ON THE ASYMPTOTIC DISTRIBUTION OF THE "PSI-SQUARED" GOODNESS OF FIT CRITERIA FOR MARKOV CHAINS AND MARKOV SEQUENCES[1]

By B. R. Bhat[2]

University of California, Berkeley

1. Introduction and summary. The use of a statistic of the algebraic form of Pearson's chi-squared as a measure of goodness of fit for frequencies from a fully specified mth order stationary Markov chain was first discussed and contrasted with the appropriate likelihood ratio criterion by Bartlett [2]. Since the distribution of the former statistic is not that of a tabular χ^2-variate, it and allied statistics, are sometimes described as "psi-squared" statistics. Patankar [14] derived the approximate asymptotic distribution (as the total number of transitions $\rightarrow \infty$) of

$$\psi_1^2 = \sum_i [(n_i - m_i)^2/m_i], \tag{1}$$

where the n_i are the marginal frequencies (1-tuples) in a large sequence from a simple stationary Markov chain and the m_i are their expected values in a new sequence of the same length. The proof is based on the fact that for a large sequence of observations the marginal frequencies are asymptotically multivariate normal and then (1) is distributed as a linear function of independent χ^2-variates. Since the latter can be approximated by a single Type III variate ([5], [15]) the approximate asymptotic distribution of (1) is completely specified by its first two moments.

Let $n_\mathfrak{u}$ be the frequency of the t-tuple $\mathfrak{u} = (u_1, u_2, \cdots, u_t)$ in a sequence of length $n + t - 1$ from an mth order stationary Markov chain; and let $m_\mathfrak{u}$ be its expected value in a new sequence of the same length. To test whether the chain has a specified transition probability matrix, in analogy with (1) one may construct the statistic

$$\psi_t^2 = \sum_\mathfrak{u} [(n_\mathfrak{u} - m_\mathfrak{u})^2/m_\mathfrak{u}] \tag{2}$$

and test the goodness of fit for $n_\mathfrak{u}$. In (2) the summation extends over those values of $\mathfrak{u}$ for which $m_\mathfrak{u}$ does not vanish.

Using methods different from those used here, Good [9] gave the asymptotic distribution of ψ_t^2 for the special case of a random sequence of digits, and showed that for an equiprobable random sequence (Markovity of order -1) having a prime number of categories, ψ_t^2 is asymptotically a linear combination of inde-

Received June 29, 1959; revised July 14, 1960.

[1] 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.

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

pendent χ^2-variates. This was generalized to the case of an arbitrary number of categories and to an arbitrary random sequence (Markovity of order 0) by Billingsley [4]. Good [11] conjectured that a similar result might be true for Markovity of any order. Following Good [10], Goodman [12] has shown that this conjecture is not true, and has proceeded to study a modification that is true. For further work in this direction and additional references see [13].

Since it is clear that the distribution of (2) does not have a simple form, we might assume that it follows the Type III form approximately. This approximation is suggested by the fact that (n_u) is asymptotically normal and hence the quadratic form (2) in (n_u) is distributed asymptotically as a linear function of χ^2-variates with one degree of freedom [5]. Since (2) is nonnegative, the coefficients of the corresponding linear function of χ^2-variates are also nonnegative. In the case when $m = 1$ or 0, the exact values of these coefficients are also known [4], [9]. The problem of approximating the distribution of a linear function of χ^2-variates has been discussed by Welch [15] and Box [5]. They observe that this Type III approximation is fairly good over a wide range of values of degrees of freedom of the different χ^2 and their coefficients, especially when these coefficients are positive. The advantage of this approximation is that it enables us to test the goodness of fit by referring to standard χ^2-tables. In Section 2 of this paper, we derive this approximate distribution of (2) by obtaining its first two moments for any m and $t \geqq m$.

Let $X_1, X_2, \cdots, X_{n+t-1}$ be a series of observations from a stationary linear Markov sequence (autoregressive) of first order;

$$X_i = \rho X_{i-1} + Y_i \quad (i = 2, 3, \cdots, n + t - 1), \tag{3}$$

where $|\rho| < 1$, and the Y_i are independent identically distributed continuous random variables with zero mean and range $(-\infty, +\infty)$. (Even though not in universal use, the term "Markov sequence" here refers to a Markov chain with continuous state space. We follow Bartlett [3] in using it.) Let these $n + t - 1$ observations be grouped into k class intervals and let n_u be the frequency of the t-tuple $(X_{u_1}, X_{u_2}, \cdots, X_{u_t})$ in this sequence, where $X_{u_1}, X_{u_2}, \cdots, X_{u_t}$ are $t(\geqq 1)$ consecutive observations belonging to the u_1th, u_2th, $\cdots$, u_tth class intervals respectively. For these frequencies, we derive the approximate distribution of the psi-squared test defined by (2), under some mild restrictions on the distribution of Y and for small class intervals, assuming ρ to be known. For the case $t = 1$, and the distribution of Y normal, Patankar [14] has obtained its distribution. We observe that, for $t = 1$ and Y arbitrarily distributed, the same distribution is obtained.

In Section 4, the distribution of the ψ_t^2 test of goodness of fit for frequencies of t-tuples $(t \geqq 2)$ in a series of observations, grouped into a finite number of class-intervals, from the stationary linear Markov sequence (autoregressive) of second order,

$$X_i = aX_{i-1} + bX_{i-2} + Y_i, \tag{4}$$

is derived, under similar restrictions on the distribution of Y. From this, the distribution of ψ_t^2 for stationary linear Markov sequences (autoregressive) of arbitrary order is deduced.

The distribution of (2) may also be used to calculate the power of the usual χ^2-test of goodness of fit for independent observations, when the alternative is serial dependence.

2. ψ^2-test for Markov chains. Let $X_1, X_2, \cdots, X_{n+t-1}$ be a sequence from a positively regular stationary Markov chain of order m with k possible states. Let a typical t-tuple $(t \geqq m)$ of states $(E_{u_1}, E_{u_2}, \cdots, E_{u_t})$ be denoted by $E_\mathfrak{u}$ and $n_\mathfrak{u}$, its observed frequency in this sequence. We shall derive the mean $m_\mathfrak{u}$, variance $\sigma_\mathfrak{u}^2$ and covariance $\sigma_{\mathfrak{u}\mathfrak{v}}$ of $n_\mathfrak{u}$ in a new sequence of the same length. For the case $m = 1, t = 1$, these formulae are derived by Patankar [14] and for the case $m = 1, t = 2$, by Gani [8].

Evidently $\mathfrak{u}$ can have k^t values, which may be viewed as k^t states of a modified simple Markov chain (cf., Bartlett, [3], p. 233). Let $\mathbf{P}_t$ be the transition probability matrix of these composite states $E_\mathfrak{u}$. It is completely specified by the transition probability matrix $\mathbf{P}_m$ of the mth order Markov chain. Thus, the probability that the t-tuple $\mathfrak{u}$ will be followed by the t-tuple $\mathfrak{v}$ in r steps is $p_t^{(r)}(\mathfrak{u}; \mathfrak{v})$, the element in the $\mathfrak{u}$th row and $\mathfrak{v}$th column of $\mathbf{P}_t^r$. Symbolically,

$$p_t^{(r)}(\mathfrak{u}; \mathfrak{v}) = \Pr\{u_1, u_2, \cdots, u_t \xrightarrow{r-t} v_1, v_2, \cdots, v_t\}.$$

Since the chain is of order m this probability is equal to

$$\begin{aligned}&\Pr(u_{t-m+1}, \cdots, u_t \xrightarrow{r-t} v_1, \cdots, v_m)\\ (5)\qquad &\qquad \cdot p_m^{(1)}(v_1, v_2, \cdots, v_m; v_2, v_3, \cdots, v_{m+1})\\ &\qquad\qquad \cdots p_m^{(1)}(v_{t-m}, \cdots, v_{m-1}; v_{t-m+1}, \cdots, v_m).\end{aligned}$$

The first factor of (5) is

$$p_m^{(r-t+m)}(u_{t-m+1}, \cdots, u_t; v_1, v_2, \cdots, v_m) = p_m^{(r-t+m)}[\mathfrak{u}_{t-m+1}(m); \mathfrak{v}_1(m)], \quad \text{(say)},$$

the element in the $\mathfrak{u}_{t-m+1}(m)$th row and the $\mathfrak{v}_1(m)$th column of $\mathbf{P}_m^{r-t+m}$. The remaining one-step transition probabilities are elements of $\mathbf{P}_m$, some of which may vanish.

Thus we see that, if the original Markov chain is positively regular and its transition probabilities are nonzero, the modified Markov chain will be positively regular. Otherwise some of the stationary probabilities, $P_\mathfrak{v}$, may vanish. From (5), $P_\mathfrak{v}$ are given by

$$(6)\qquad P_\mathfrak{v} = P_{\mathfrak{v}_1(t)} = P_{\mathfrak{v}_1(m)} \cdot T(\mathfrak{v}),$$

where $\mathfrak{v}_1(t) = [\mathfrak{v}_1(m), v_{m+1}, \cdots, v_t]$ and $T(\mathfrak{v})$ is the product of one-step transition probabilities in (5).

To derive $m_\mathfrak{u}$, $\sigma_\mathfrak{u}^2$ and $\sigma_{\mathfrak{u}\mathfrak{v}}$, we follow the procedure of Fréchet [7], Patankar

[14] and Gani [8] for the simple Markov chain. Let X_u^i be a random variable such that its value is 1 if the t-tuple starting with the ith observation is E_u $(i = 1, 2, \cdots, n)$ and 0 otherwise. Evidently

$$n_u = \textstyle\sum_i X_u^i, \qquad \sum_u n_u = n.$$

Since the chain is stationary,

$$\begin{aligned} E(X_u^i) &= P_u, \\ \text{Var}\,(X_u^i) &= P_u(1 - P_u), \\ \text{Cov}\,(X_u^i, X_u^j) &= \Pr\,(X_u^i = 1)\cdot\Pr\,(X_u^j = 1 \mid X_u^i = 1) - P_u^2, \\ &= P_u p_t^{(j-i)}(u; u) - P_u^2. \end{aligned}$$

Similarly

$$\text{Cov}\,(X_u^i, X_v^j) = P_u p_t^{(j-i)}(u; v) - P_u P_v \qquad (i < j).$$

Thus

$$E(n_u) = m_u = nP_u,$$

$$\text{(7)} \qquad \begin{aligned} \text{Var}(n_u) &= \sigma_u^2 = m_u - m_u^2 + 2m_u \sum_{s=1}^{n-1} \frac{n-s}{n} p_t^{(s)}(u; u), \\ &= m_u(1 - m_u + 2S_{uu}), \end{aligned}$$

and

$$\text{(8)} \qquad \sigma_{uv} = m_u S_{uv} + m_v S_{vu} - m_u m_v,$$

where

$$S_{uv} = \sum_{s=1}^{n-1} \frac{n-s}{n} p_t^{(s)}(u; v).$$

Now we can obtain the distribution of

$$\psi_t^2 = \textstyle\sum_u [(n_u - m_u)^2/m_u]$$

from (7) and (8), since for large values of n the joint distribution of n_u may be assumed multivariate normal [3]. Thus

$$\text{(9)} \qquad E(\psi_t^2) = \textstyle\sum_u \sigma_u^2/m_u$$

and

$$\text{(10)} \qquad \begin{aligned} \text{Var}(\psi_t^2) &= E(\psi_t^2)^2 - [E(\psi_t^2)]^2, \\ &= \sum_{u,v} \frac{\sigma_u^2\sigma_v^2 + 2\sigma_{uv}^2}{m_u m_v} - \sum_{u,v} \frac{\sigma_u^2\sigma_v^2}{m_u m_v}, \\ &= 2\sum_{u,v} \sigma_{uv}^2/m_u m_v, \end{aligned}$$

(cf., Anderson [1], p. 39).

It may be noted that, when $m_{\mathfrak{u}}$ vanishes, $\sigma_{\mathfrak{u}}^2$ and $\sigma_{\mathfrak{u}\mathfrak{v}}$ vanish. Thus (9) and (10) are valid even when some $m_{\mathfrak{u}}$ vanish, in which case as before the summation extends over those values of $\mathfrak{u}$ for which $m_{\mathfrak{u}}$ does not vanish.

Substituting the values of $\sigma_{\mathfrak{u}}^2$ from (7) we have

$$
\begin{aligned}
E(\psi_t^2) &= \textstyle\sum_{\mathfrak{u}}(1 - m_{\mathfrak{u}} + 2S_{\mathfrak{u}\mathfrak{v}}) \\
&= k_t - n + 2\textstyle\sum_{\mathfrak{u}} S_{\mathfrak{u}\mathfrak{v}} \\
&= k_t - n + 2\sum_{s=1}^{n-1} \frac{n-s}{n} \operatorname{tr}(\mathbf{P}_t^s),
\end{aligned}
\tag{11}
$$

where tr(**A**) is the trace of the matrix **A**, and k_t is the number of t-tuples for which $m_{\mathfrak{u}}$ does not vanish. (Cf., Goodman [13].) But from (5)

$$
\begin{aligned}
\operatorname{tr}(\mathbf{P}_t^s) &= \textstyle\sum_{\mathfrak{u}} p_t^{(s)}(\mathfrak{u};\mathfrak{u}) \\
&= \sum_{\mathfrak{u}_{t-m+1}(m)} \sum_{\mathfrak{u}_1(t-m)} p_m^{(s-t+m)}[\mathfrak{u}_{t-m+1}(m); \mathfrak{u}_1(m)] \cdot T(\mathfrak{u}) \\
&= \sum_{\mathfrak{u}_{t-m+1}(m)} p_m^{(s)}[\mathfrak{u}_{t-m+1}(m); \mathfrak{u}_{t-m+1}(m)] \\
&= \operatorname{tr}(\mathbf{P}_m^s).
\end{aligned}
\tag{12}
$$

Thus (11) can be evaluated if we know the trace of the powers of the transition probability matrix of the mth order Markov chain. Similarly, substituting (7) and (8) in (10), we have

$$
\begin{aligned}
\operatorname{Var}(\psi_t^2) &= 2\sum_{\mathfrak{u}} (1 - m_{\mathfrak{u}} + 2S_{\mathfrak{u}\mathfrak{u}})^2 \\
&\quad + 4\sum_{\mathfrak{u}\neq\mathfrak{v}} \left(m_{\mathfrak{u}}m_{\mathfrak{v}} - 2(m_{\mathfrak{u}}S_{\mathfrak{u}\mathfrak{v}} + m_{\mathfrak{v}}S_{\mathfrak{v}\mathfrak{u}}) + \frac{(m_{\mathfrak{u}}S_{\mathfrak{u}\mathfrak{v}} + m_{\mathfrak{v}}S_{\mathfrak{v}\mathfrak{u}})^2}{m_{\mathfrak{u}}m_{\mathfrak{v}}}\right) \\
&= 2\Bigg\{ k_t + n^2 - 2n + 4\sum_{\mathfrak{u}} S_{\mathfrak{u}\mathfrak{u}} \\
&\quad + \sum_{\mathfrak{u},\mathfrak{v}} \left[\frac{(m_{\mathfrak{u}}S_{\mathfrak{u}\mathfrak{v}} + m_{\mathfrak{v}}S_{\mathfrak{v}\mathfrak{u}})^2}{m_{\mathfrak{u}}m_{\mathfrak{v}}} - 2(m_{\mathfrak{u}}S_{\mathfrak{u}\mathfrak{v}} + m_{\mathfrak{v}}S_{\mathfrak{v}\mathfrak{u}})\right]\Bigg\}.
\end{aligned}
$$

Since

$$
\sum_{\mathfrak{v}} S_{\mathfrak{u}\mathfrak{v}} = \sum_{s=1}^{n-1} \frac{n-s}{n} = \frac{n-1}{2},
$$

$$
\operatorname{Var}(\psi_t^2) = 2\left\{k_t - n^2 + 4\sum_{\mathfrak{u}} S_{\mathfrak{u}\mathfrak{u}} + \sum_{\mathfrak{u},\mathfrak{v}} \frac{(m_{\mathfrak{u}} S_{\mathfrak{u}\mathfrak{v}} + m_{\mathfrak{u}} S_{\mathfrak{u}\mathfrak{v}})^2}{m_{\mathfrak{u}} m_{\mathfrak{v}}}\right\}. \tag{13}
$$

If the chain is reversible (for definition, see Burke and Rosenblatt [6]),

$$
m_{\mathfrak{u}}S_{\mathfrak{u}\mathfrak{v}} = m_{\mathfrak{v}}S_{\mathfrak{v}\mathfrak{u}}.
$$

Then (13) can be simplified to

$$(14)\quad \begin{aligned} &2\{k_t - n^2 + 4 \sum_{\mathfrak{u}} S_{\mathfrak{uu}} + 4 \sum_{\mathfrak{u},\mathfrak{v}} S_{\mathfrak{uv}} S_{\mathfrak{vu}}\} \\ &= 2\left\{k_t - n^2 + 4 \sum_s \frac{n-s}{n} \operatorname{tr}(\mathbf{P}_m^s) + 4 \sum_{s,t} \frac{n-s}{n} \cdot \frac{n-t}{n} \operatorname{tr}(\mathbf{P}_m^{s+t})\right\}. \end{aligned}$$

On the assumption that ψ_t^2 may be approximated by a Type III variate, we have derived its distribution. It may be noted that (14) is the variance of ψ_t^2 when the chain is reversible, while (11) is the mean, without any such restriction.

3. First order Markov sequence (autoregressive). Let $X_1, X_2, \cdots, X_{n+t-1}$ be a sequence of observations from (3). In this section we shall derive the approximate asymptotic distribution of ψ_t^2 test of goodness of fit for the frequencies of t-tuples, defined in Section 1.

Since the sequence is assumed stationary, the joint probability density function (p.d.f.) (assumed to exist and to be continuous) of $X_1, X_2, \cdots, X_{n+t-1}$ is

$$(15)\qquad p(x_1, x_2, \cdots, x_{n+t-1}) = p(x_1) \prod_{r=2}^{n+t-1} p_1(x_r \mid x_{r-1}),$$

where $p(x)$ is the stationary p.d.f., and $p_k(x \mid y)$ is the conditional density function of X, the $(r+k)$th observation, given Y, the rth. Further, the probability that X belongs to the ith class interval is

$$(16)\qquad P_i = \textstyle\int_i p(x)\, dx \qquad (i = 1, 2, \cdots, k),$$

where the integration is performed over the ith class. But, since $p(x)$ is continuous,

$$(17)\qquad P_i = p(\xi_i)\Delta\xi_i,$$

where ξ_i is some fixed point in the ith class, $\Delta\xi_i$ being its length. If the interval is of infinite length, $\Delta\xi_i$ may be chosen such that (17) is satisfied for a certain fixed point ξ_i in the ith class interval.

The probability that X_{t+r} belongs to the jth class, given that X_t belongs to ith, is

$$(18)\qquad p_{ij}^{(r)} = (1/P_i) \textstyle\int_i \int_j p(x_t) p_r(x_{t+r} \mid x_t)\, dx_{t+r}\, dx_t \; (r = 1, 2, \cdots).$$

For sufficiently small class intervals, (18) is approximately equal to

$$(19)\quad \begin{aligned} (1/P_i)p(\xi_i)p_r(\xi_j \mid \xi_i)\Delta\xi_i\Delta\xi_j &= p_r(\xi_j \mid \xi_i)\Delta\xi_j \\ &\cong \textstyle\int_j p_r(x_{t+r} \mid \xi_i)\, dx_{t+r}, \end{aligned}$$

for all values of i and j. Since ξ_i and ξ_j are fixed points, from (19) we observe that the "transition probabilities" $p_{ij}^{(n)}$ are independent of the values of x_{t+r} and x_t in the jth and ith class intervals respectively. Thus from Theorem 4, Corollary 3 of Burke and Rosenblatt [6], the observations retain their Markovian property, even though in general this property is lost by grouping. Hence, we may consider $X_1, X_2, \cdots, X_{n+t-1}$ as a sequence of observations from a simple Markov chain with rth transition probabilities given by (18). From the results of the previous section we may at once write down the mean and variance of ψ_t^2 as (11) and (13) respectively. We note that (11) and (14) can be calculated if we know $\mathrm{tr}(\mathbf{P}_1^r)$.

From (3)

$$X_{t+r} = \rho^r X_t + \rho^{r-1} Y_{t+1} + \cdots + \rho Y_{t+r-1} + Y_{t+r} .$$

The conditional distribution of X_{t+r} for $X_t = x_t$ is

$$\begin{aligned} \Pr[X_{t+r} \leqq x_{t+r} \mid X_t = x_t] & \\ &= \Pr[\rho^{r-1} Y_{t+1} + \cdots + Y_{t+r} \leqq x_{t+r} - \rho^r x_t] \\ &= F_r(x_{t+r} - \rho^r x_t), \end{aligned}$$

where F_r is the distribution function of $\rho^{r-1} Y_{t+1} + \cdots + Y_{t+r}$, which is independent of t and x_t. Because F_r is absolutely continuous, its derivative with respect to x_{t+r},

$$f_r(x_{t+r} - \rho^r x_t),$$

exists at any point $X_{t+r} = x_{t+r}$, and f_r does not depend on t and x_t. Therefore the probability that X_{t+r} lies in a small interval of length δx_t around the point $X_{t+r} = x_t$, under the condition that $X_t = x_t$ is given by

$$p_r(x_t \mid x_t) \delta x_t = f_r(x_t - \rho^r x_t) \delta x_t .$$

Hence,

$$\begin{aligned} \mathrm{tr}(\mathbf{P}_1^r) &= \textstyle\sum_i p_r(\xi_i \mid \xi_i) \Delta \xi_i \\ &= \textstyle\sum_i f_r(\xi_i - \rho^r \xi_i) \Delta \xi_i \qquad (20) \\ &\cong \textstyle\int_{-\infty}^{\infty} f_r(x - \rho^r x)\, dx; \end{aligned}$$

and because f_r is a density function, (20) equals $(1 - \rho^r)^{-1}$. (20) can be interpreted as the probability that $X_{t+r} = X_t$, for any given X_t, and is the continuous analogue of the trace of the rth power of transition probability matrix.

Because all the expected frequencies may be assumed non-zero, from (20), we have

$$E\psi_t^2 = k^t - n + 2 \sum_{r=1}^{n-r} \frac{n-r}{n} (1 - \rho^r)^{-1}. \qquad (21)$$

If

$$\textstyle\int_i \int_j p(x_t) p_r(x_{t+r} \mid x_t)\, dx_{t+r}\, dx_t = \int_j \int_i p(x_t) p_r(x_{t+r} \mid x_t)\, dx_{t+r}\, dx_t$$

(reversibility condition), from (14) we have

$$\mathrm{Var}(\psi_t^2) = 2\left\{k^t - n^2 + \sum_s \frac{n-s}{n} \frac{1}{1-\rho^s} + 4 \sum_{s,t} \frac{n-s}{n} \frac{n-t}{n} \frac{1}{1-\rho^{s+t}}\right\}. \qquad (22)$$

The reversibility condition is satisfied if the joint p.d.f. of X_t and X_{t+r} is symmetric in X_t and X_{t+r}, as in the normal case. It may be noted that (21) and (22) are the same as those derived by Patankar [14] for the special case when the X's are normal variates and the class intervals equal. Thus his results are true even when the X's follow a general class of continuous distributions.

4. Second order Markov sequence (autoregressive). As before, let $X_1, X_2, \cdots, X_{n+t-1}$ be a sequence of observations from (4). Their joint p.d.f. is

$$p(x_1, x_2, \cdots, x_{n+t-1}) = p(x_1, x_2) \prod_{r=3}^{n+t-1} p_1(x_r \mid x_{r-1}, x_{r-2}), \tag{23}$$

where $p(x, y)$ is the stationary p.d.f. of two consecutive observations in the sequence and $p_k(x \mid y, z)$ is the conditional p.d.f. of X, the $r + k$-th observation, given the rth observation y and $(r - 1)$th observation z. As in Section 3, the stationary probability that two consecutive observations belong to the ith and jth class intervals respectively, may be written as

$$P_{ij} = p(\xi_i, \xi_j)\Delta\xi_i\Delta\xi_j,$$

where ξ_i and ξ_j are some fixed points in the ith and jth class intervals. As before we assume the class intervals to be small. The probability that the 2-tuple (X_{t+r-1}, X_{t+r}) is (i', j'), given that (X_{t-2}, X_{t-1}) is (i, j), is $p_{ij,i'j'}^{(r+1)}$

$$\begin{aligned} &= P_{ij}^{-1} \textstyle\int_i \int_j \int_{i'} \int_{j'} p(x_{t-2}, x_{t-1})\bar{p}_{r+1}(x_{t+r-1}, x_{t+r} \mid x_{t-2}, x_{t-1}) \\ &\qquad dx_{t+r}\, dx_{t+r-1}\, dx_{t-1}\, dx_{t-2} \\ &\cong \bar{p}_{r+1}(\xi_{i'}, \xi_{j'} \mid \xi_i, \xi_j)\Delta\xi_{i'}\Delta\xi_{j'} \end{aligned} \tag{24}$$

where $\bar{p}_{r+1}$ is the conditional joint p.d.f. of (X_{t+r-1}, X_{t+r}) given (X_{t-2}, X_{t-1}).

Since (24) is independent of the values of $X_{t-2}, X_{t-1}, X_{t+r-1}, X_{t+r}$ in their respective class intervals we may consider the frequencies of t-tuples, n_u, as frequencies in a sequence of length $n + t - 1$ from a second order Markov chain and the mean and variance of ψ_t^2 can be obtained from (11) and (13). As in Section 3, we shall get the expression for $\mathrm{tr}(\mathbf{P}_2^r)$ in terms of a and b of equation (4).

Let the solution of the difference equation $u_r = au_{r-1} + bu_{r-2}$, for given u_1 and u_2, be $u_{k+2} = A_k u_2 + B_k u_1$. Then u_{k+2} for given u_1, u_2 and u_{k+2} is

$$au_{k+2} + b(A_{k-1}u_2 + B_{k-1}u_1).$$

The conditional joint p.d.f., f_{r+1} (say), of (X_{t+r-1}, X_{t+r}) given that $X_{t-1} = x_{t-1}$ and $X_{t-2} = x_{t-2}$, is the joint p.d.f. of

$$\eta_1 = X_{t+r-1} - A_r x_{t-1} - B_r x_{t-2}$$

and

$$\eta_2 = X_{t+r} - ax_{t+r-1} - bA_{r-1}x_{t-1} - bB_{r-1}x_{t-2},$$

for given (X_{t-1}, X_{t-2}). Hence as in Section 3,

$$\text{(25)} \qquad \begin{aligned} \operatorname{tr}(\mathbf{P}_2^{r+1}) &= \textstyle\sum_i \sum_j \bar{p}_{r+1}(\xi_i, \xi_j \mid \xi_i, \xi_j)\Delta\xi_i\Delta\xi_j \\ &\cong \textstyle\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{r+1}(\eta_1', \eta_2')\, dx_1\, dx_2 \end{aligned}$$

where η_1' and η_2' are η_1 and η_2 after substituting

$$X_{t+r-1} = x_{t-2} = x_1,$$

$$X_{t+r} = x_{t-1} = x_2.$$

We note that f_{r+1} does not depend on (x_{t-1}, x_{t-2}) except in the expression for η_1 and η_2. The Jacobian J of the transformation from (η_1', η_2') to (x_1, x_2) is a constant,

$$\text{(26)} \qquad |J| = (1 - bA_{r-1})(1 - B_r) - A_r(a + bB_{r-1}).$$

Thus (25) equals $|J|^{-1}$. Using (12) and substituting for tr $(\mathbf{P}_2^s)$ and $\operatorname{tr}(\mathbf{P}_2^{s+t})$ in (11) and (14) from (26), we get the mean and variance of ψ_t^2.

In general, for linear Markov sequences of arbitrary order m, the mean and variance of $\psi_t^2 (t \geqq m)$, can be obtained from the Jacobian J of the transformation from $\eta_1', \eta_2', \cdots, \eta_m'$ to $x_1, x_2, \cdots, x_m$ since it can be verified that

$$\operatorname{tr}(\mathbf{P}_m^{r+1}) = |J|^{-1},$$

where $\eta_1', \eta_2', \cdots, \eta_m'$ are defined in the same manner as η_1' and η_2' in (25), viz.,

$\eta_1 = X_{t+r-m+1}$ adjusted for given $x_{t-1}, x_{t-2}, \cdots, x_{t-m}$,

$\eta_2 = X_{t+r-m+2}$ adjusted for given $x_{t+r-m+1}, x_{t-1}, \cdots, x_{t-m}$,

. . . .,

$\eta_m = X_{t+r}$ adjusted for given $x_{t+r-1}, \cdots, x_{t+r-m+1}, x_{t-1}, \cdots, x_{t-m}$;

and $\eta_1', \eta_2', \cdots, \eta_m'$ are $\eta_1, \eta_2, \cdots, \eta_m$ with

$$X_{t+r-m+1} = x_{t-m} = x_1,$$

$$X_{t+r-m+2} = x_{t-m+1} = x_2,$$

$$X_{t+r} = x_{t-1} = x_m.$$

It is interesting to note that the expectation and variance of ψ_t^2 for first order Markov sequences (3) depend only on ρ and for second order sequences (4), only on a and b. They are independent of the distribution Y and also of the nature of grouping. The first property can be verified to be true for all linear Markov sequences, but does not appear to hold for Markov sequences in general.

In this paper we have assumed that the transition probabilities in Section 2, ρ, a and b in Sections 3 and 4, are completely specified. If they involve some unknown parameters the above formulae for the mean and variance of ψ_t^2 require modification.

5. Acknowledgment. The author is indebted to Messrs. J. Gani, P. A. P. Moran, and N. U. Prabhu for many useful suggestions. Comments and criticisms of Messrs. Leo A. Goodman and I. J. Good have weeded out the errors in an earlier draft of the paper and I am grateful to them.

REFERENCES

[1] T. W. ANDERSON, *Introduction to Multivariate Statistical Analysis*, John Wiley and Sons, New York, 1958.

[2] M. S. BARTLETT, "The frequency goodness of fit tests for probability chains," *Proc. Camb. Phil. Soc.*, Vol. 47 (1951), pp. 86–95.

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

[4] P. BILLINGSLEY, "Asymptotic distributions of two goodness of fit criteria," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 1123–9.

[5] G. E. P. BOX, "Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 290–302.

[6] C. J. BURKE AND M. ROSENBLATT, "A Markovian function of a Markov chain," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 1112–1122.

[7] MAURICE FRÉCHET, *Traité du Calcul des Probabilités et de ses Applications*, Vol. 2, Part III, Gauthier-Villars, Paris, 1952.

[8] J. GANI, "Some theorems and sufficiency conditions for the maximum likelihood estimator of an unknown parameter in a simple Markov chain," *Biometrika*, Vol. 42 (1955), pp. 342–359.

[9] I. J. GOOD, "The serial test for sampling numbers and other tests for randomness," *Proc. Camb. Phil. Soc.*, Vol. 49 (1953), pp. 276–284.

[10] I. J. GOOD, "The likelihood ratio test for Markoff chains," *Biometrika*, Vol. 42 (1955), pp. 531–533; Corrigenda, Vol. 44 (1957), p. 301.

[11] I. J. GOOD, "Review of P. Billingsley's 'Asymptotic distributions of two goodness of fit criteria'," *Math. Reviews*, Vol. 18 (1957), p. 607.

[12] LEO A. GOODMAN, "Asymptotic distributions of 'Psi-Squared' goodness of fit criteria for *m*th order Markov chains," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 1123–1133.

[13] LEO A. GOODMAN, "On some statistical tests for *m*th order Markov chains," *Ann. Math. Stat.*, Vol. 30 (1959), pp. 154–164.

[14] V. N. PATANKAR, "The goodness of fit of frequency distributions obtained from stochastic processes," *Biometrika*, Vol. 41 (1954), pp. 450–462.

[15] B. L. WELCH, "On linear combinations of several variances," *J. Amer. Stat. Assn.*, Vol. 51 (1956), pp. 132–48.