
$$(24) \qquad T_a + U_a - T - U = -p_y P_{\gamma a}\,,$$
$$U_a - U = -p_x p_y P_{x;y}\,,$$

where $P_{x;y}$ is the coefficient of $p_x p_y$ in M.

Substituting (23) and (24) in (22), we have

$$(25) \qquad E(n_x n_y) = \frac{-\,p_x p_y P_{x;y}}{D(1)} + \frac{p_x P_x p_y P_{\gamma a} + p_y P_y p_x P_{\alpha a}}{[D(1)]^2}$$

for all values of x and y.

CASE (i): $\alpha \neq \gamma$, $\beta \neq \delta$. In this case we have

$$(26) \quad E(n_x n_{\gamma\cdot}) = -\,\frac{p_x(P_{\alpha a} + P_{x;\gamma})}{D(1)} + \frac{p_x P_x P_{\gamma a} + [D(1) + P_{\gamma\gamma}]p_x P_{\alpha a}}{[D(1)]^2}$$

since $\sum_\delta p_y P_{x;y} = P_{\alpha a} + P_{x;\gamma}$ and $\sum_\delta p_y P_y = D(1) + P_{\gamma\gamma}$, where $P_{x;\gamma}$ is the coefficient of $p_x p_{\gamma\gamma}$ in M.

Similarly we can derive

$$(27) \quad E(n_y n_{\alpha\cdot}) = -\,\frac{p_y(P_{\gamma a} + P_{x;\alpha})}{D(1)} + \frac{p_y P_y P_{\alpha a} + [D(1) + P_{\alpha\alpha}]p_y P_{\gamma a}}{[D(1)]^2}$$

and

$$(28) \quad E(n_{\alpha\cdot} n_{\gamma\cdot}) = -\,\frac{P_{\alpha a} + P_{\gamma a} + P_{\alpha;\gamma}}{D(1)} + \frac{P_{\gamma a}[D(1) + P_{\alpha a}] + P_{\alpha a}[D(1) + P_{\gamma\gamma}]}{[D(1)]^2}$$

where $P_{\alpha;\gamma}$ and $P_{x;\alpha}$ are defined similarly to $P_{x;\gamma}$.

Now $P_{x;y}$, $P_{x;\gamma}$, $P_{y;\alpha}$ and $P_{\alpha;\gamma}$ are determinants having all but two columns alike. Transferring these common columns to the same place as those in $P_{\alpha;\gamma}$ we get $\bar{P}_{x;y}$, $\bar{P}_{x;\gamma}$ and $\bar{P}_{y;\alpha}$ respectively.

Since

$$P_{x;\gamma} = -\,|\bar{P}_{x;\gamma}|,$$
$$P_{y;\alpha} = -\,|\bar{P}_{y;\alpha}|,$$
$$P_{x;y} = +\,|\bar{P}_{x;y}|,$$

we see that

$$P_{\alpha;\gamma} + P_{x;y} - P_{x;\gamma} - P_{y;\alpha} = \{|P_{\alpha;\gamma}| + |\bar{P}_{x;\gamma}|\} + \{|\bar{P}_{y;\alpha}| + |\bar{P}_{x;y}|\}.$$
$$= |\bar{Q}_{\alpha;\gamma}| + |\bar{Q}_{y;\alpha}|$$
$$= |\bar{R}_{\alpha;\gamma}|, \text{ say,}$$

where $\bar{R}_{\alpha;\gamma}$ is a determinant with the sum of elements in any row equal to that in M, that is, zero. Hence

(29) $$|\bar{R}_{\alpha;\gamma}| = 0.$$

From (25)–(29) we have

(30) $$E\{(n_x - p_x n_{\alpha\cdot})(n_y - p_y n_{\gamma\cdot})\} = 0.$$

Case (ii): $\alpha = \gamma$, $\beta \neq \delta$. In this case U_a and U vanish.

(31) $$E(n_x n_y) = \frac{p_x p_y P_{\alpha\alpha}(P_x + P_y)}{[D(1)]^2}$$

$$E(n_x n_{\alpha .}) = E(n_x^2) + \sum_{\delta \neq \beta} E(n_x n_y),$$

(32) $$= \frac{-p_x P_{\alpha\alpha}}{D(1)} + \sum_{\delta} \frac{p_x p_y P_{\alpha\alpha}(P_x + P_y)}{[D(1)]^2}$$
$$= \frac{-p_x P_{\alpha\alpha}}{D(1)} + \frac{p_x P_{\alpha\alpha}\{P_x + D(1) + P_{\alpha\alpha}\}}{[D(1)]^2}$$

since from (20),

$$E(n_x^2) = \frac{-p_x P_{\alpha\alpha}}{D(1)}\left[1 - \frac{2p_x P_x}{D(1)}\right].$$

Further

$$E(n_{\alpha .}^2) = \frac{P_{\alpha\alpha}}{D(1)}\left(1 + \frac{2P_{\alpha\alpha}}{D(1)}\right).$$

Hence

(33) $$E\{(n_x - p_x n_{\alpha .})(n_y - p_y n_{\alpha .})\} = \frac{p_x p_y P_{\alpha\alpha}}{D(1)}.$$

Case iii: $\alpha = \gamma$, $\beta = \delta$. When $\alpha = \gamma$, $\beta = \delta$ we obtain

$$E(n_x - p_x n_{\alpha .})^2 = \operatorname{Var}(n_x - p_x n_{\alpha .})$$

(34) $$= \frac{-p_x P_{\alpha\alpha}(1 - p_x)}{D(1)}.$$

Comparing (30), (33) and (34) we may write

$$\operatorname{Cov}(n_x - p_x n_{\alpha .}, n_y - p_y n_{\delta .})$$

(35) $$= \frac{-\delta_{\alpha\gamma}(\delta_{\beta\delta} p_x - p_x p_y) P_{\alpha\alpha}}{D(1)} \quad (\alpha, \beta, \gamma, \delta = 0, 1, \cdots, s).$$

It is interesting to note that (35) is of the same form as that obtained in the case when n is nonrandom [1].

5. Inference. Statistical inference for Markov chains, when the number of transitions in any sequence is nonrandom is considered by Anderson and Goodman [1]. In this section we shall give some analogous results for the case when the number of transitions is a random variable.

(i) *Estimation of* p_{ij}. Let the transition probabilities p_{ij} be unknown. They are to be estimated when there are a large number of sequences starting from the same given state E_a. Let $S_1, S_2, \cdots, S_m$ be m such sequences and n_{ijk} be the number of transitions from E_i to E_j in S_k $(k = 1, 2, \cdots, m)$. Since $n_{ijk} = 0$ when $p_{ij} = 0$, without loss of generality we shall assume that $p_{ij} > 0$.

Since n_{ijk} $(k = 1, 2, \cdots, m)$ are independently and identically distributed with finite variance, by the law of large numbers,

$$\frac{1}{m} \sum_k n_{ijk} = \bar{n}_{ij}$$

tends to $E(n_{ijk})$ in probability as $m \to \infty$. Hence the maximum likelihood estimate,

$$\hat{p}_{ij} = \bar{n}_{ij}/\bar{n}_{i\cdot}, \tag{36}$$

where $\bar{n}_{i\cdot} = \sum_j \bar{n}_{ij}$, tends to

$$E(n_{ijk})/E(n_{i\cdot k}) = p_{ij}$$

in probability (cf., Cramér [4]). This result may be compared with that for the positively regular case [3].

Further, as $m \to \infty$, since $\bar{n}_{i\cdot}$ tends to

$$E(n_{i\cdot k}) = -P_{ia}/D(1) > 0 \tag{37}$$

in probability,

$$(m)^{\frac{1}{2}}(\hat{p}_{ij} - p_{ij}) = -[(m)^{\frac{1}{2}}(\bar{n}_{ij} - p_{ij}\bar{n}_{i\cdot})]/n_{i\cdot}$$

has the same limiting distribution as

$$-\frac{(n_{ij\cdot} - p_{ij}\, n_{i\cdot\cdot})}{(m)^{\frac{1}{2}}[P_{ia}/D(1)]}, \tag{38}$$

where $n_{ij\cdot} = \sum_k n_{ijk}$, $n_{i\cdot\cdot} = \sum_j \sum_k n_{ijk}$.

But the numerator of (38) is a sum of m independent linear functions, all following the same distribution with mean zero and variance-covariance matrix

$$-\frac{\delta_{ii'}(\delta_{jj'}\, p_{ij} - p_{ij}\, p_{i'j'})P_{ia}}{D(1)}, \qquad \begin{pmatrix} i, i' = 1, 2, \cdots, s \\ j, j' = 0, 1, \cdots, s \end{pmatrix}$$

from (35). Hence the $(m)^{\frac{1}{2}}(\hat{p}_{ij} - p_{ij})$ have an asymptotic multivariate normal distribution with mean zero and variance-covariance matrix

$$-[\delta_{ii'}(\delta_{jj'}p_{ij} - p_{ij'}p_{ij'})D(1)]/P_{ia} \tag{39}$$

(cf., Cramér [5]).

(ii) *Testing of hypotheses.* If p_{ij}^0 is the true value of p_{ij}, from (37) and (39), it is clear that, for each $i = 1, 2, \cdots, s$, the

$$(m\bar{n}_{i\cdot})^{\frac{1}{2}}(\hat{p}_{ij} - p_{ij}^0) \qquad (j = 0, 1, \cdots, s)$$

have an asymptotic normal distribution with variances and covariances depending on p_{ij}^0 in the same way as those obtained for multinomial estimates. Using this limiting distribution we can test hypotheses about one or more p_{ij} or determine a confidence region for one or more p_{ij}.

(iii) *Test of the hypothesis that several sets of sequences are from the same Markov chain.* Let there be t sets of $m_1, m_2, \cdots, m_t$ sequences, each starting from the same given state E_a, from Markov chains possibly with different transition probability matrices. If $(p_{ij}^{(h)})$ is the transition probability matrix for the hth set $(h = 1, 2, \cdots, t)$, we want to test the hypothesis that

$$p_{ij}^{(h)} = p_{ij} \qquad (i, j = 0, 1, \cdots, s; h = 1, 2, \cdots, t).$$

For this we may use the likelihood ratio criteria or equivalently χ^2-test of goodness of fit, as in [1]. Let

$$\hat{p}_{ij}^{(h)} = \frac{\sum_{k=1}^{m_h} n_{ijk}^{(h)}}{\sum_{j,k} n_{ijk}^{(h)}} \qquad (h = 1, 2, \cdots, t; i, j = 0, 1, \cdots, s)$$

and

$$\hat{p}_{ij} = \frac{\sum_{k,h} n_{ijk}^{(h)}}{\sum_{k,h,j} n_{ijk}^{(h)}}, \qquad (i, j = 0, 1, \cdots, s)$$

where $n_{ijk}^{(h)}$ is the number of transitions from E_i to E_j in the kth sequence of hth set. From (ii), the required criterion

$$\chi^2 = \sum_{i=1}^{s} \sum_{h=1, j=0}^{t, s} \frac{(\hat{p}_{ij}^{(h)} - \hat{p}_{ij})^2 \cdot \sum_{k,j} n_{ijk}^{(h)}}{\hat{p}_{ij}}$$

has a limiting χ^2-distribution with $s^2(t - 1)$ degrees of freedom, for large m_h. It may be mentioned that in the present case, (40) will have a χ^2-distribution only if we have a large number of sequences in each set, unlike in the case when the total number of transitions n is nonrandom, when m_h might be equal to one.

My thanks are due to Dr. J. Gani for suggesting the field of inference in Markov chains as a topic for research and for criticizing an earlier draft of the paper. I am also indebted to Professor L. A. Goodman of the University of Chicago for pointing out an error in an earlier draft of the paper, and to the referee for his constructive criticisms of the paper.

APPENDIX

The following lemma is well known; it is given here for immediate reference in the paper.

LEMMA. *If a square matrix* $\mathbf{A} = (a_{ij})(i, j = 1, 2, \cdots, k)$ *is known to be of rank* $k - 1$, *and if for a vector* $(l_1, l_2, \cdots, l_k)$

$$l_1 a_{i1} + l_2 a_{i2} + \cdots + l_k a_{ik} = 0 \qquad (i = 1, 2, \cdots, k),$$

then the cofactors $A_{i1}, A_{i2}, \cdots, A_{ik}$ *are proportional to* $l_1, l_2, \cdots, l_k$ *respectively, for* $i = 1, 2, \cdots, k$.

PROOF. Since $\mathbf{A}$ is of rank $k - 1$,

$$|\mathbf{A}| = 0.$$

Hence for a particular value of j,

$$a_{i1}A_{j1} + a_{i2}A_{j2} + \cdots + a_{ik}A_{jk} = 0 \qquad (i = 1, 2, \cdots, k).$$

Thus $(A_{j1}, A_{j2}, \cdots, A_{jk})$ is a vector orthogonal to all row-vectors of $\mathbf{A}$. But $(l_1, l_2, \cdots, l_k)$ is also such a vector. Since the rank of $\mathbf{A}$ is $k - 1$, the two vectors must be proportional, and the lemma follows.

REFERENCES

[1] T. W. ANDERSON AND LEO A. GOODMAN, "Statistical inference about Markov chains," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 89–110.

[2] M. S. BARTLETT, *An Introduction to Stochastic Processes*, The University Press, Cambridge, 1956.

[3] B. R. BHAT, "Maximum likelihood estimation for positively regular Markov chain," *Sankhyā*, Vol. 22 (1960), pp. 339–344.

[4] HARALD CRAMÉR, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, 1946.

[5] HARALD CRAMÉR, *Random Variables and Probability Distributions*, Cambridge Tracts in Mathematics, Cambridge, 1937.

[6] WILLIAM FELLER, *An Introduction to Probability Theory and Its Applications*, Vol. I, 2nd ed., John Wiley and Sons, New York, 1957.

[7] MAURICE FRÉCHET, "Recherches modernes sur la théorie des probabilités," *Traité du Calcul des Probabilités* (ed. E. Borel), Vol. 1, No. 3, Paris, 1937–38.

[8] LEO A. GOODMAN, "Exact probabilities and asymptotic relationships for some statistics from mth order Markov chains," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 476–490.

[9] DAVID G. KENDALL, "Some problems in the theory of queues," *J. Roy. Stat. Soc.*, Ser. B, Vol. 13 (1951), pp. 151–185.

[10] V. I. ROMANOVSKY, *Discretnie Tsepi Markova* (Discrete Markov Chains), Moscow-Leningrad, 1949.

[11] P. WHITTLE, "Some distributions and moment formulae for the Markov chain," *J. Roy. Stat. Soc.*, Ser. B, Vol. 17 (1955), pp. 235–242.