# TESTS AUXILIARY TO $\chi^2$ TESTS IN A MARKOV CHAIN[1]

By Ruth Z. Gold

*Columbia University*

**0. Summary.** In testing certain hypotheses about finite Markov chains, asymptotic $\chi^2$ tests have been derived which simulate $\chi^2$ tests in contingency tables. This paper considers two methods of augmenting such tests. In Section 3 a $\chi^2$ statistic proposed by Anderson and Goodman [2] for testing the hypothesis that the transition probabilities are homogeneous over time is decomposed into asymptotically independent single degree components by an extension of the decomposition given by Lancaster [11] for contingency tables. An interesting corollary of the proof given here is a nested hypothesis theorem.

In Section 4, an asymptotic method of judging simultaneously all linear combinations of multinomial probabilities in independent sequences of trials is derived. The extension to transition probabilities in a Markov chain is given in Section 5.

**1. Introduction.** The large sample theory of inference about Markov chains based on a single observation on a long chain or on repeated observations from a chain has provided tests about the order of the chain and about the homogeneity of the transition probabilities based on chi-square methods. Billingsley [4] has presented a survey of some of these results and the reader is referred to his paper for an extensive bibliography. Of particular interest here are the results of Anderson and Goodman [2], among which is an asymptotic $\chi^2$ test of homogeneity of the transition probabilities over time. The interest in the present paper is in methods which may help to pinpoint where (in time and/or over which states) the departures from homogeneity take place if heterogeneity is indicated.

The problem of detecting points of departure from homogeneity in Markov chains has been considered by Anderson [1] and by Goodman [7]. Both have proposed a sequence of $\chi^2$ tests which are not independent, Anderson's based on likelihood ratio statistics and Goodman's on the usual $\chi^2$ statistics in contingency tables. The two methods are asymptotically equivalent and Goodman [7] shows that for a particular example the numerical results are quite close. The decomposition of the $\chi^2$ statistic given in this paper provides a sequence of tests which are asymptotically independent. While this property does not in any way imply the superiority of this method, it does provide certain obvious conveniences based on the properties of independent $\chi^2$ statistics.

The test statistic proposed by Anderson and Goodman [2] for the hypothesis of homogeneity of the transition probabilities is computationally the same as

Received November 7, 1960; revised May 15, 1962.

[1] Part of a Ph.D. thesis presented to the Faculty of Political Science of Columbia University, 1960. Part of this research was done while under a grant from the Biostatistics Training Program of Columbia University.

the usual $\chi^2$ statistic in contingency tables and is computed from arrays which formally resemble contingency tables. The crucial difference is that the number which plays the role of the number of observations in the arrays arising from Markov chain sampling is a random variable rather than a fixed number as in the multinomial case. In the $r \times c$ contingency table, the decomposition of the $\chi^2$ statistic into $(r - 1)(c - 1)$ single components has been studied by Lancaster [11] and Irwin [10] and their results were further elucidated by Spencer [13]. They obtained two decompositions or partitions; the first being exact in the sense that the sum of the components is exactly the value of the $\chi^2$ statistic for the entire $r \times c$ array, while the second, which is the relevant one here, leads to an asymptotic equivalence. One may then test the individual components in order to detect departures from independence.

Lancaster's partition [11] is based on the $\chi^2$ statistics derived from $(r - 1)(c - 1)$ four-fold tables which are derived from the $r \times c$ contingency table. The decomposition given here is derived from the analogous four-fold tables in the Markov chain array. The proofs for contingency tables [11], [10] and [13] depend on the number of observations being fixed and the equivalence shown is between the sum of all the individual components and the usual $\chi^2$ statistic for the entire $r \times c$ array. In Section 3, it is shown that an asymptotic equivalence exists between the usual $\chi^2$ statistic for any $r' \times c'$ $(r' \leqq r, c' \leqq c)$ subset of the $r \times c$ Markov chain array and the sum of appropriate individual components. The asymptotic decomposition of the entire $r \times c$ array is thus a special case of the theorem proved. While this proof is given in terms of the $\chi^2$ statistic appropriate for testing the homogeneity of the transition probabilities, it is readily seen to apply to $\chi^2$ tests of other hypotheses about the chain which resemble tests of homogeneity in independent sequences of multinomial trials (see, for example, the test proposed in [2] p. 99 for the hypothesis that the chain is of a given order), as well as to $\chi^2$ tests of homogeneity in independent sequences of multinomial trials.

Goodman [8] and [9], in considering tests about the order of a chain, decomposes the likelihood ratio for certain contingency tables into asymptotically independent likelihood ratio statistics, implying an analogous decomposition for the corresponding asymptotically equivalent goodness of fit statistics. The proof given here is based directly on the usual $\chi^2$ statistics themselves.

An alternative method of augmenting $\chi^2$ tests of homogeneity in Markov chains is also presented here. The procedure derived enables one to construct simultaneous confidence intervals for all linear combinations of the transition probabilities in a finite Markov chain. The proof is based on Schwarz's inequality. The idea of using this inequality to prove Theorems 3 and 4 came from an analogous application of it extending the theorem of Scheffé [12] on simultaneous confidence intervals for all contrasts in the analysis of variance to all linear combinations of the unknown parameters. Unfortunately, the originator of this trick is unknown to this writer so that proper acknowledgment cannot be made.

In the analysis of variance, the remarkable procedures of both Scheffé [12] and Tukey [14] for simultaneously judging all contrasts among the parameters have the very desirable property that rejection of the homogeneity hypothesis by the appropriate $F$ or Studentized range test implies the existence of at least one relevant confidence interval which does not cover zero. An analogous result has not yet been obtained for the procedures derived in Sections 4 and 5. The difficulty seems to be that if the homogeneity test in the case of multinomial trials and of Markov chains is based on a $\chi^2$ statistic with say $\nu$ degrees of freedom, the relevant confidence intervals are proportional to square root of a cut-off point of the $\chi^2$ distribution with $\nu + 1$ degrees of freedom.

**2. The model.** Let the states be $i = 1, 2, \cdots, m$ and the times of observation $t = 0, 1, \cdots, T$. Denote by $p_{ij}(t)$ the probability of state $j$ at time $t$, given state $i$ at time $t - 1$. It is assumed for convenience that $p_{ij}(t) > 0$ $(i, j = 1, \cdots, m; t = 1, \cdots, T)$. A random sample of $N$ individuals is observed from the chain, an observation on an individual consisting of the sequence of states occupied by the individual at $t = 0, 1, \cdots, T$. Let $n_i(t)$ represent the number of individuals in state $i$ at time $t$. It is assumed that the sampling is such that, as $N \to \infty$, $n_i(0)/N \to_p \eta_i$ $(i = 1, \cdots, m)$, where $0 < \eta_i < 1$, $\sum_{i=1}^m \eta_i = 1$. Denote by $n_{ij}(t)$ the number of individuals in state $j$ at time $t$ who were in state $i$ at time $t - 1$. Let $\hat{p}_{ij}(t) = n_{ij}(t)/n_i(t-1)$. Thus [2] the $\hat{p}_{ij}(t)$ are the maximum likelihood estimators of the $p_{ij}(t)$.

**3. Decomposing the $\chi^2$ statistic in a finite Markov chain.** It has been shown [2] that an appropriate statistic for testing the hypothesis that the transition probabilities are homogeneous, i.e., $H_0 : p_{ij}(t) = p_{ij}$, $i, j = 1, \cdots, m$; $t = 1, \cdots, T$, is

$$\chi^2 = \sum_{i=1}^{m} \sum_{t=1}^{T} \sum_{j=1}^{m} n_i(t-1)[(\hat{p}_{ij}(t) - \hat{p}_{ij})^2/p_{ij}], \tag{3.1}$$

where $\hat{p}_{ij} = \sum_{t=1}^T n_{ij}(t)/\sum_{t=0}^{T-1} n_i(t)$. Anderson and Goodman proved [2] that under $H_0$, the statistic

$$\chi_i^2 = \sum_{t,j} n_i(t-1)[(\hat{p}_{ij}(t) - \hat{p}_{ij})^2/\hat{p}_{ij}] \tag{3.2}$$

is distributed in the limit as $\chi^2$ with $(m-1)(T-1)$ degrees of freedom and that, since the $\hat{p}_{ij}(t)$ and the $\hat{p}_{ij}$ for different $i$ and $t$ are asymptotically independent, $\chi^2 = \sum_i \chi_i^2$ has a limiting $\chi^2$ distribution with $m(m-1)(T-1)$ degrees of freedom.

Consider for $i$ fixed the $T \times m$ array

(3.3)

| $t\backslash j$ | 1 | 2 | $\cdots$ | $m$ | Total |
|---|---|---|---|---|---|
| 1 | $n_{i1}(1)$ | $n_{i2}(1)$ | | $n_{im}(1)$ | $n_i(0)$ |
| 2 | $n_{i1}(2)$ | $n_{i2}(2)$ | | $n_{im}(2)$ | $n_i(1)$ |
| $\vdots$ | | | | | |
| $T$ | $n_{i1}(T)$ | $n_{i2}(T)$ | | $n_{im}(T)$ | $n_i(T-1)$ |
| Total | $\sum_{t=1}^T n_{i1}(t)$ | $\sum_{t=1}^T n_{i2}(t)$ | | $\sum_{t=1}^T n_{im}(t)$ | $\sum_{t=1}^T n_i(t-1)$. |

The statistic given by (3.2) is the one that would be used if one were performing the usual test for homogeneity on the data of the array (3.3) in spite of the fact that not only the marginal totals, the $n_i(t-1)$, but also the number $\sum_{t=1}^{T} n_i(t-1)$, analogous to the total number of observations in a contingency table, is a random variable.

Consider now the $(m-1)(T-1)$ four-fold tables derived from (3.3) as follows:

$$\begin{array}{cc|} n_{i1}(1) & n_{i2}(1) \\ n_{i1}(2) & n_{i2}(2) \\ \hline \end{array} \qquad \begin{array}{cc|} n_{i1}(1) + n_{i2}(1) & n_{i3}(1) \\ n_{i1}(2) + n_{i2}(2) & n_{i3}(2) \\ \hline \end{array} \qquad \cdots \qquad \begin{array}{cc|} \sum_{\alpha=1}^{m-1} n_{i\alpha}(1) & n_{im}(1) \\ \sum_{\alpha=1}^{m-1} n_{i\alpha}(2) & n_{im}(2) \\ \hline \end{array}$$

$$\begin{array}{cc|} n_{i1}(1) + n_{i1}(2) & n_{i2}(1) + n_{i2}(2) \\ n_{i1}(3) & n_{i2}(3) \\ \hline \end{array} \qquad \qquad \vdots$$

$$\vdots$$

$$\begin{array}{cc|} \sum_{\beta=1}^{T-1} n_{i1}(\beta) & \sum_{\beta=1}^{T-1} n_{i2}(\beta) \\ n_{i1}(T) & n_{i2}(T) \\ \hline \end{array} \qquad \qquad \begin{array}{cc|} \sum_{\beta=1}^{T-1} \sum_{\alpha=1}^{m-1} n_{i\alpha}(\beta) & \sum_{\beta=1}^{T-1} n_{im}(\beta) \\ \sum_{\alpha=1}^{m-1} n_{i\alpha}(T) & n_{im}(T) \\ \hline \end{array}$$

The general four-fold table is

$$\begin{array}{lcc|c} & N_{ij}(t) & \sum_{\beta=1}^{t} n_{i,j+1}(\beta) & N_{i,j+1}(t) \\ & \sum_{\alpha=1}^{j} n_{i\alpha}(t+1) & n_{i,j+1}(t+1) & \sum_{\alpha=1}^{j+1} n_{i\alpha}(t+1) \\ \hline \text{Total} & N_{ij}(t+1) & \sum_{\beta=1}^{t+1} n_{i,j+1}(\beta) & N_{i,j+1}(t+1) \end{array} \tag{3.4}$$

$$j = 1, \cdots, m; \quad t = 1, \cdots, T-1,$$

where $N_{ih}(\tau) = \sum_{\beta=1}^{\tau} \sum_{\alpha=1}^{h} n_{i\alpha}(\beta)$. These are analogous to the tables considered by Lancaster [11] for contingency tables.

Let

$$\chi_{i:jt} = \frac{[N_{i,j+1}(t+1)/N]^{\frac{1}{2}} N^{\frac{1}{2}} \left[ N_{ij}(t) n_{i,j+1}(t+1)/N^2 - \sum_{\beta=1}^{t} n_{i,j+1}(\beta) \cdot \sum_{\alpha=1}^{j} n_{i\alpha}(t+1)/N^2 \right]}{\left[ N_{ij}(t+1) \sum_{\beta=1}^{t+1} n_{i,j+1}(\beta) N_{i,j+1}(t) \sum_{\alpha=1}^{j+1} n_{i\alpha}(t+1)/N^4 \right]^{\frac{1}{2}}}, \tag{3.5}$$

$$j = 1, \cdots, m-1; \; t = 1, \cdots, T-1.$$

In Theorem 1 it is shown that under $H_0$, the statistics $\chi_{i:jt}$ associated with the $(m-1)(T-1)$ four-fold tables (3.4) are asymptotically distributed as independent standard normal variates. The asymptotic equivalence between the "usual" $\chi^2$ for any subset of the array (3.3) and the sum of an appropriate subset of squares of the variates $\chi_{i:jt}$ is proved in the second theorem.

THEOREM 1. *Under the hypothesis* $H_0 : p_{ij}(t) = p_{ij}$ $(i, j = 1, \cdots, m;\ t = 1, \cdots, T)$, *the statistics* $\chi_{i:jt}$ $(i, j = 1, \cdots, m;\ t = 1, \cdots, T)$ *defined by* (3.5) *are asymptotically distributed as independent normal variables each with zero mean and unit variance.*

PROOF. Under $H_0$, we have from [2] the asymptotic joint normality of the $n_{ij}(t)$ and that

$$\mathcal{E} n_{ij}(t)/N = m_i(t-1)p_{ij}, \tag{3.6}$$

where

$$m_i(t-1) = \mathcal{E} n_i(t-1)/N = \sum_{k=1}^{m} \eta_k p_{ki}^{[t-1]}, \tag{3.7}$$

$p_{ki}^{[t-1]}$ denoting the $(ki)$th element of the matrix $\|p_{ij}\|$ $(i, j = 1, \cdots, m)$ of transition probabilities raised to the power $t-1$. Also, from [1] we have that under $H_0$,

$$\operatorname{Var}(n_{ij}(t)/N) = p_{ij}(1-p_{ij})m_i(t-1)/N, \tag{3.8}$$

and

$$\operatorname{Cov}\{n_{ij}(t)/N, n_{ik}(t)/N\} = -p_{ij}p_{ik}m_i(t-1)/N. \tag{3.9}$$

Let

$$y_{i:jt} = \frac{N^{\frac{1}{2}}\left[N_{ij}(t)n_{i,j+1}(t+1)/N^2 - \sum_{\beta=1}^{t} n_{i,j+1}(\beta) \sum_{\alpha=1}^{j} n_{i\alpha}(t+1)/N^2\right]}{\left[\sum_{\alpha=1}^{j} p_{i\alpha} \sum_{\alpha=1}^{j+1} p_{i\alpha}p_{i,j+1} \sum_{\beta=0}^{t-1} m_i(\beta) \sum_{\beta=0}^{t} m_i(\beta)m_i(t)\right]^{\frac{1}{2}}}, \tag{3.10}$$

$$j = 1, \cdots, m-1;\ t = 1, \cdots, T-1,$$

and let

$$\begin{aligned} x_{i:jt} = (c_{i:jt}N)^{\frac{1}{2}}\Bigg\{&p_{i,j+1}\left[N_{ij}(t)/N\sum_{\beta=0}^{t-1} m_i(\beta) - \sum_{\alpha=1}^{j} n_{i\alpha}(t+1)/N\, m_i(t)\right] \\ &+ \sum_{\alpha=1}^{j} p_{i\alpha}\left[n_{i,j+1}(t+1)/N\, m_i(t) - \sum_{\beta=1}^{t} n_{i,j+1}(\beta)/N \sum_{\beta=0}^{t-1} m_i(\beta)\right]\Bigg\}, \end{aligned} \tag{3.11}$$

$$j = 1, \cdots, m-1;\ t = 1, \cdots, T-1,$$

where

$$c_{i:jt} = m_i(t) \sum_{\beta=0}^{t-1} m_i(\beta) \Big/ \sum_{\beta=0}^{t} m_i(\beta) \sum_{\alpha=1}^{j} p_{i\alpha} \sum_{\alpha=1}^{j+1} p_{i\alpha}p_{i,j+1}. \tag{3.12}$$

We shall show that the $x_{i:jt}$ are asymptotically independently normally distributed with 0 means and unit variances, that the $y_{i:jt}$ are asymptotically equivalent to the $x_{i:jt}$ and that the $\chi_{i:jt}$ are in turn asymptotically equivalent to the $y_{i:jt}$. It will then follow that the limiting distribution of the $\chi_{i:jt}$ is that of independent standard normal variates and that their squares are asymptotically independently distributed as single degree $\chi^2$'s.

Since the $x_{i:jt}$ are linear functions of asymptotically jointly normal variables, the $n_{ij}(t)$, the limiting joint distribution of the $x_{i:jt}$ is normal. To calculate the moments of the limiting distribution, we have from (3.6)

$$\mathcal{E}(x_{i:jt}) = 0 \quad j = 1, \cdots, m-1; t = 1, \cdots, T-1. \tag{3.13}$$

Because of the asymptotic independence for different $\beta$ of the $n_{i\alpha}(\beta)/N$[1],

$$\begin{aligned}\sigma^2(x_{i:jt}) = c_{i:jt}/N \Bigg\{\sigma^2\Bigg[p_{i,j+1} N_{ij}(t) \Big/ \sum_{\beta=0}^{t-1} m_i(\beta) \\ - \sum_{\alpha=1}^{j} p_{i\alpha} \sum_{\beta=1}^{t} n_{i,j+1}(\beta) \Big/ \sum_{\beta=0}^{t-1} m_i(\beta)\Bigg] \\ + \sigma^2\Bigg[\sum_{\alpha=1}^{j} p_{i\alpha} n_{i,j+1}(t+1)/m_i(t) - p_{i,j+1} \sum_{\alpha=1}^{j} n_{i\alpha}(t+1)/m_i(t)\Bigg]\Bigg\}, \\ j = 1, \cdots, m-1; t = 1, \cdots, T-1,\end{aligned} \tag{3.14}$$

where the symbol $\sigma^2(z)$ denotes the variance of the limiting distribution of $z$. Utilizing (3.8) and (3.9), we find $\sigma^2(x_{i:jt}) = 1$ $(j = 1, \cdots, m-1; t = 1, \cdots, T-1)$.

We next calculate Cov $(x_{i:jt}, x_{i:hk})$. Considering separately the cases $1 - j < h, t = k$; $2 - j = h, t < k$; $3 - j < h, t < k$ and $4 - j < h, t > k$ and utilizing (3.8) and (3.9) along with the asymptotic independence for different $\beta$ of the $n_{i\alpha}(\beta)$, it can be shown that in all cases Cov $(x_{i:jt}, x_{i:hk}) = 0$ $(j \neq h$, and/or $t \neq k)$. The variables $x_{i:jt}$ $(j = 1, \cdots, m-1; t = 1, \cdots, T-1)$ are thus asymptotically distributed as independent standard normal variables.

We show now that

$$p \lim (y_{i:jt} - x_{i:jt}) = 0. \tag{3.15}$$

Let

$$y'_{i:jt} = \frac{N^{\frac{1}{2}}\left[N_{ij}(t) n_{i,j+1}(t+1)/N^2 - \sum_{\beta=1}^{t} n_{i,j+1}(\beta) \sum_{\alpha=1}^{j} n_{i\alpha}(t+1)/N^2\right]}{m_i(t) \sum_{\beta=0}^{t-1} m_i(\beta)}, \tag{3.16}$$

$$j = 1, \cdots, m-1; t = 1, \cdots, T-1,$$

and let

$$
(3.17)\quad \begin{aligned} x'_{i:jt} = N^{\frac{1}{2}} \Big\{ & p_{i,j+1} \Big[ N_{ij}(t)/N \sum_{\beta=0}^{t-1} m_i(\beta) - \sum_{\alpha=1}^{j} n_{i\alpha}(t+1)/N\, m_i(t) \Big] \\ & + \sum_{\alpha=1}^{j} p_{i\alpha} \Big[ n_{i,j+1}(t+1)/N\, m_i(t) - \sum_{\beta=1}^{t} n_{i,j+1}(\beta)/N \sum_{\beta=0}^{t-1} m_i(\beta) \Big] \Big\}. \end{aligned}
$$

Then

$$
(3.18)\quad y_{i:jt} = (c_{i:jt})^{\frac{1}{2}} y'_{i:jt},
$$

and

$$
(3.19)\quad x_{i:jt} = (c_{i:jt})^{\frac{1}{2}} x'_{i:jt}.
$$

Since $c_{i:jt}$ is a constant,

$$
(3.20)\quad p \lim (y_{i:jt} - x_{i:jt}) = (c_{i:jt})^{\frac{1}{2}} p \lim (y'_{i:jt} - x'_{i:jt}).
$$

Both $y'_{i:jt}$ and $x'_{i:jt}$ are continuous functions of the $n_{i\alpha}(\beta)/N$ and have continuous partial derivatives with respect to these variables so that a Taylor's expansion in terms of differences $n_{i\alpha}(\beta)/N - m_i(\beta - 1)p_{i\alpha}$ ($\alpha = 1, \cdots, j+1$; $\beta = 1, \cdots, t+1$) may be obtained. It is easily verified that both $y'_{i:jt}$ and $x'_{i:jt}$ vanish at the point $n_{i\alpha}(\beta)/N = m_i(\beta - 1)p_{i\alpha}$ ($\alpha = 1, \cdots, j+1$; $\beta = 1, \cdots, t+1$) and that all corresponding first-order partial derivatives evaluated at this point coincide.

All second-order partial derivatives of $x'_{i:jt}$ vanish while the order of magnitude of the second-order terms of the expansion of $y'_{i:jt}$ is that of

$$
(3.21)\quad \begin{aligned} & N^{\frac{1}{2}}[n_{i\alpha}(\beta)/N - m_i(\beta - 1)p_{i\alpha}][n_{i\gamma}(\delta)/N - m_i(\delta - 1)p_{i\delta}], \\ & \qquad \alpha, \gamma = 1, \cdots, j+1; \beta, \delta = 1, \cdots, t+1. \end{aligned}
$$

Because of the asymptotic joint normality [2] of terms such as $N^{\frac{1}{2}}[n_{i\alpha}(\beta)/N - m_i(\beta - 1)p_{i\alpha}]$, these terms are bounded in probability while $p \lim [n_{i\gamma}(\delta)/N - m_i(\delta - 1)p_{i\gamma}] = 0$. Thus the second-order terms of $y'_{i:jt}$ (of which there are a finite number) tend to zero in probability. It follows that

$$
(3.22)\quad p \lim (y'_{i:jt} - x'_{i:jt}) = 0,
$$

and because of (3.20) this establishes (3.15). It follows, ([6], p. 254), that the limiting distribution of the $y_{i:jt}$ is that of the $x_{i:jt}$. Thus, ([5], Theorem 2) the limiting distribution of $y^2_{i:jt}$ is that of $\chi^2$ with one degree of freedom and because of the asymptotic independence of the $y_{i:jt}$, the limiting distribution of $\sum_{t=1}^{T-1} \sum_{j=1}^{m-1} y^2_{i:jt}$ is that of $\chi^2$ with $(m-1)(T-1)$ degrees of freedom.

We shall show now that

$$
(3.23)\quad p \lim (\chi_{i:jt} - y_{i:jt}) = 0.
$$

Write

$$
(3.24)\quad \chi_{i:jt} = y_{i:jt} z_{i:jt},
$$

where

$$z_{i:jt} = \left[ \frac{N_{i,j+1}(t+1)/N \sum_{\beta=0}^{t} m_i(\beta) \sum_{\alpha=1}^{j} p_{i\alpha}}{\sum_{\beta=0}^{t} m_i(\beta) \sum_{\alpha=1}^{j+1} p_{i\alpha}} \frac{p_{i,j+1} \sum_{\beta=0}^{t} m_i(\beta)}{N_{ij}(t+1)/N \sum_{\beta=1}^{t+1} n_{i,j+1}(\beta)/N} \right. \tag{3.25}$$

$$\left. \cdot \frac{\sum_{\beta=0}^{t-1} m_i(\beta) \sum_{\alpha=1}^{j+1} p_{i\alpha}}{N_{i,j+1}(t)/N} \frac{m_i(t) \sum_{\alpha=1}^{j+1} p_{i\alpha}}{\sum_{\alpha=1}^{j+1} n_{i\alpha}(t+1)/N} \right]^{\frac{1}{2}}.$$

From the fact that $p \lim [n_{ij}(t)/N - m_i(t-1)p_{ij}] = 0$ and ([6], p. 255), it follows that

$$p \lim [z_{i:jt} - 1] = 0. \tag{3.26}$$

Therefore, the limiting distribution of the $\chi_{i:jt}$ is that of the $y_{i:jt}$, namely, that of independent standard normal variables from which it follows that $\chi^2_{i:jt}$ is asymptotically distributed as $\chi^2$ with one degree of freedom and that $\sum_{t=1}^{T-1} \sum_{j=1}^{m-1} \chi^2_{i:jt}$ is distributed in the limit as $\chi^2$ with $(m-1)(T-1)$ degrees of freedom.

Consider now the following subset of (3.3):

$$\begin{matrix} n_{i1}(1) & n_{i2}(1) & \cdots & n_{i,j+1}(1) \\ n_{i1}(2) & n_{i2}(2) & \cdots & n_{i,j+1}(2) \\ \vdots & & & \\ n_{i1}(t+1) & n_{i2}(t+1) & \cdots & n_{i,j+1}(t+1) \end{matrix} \quad , j \leqq m-1; t \leqq T-1. \tag{3.27}$$

Compute the usual $\chi^2$ associated with such an array, i.e.,

$$\chi^2 [i:j+1, t+1] = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}, \tag{3.28}$$

where the "expected" corresponding to $n_{iq}(p)$ is

$$\sum_{\alpha=1}^{q} n_{i\alpha}(p) \sum_{\beta=1}^{p} n_{iq}(\beta)/N_{i,j+1}(t+1) \qquad p = 1, \cdots, t+1; q = 1, \cdots, j+1. \tag{3.29}$$

THEOREM 2. *Under* $H_0 : p_{ij}(t) = p_{ij}$ $(i, j = 1, \cdots, m;\ t = 1, \cdots, T)$, $p \lim (\chi^2[i:j+1, t+1] - \sum_{k=1}^{t} \sum_{h=1}^{j} y^2_{i:hk}) = 0$, *where* $\chi^2[i:j+1, t+1]$ *is defined by* (3.28) *and* (3.29) *and* $y_{i:hk}$ *by* (3.10).

PROOF: By induction. For $j = t = 1$,

$$\chi^2[i:2, 2] = \chi^2_{i:1,1} \tag{3.30}$$

where $\chi^2_{i:1,1}$ is defined by (3.5). But by (3.24) and (3.25)

$$\chi^2_{i:1,1} = y^2_{i:1,1} z^2_{i:1,1}, \tag{3.31}$$

and since $p \lim (z^2_{i:1,1} - 1) = 0$ by (3.26) and ([6], p. 255), the theorem is true for $j = t = 1$.

Now assume that

$$p \lim \left( \chi^2 [i:j, t] - \sum_{k=1}^{t-1} \sum_{h=1}^{j-1} y^2_{i:hk} \right) = 0. \tag{3.32}$$

We show next that then

$$p \lim \left( \chi^2 [i:j+1, t] - \sum_{k=1}^{t-1} \sum_{h=1}^{j} y^2_{i:hk} \right) = 0, \tag{3.33}$$

and that

$$p \lim \left( \chi^2 i: j, t+1] - \sum_{k=1}^{t} \sum_{h=1}^{j-1} y^2_{i:hk} \right) = 0. \tag{3.34}$$

To prove (3.33), we write

$$\chi^2[i:j+1, t] = a + b + c + d, \tag{3.35}$$

where

$$a = N_{i,j+1}(t) \sum_{\beta=1}^{t} \sum_{\alpha=1}^{j} \frac{\left[ n_{i\alpha}(\beta) - \sum_{p=1}^{t} n_{i\alpha}(p) \sum_{q=1}^{j} n_{iq}(\beta)/N_{ij}(t) \right]^2}{\sum_{p=1}^{t} n_{i\alpha}(p) \sum_{q=1}^{j+1} n_{iq}(\beta)}, \tag{3.36}$$

$$b = 2N_{i,j+1}(t) \sum_{\beta=1}^{t} \sum_{\alpha=1}^{j} \frac{\left[ n_{i\alpha}(\beta) - \sum_{p=1}^{t} n_{i\alpha}(p) \sum_{q=1}^{j} n_{iq}(\beta)/N_{ij}(t) \right] \cdot \left[ \sum_{q=1}^{j} n_{iq}(\beta)/N_{ij}(t) - \sum_{q=1}^{j+1} n_{iq}(\beta)/N_{i,j+1}(t) \right]}{\sum_{q=1}^{j+1} n_{iq}(\beta)}, \tag{3.37}$$

$$c = N_{i,j+1}(t) \sum_{\beta=1}^{t} \sum_{\alpha=1}^{j} \cdot \frac{\left[ \sum_{p=1}^{t} n_{i\alpha}(p) \sum_{q=1}^{j} n_{iq}(\beta)/N_{ij}(t) - \sum_{p=1}^{t} n_{i\alpha}(p) \sum_{q=1}^{j+1} n_{iq}(\beta)/N_{i,j+1}(t) \right]^2}{\sum_{p=1}^{t} n_{i\alpha}(p) \sum_{q=1}^{j+1} n_{iq}(\beta)} \tag{3.38}$$

and

$$d = \sum_{\beta=1}^{t} \frac{\left[ n_{i,j+1}(\beta) - \sum_{\alpha=1}^{j+1} n_{i\alpha}(\beta) \sum_{p=1}^{t} n_{i,j+1}(p)/N_{i,j+1}(t) \right]^2}{\sum_{\alpha=1}^{j+1} n_{i\alpha}(\beta) \sum_{p=1}^{t} n_{i,j+1}(p)/N_{i,j+1}(t)}. \tag{3.39}$$

We shall show that

$$p \lim (a - \chi^2[i:j, t]) = 0, \tag{3.40}$$

$$b = 0 \tag{3.41}$$

and

$$p \lim \left( c + d - \sum_{k=1}^{t-1} y_{i:jk}^2 \right) = 0. \tag{3.42}$$

We may write

$$a = N_{ij}(t) \sum_{\beta=1}^{t} \sum_{\alpha=1}^{j} \frac{\left[ n_{i\alpha}(\beta) - \sum_{p=1}^{t} n_{i\alpha}(p) \sum_{q=1}^{j} n_{iq}(\beta)/N_{ij}(t) \right]^2}{\sum_{p=1}^{t} n_{i\alpha}(p) \sum_{q=1}^{j} n_{iq}(\beta)} \epsilon_\beta , \tag{3.43}$$

where

$$\epsilon_\beta = \frac{N_{i,j+1}(t)/N}{\sum_{q=1}^{j+1} n_{iq}(\beta)/N} \cdot \frac{\sum_{q=1}^{j} n_{iq}(\beta)/N}{N_{ij}(t)/N} \tag{3.44}$$

It is easily verified that

$$p \lim (\epsilon_\beta - 1) = 0 \qquad \beta = 1, \cdots, t. \tag{3.45}$$

We may therefore write

$$a = \chi^2[i:j, t] + R_{i:jt} , \tag{3.46}$$

where

$$R_{i:jt} = N_{ij}(t) \sum_{\beta=1}^{t} \sum_{\alpha=1}^{j} \frac{\left[ n_{i\alpha}(\beta) - \sum_{p=1}^{t} n_{i\alpha}(p) \sum_{q=1}^{j} n_{iq}(\beta)/N_{ij}(t) \right]^2}{\sum_{p=1}^{t} n_{i\alpha}(p) \sum_{q=1}^{j} n_{iq}(\beta)} \delta_\beta , \tag{3.47}$$

and

$$\delta_\beta = \epsilon_\beta - 1. \tag{3.48}$$

Because of (3.45),

$$p \lim \delta_\beta = 0. \tag{3.49}$$

Since each term in $\chi^2[i:j, t]$ is non-negative and their sum is bounded in probability by virtue of (3.32) and the fact that $\sum_{k=1}^{t-1} \sum_{h=1}^{j-1} y_{i:hk}^2$ tends to a limiting distribution, each of the terms must be bounded in probability. Thus

$$R_{i:jt} = \sum_{\beta=1}^{t} \sum_{\alpha=1}^{j} O_p(1) o_p(1) = \sum_{\beta=1}^{t} \sum_{\alpha=1}^{j} o_p(1) = o_p(1). \tag{3.50}$$

From (3.50) and (3.46), we have (3.40).

Now, since

$$\sum_{\alpha=1}^{j}\left[n_{i\alpha}(\beta) - \sum_{p=1}^{t} n_{i\alpha}(p)\sum_{q=1}^{j} n_{iq}(\beta)/N_{ij}(t)\right] = \sum_{\alpha=1}^{j} n_{i\alpha}(\beta) - N_{ij}(t)\sum_{q=1}^{j} n_{iq}(\beta)/N_{ij}(t) = 0, \tag{3.51}$$

it follows that $b = 0$.

Adding $c$ to $d$ and simplifying, we have

$$c + d = \sum_{\beta=1}^{t} G_\beta \gamma_\beta , \tag{3.52}$$

where

$$G_\beta = \frac{\left[N_{ij}(t) n_{i,j+1}(\beta) - \sum_{k=1}^{t} n_{i,j+1}(k)\sum_{q=1}^{j} n_{iq}(\beta)\right]^2}{N^3 \sum_{\tau=0}^{t-1} m_i(\tau)\sum_{\alpha=1}^{j} p_{i\alpha}\sum_{\alpha=1}^{j+1} p_{i\alpha}\, p_{i,j+1}\, m_i(\beta - 1)}, \tag{3.53}$$

and

$$\gamma_\beta = \frac{\sum_{\alpha=1}^{j} p_{i\alpha}\sum_{\tau=0}^{t-1} m_i(\tau)}{N_{ij}(t)/N}\;\frac{\sum_{\tau=0}^{t-1} m_i(\tau) p_{i,j+1}}{\sum_{p=1}^{t} n_{i,j+1}(p)/N}\;\frac{m_i(\beta - 1)\sum_{\alpha=1}^{j+1} p_{i\alpha}}{\sum_{q=1}^{j+1} n_{iq}(\beta)/N}. \tag{3.54}$$

It is easily verified that

$$p \lim\, (\gamma_\beta - 1) = 0. \tag{3.55}$$

Setting

$$\omega_\beta = \gamma_\beta - 1, \tag{3.56}$$

we may write

$$c + d = \sum_{\beta=1}^{t} G_\beta(1 + \omega_\beta). \tag{3.57}$$

It will be shown below that $\sum_{\beta=1}^{t} G_\beta$ tends to a limiting distribution, from which it will follow that

$$p \lim \sum_{\beta=1}^{t} G_\beta \omega_\beta = 0. \tag{3.58}$$

We shall then have that

$$p \lim \left(c + d - \sum_{\beta=1}^{t} G_\beta\right) = 0. \tag{3.59}$$

We have only to show that

$$p \lim \left(\sum_{\beta=1}^{t} G_\beta - \sum_{k=1}^{t-1} y_{i:jk}^2\right) = 0, \tag{3.60}$$

and (3.42) will be established.

Write

$$\sum_{k=1}^{t-1} y_{i:jk}^2 = F'C, \tag{3.61}$$

where

$$C = 1/\sum_{\alpha=1}^{j} p_{i\alpha} \sum_{\alpha=1}^{j+1} p_{i\alpha}p_{i,j+1}\,, \tag{3.62}$$

and

$$F' = N \sum_{k=1}^{t-1} \cdot \frac{\left[N_{ij}(k)n_{i,j+1}(k+1)/N^2 - \sum_{\beta=1}^{k} n_{i,j+1}(\beta) \sum_{\alpha=1}^{j} n_{i\alpha}(k+1)/N^2\right]^2}{m_i(k) \sum_{\tau=0}^{k-1} m_i(\tau) \sum_{\tau=0}^{k} m_i(\tau)}. \tag{3.63}$$

Write also

$$\sum_{\beta=1}^{t} G_\beta = G'C, \tag{3.64}$$

where $C$ is defined by (3.62) and

$$G' = \frac{N}{\left[\sum_{\tau=0}^{t-1} m_i(\tau)\right]^2} \sum_{\beta=1}^{t} \cdot \frac{\left[N_{ij}(t)n_{i,j+1}(\beta)/N^2 - \sum_{k=1}^{t} n_{i,j+1}(k) \sum_{q=1}^{j} n_{iq}(\beta)/N^2\right]^2}{m_i(\beta - 1)}. \tag{3.65}$$

Then

$$p \lim \left(\sum_{\beta=1}^{t} G_\beta - \sum_{k=1}^{t-1} y_{i:jk}^2\right) = Cp \lim (G' - F'). \tag{3.66}$$

Writing the Taylor's expansions of $G'$ and of $F'$ in terms of differences $n_{i\alpha}(\beta)/N - m_i(\beta - 1)p_{i\alpha}$, it is easily verified that up to first-order terms, all terms vanish and that the second-order terms of $G'$ are identical with the corresponding terms of $F'$. Thus, $F'$ and $G'$ can differ at most by third-order terms which can be shown to tend to zero in probability. It follows that

$$p \lim (G' - F') = 0, \tag{3.67}$$

and because of (3.66), we have (3.60). Thus, by (3.60), (3.59), (3.51), (3.40) and (3.35), we may write

$$\begin{aligned} &p \lim \left(\chi^2[i:j+1, t] - \sum_{k=1}^{t-1}\sum_{h=1}^{j-1} y_{i:hk}^2 - \sum_{k=1}^{j} y_{i:jk}^2\right) \\ &\qquad = p \lim \left(\chi^2[i:j+1, t] - \sum_{k=1}^{t-1}\sum_{h=1}^{j} y_{i:hk}^2\right) = 0. \end{aligned} \tag{3.68}$$

The induction on $t$ is completely analogous with the roles of the subscripts and of the $m_i(\beta)$ and $p_{i\alpha}$ interchanged. Thus, given (3.32), (3.34) as well as (3.33) follows, completing the induction and the proof. Therefore, the usual $\chi^2$ defined by $\chi^2[i:j+1, t+1]$ associated with the array (3.27) is asymptotically distributed as $\chi^2$ with $jt$ degrees of freedom. It also follows from (3.24) and (3.26) that

$$p \lim \left(\chi^2[i:j+1, t+1] - \sum_{k=1}^{t} \sum_{h=1}^{j} \chi^2_{i:hk}\right) = 0, \tag{3.69}$$

where $\chi^2_{i:hk}$ is the usual asymptotic single degree $\chi^2$ variable associated with the appropriate four-fold table defined by (3.4).

Some corollaries, the first of which has been proved alternatively by Anderson and Goodman [2] follow immediately:

COROLLARY 1. $\chi^2[i:m, T]$ *associated with the entire matrix of observations* (3.3) *for* $i$ *fixed is asymptotically distributed as* $\chi^2$ *with* $(m-1)(T-1)$ *degrees of freedom.* Each of the $(m-1)(T-1)$ components may be identified with one of the four-fold tables of the partition given by (3.4).

COROLLARY 2. $\chi^2[i:j, t+g] - \chi^2[i:j, t] (g > 0)$ *is distributed in the limit as* $\chi^2$ *with* $g(j-1)$ *degrees of freedom.* This difference is asymptotically independent of $\chi^2[i:j, t]$ and its $g(j-1)$ components can be identified with the appropriate four-fold tables of (3.4).

COROLLARY 3. $\chi^2[i:j+r, t] - \chi^2[i:j, t] (r > 0)$ *is distributed in the limit as* $\chi^2$ *with* $r(t-1)$ *degrees of freedom and is asymptotically independent of* $\chi^2[i:j, t]$.

COROLLARY 4. *On nested hypotheses. Let* $\{H_s\}$, $s = 1, \cdots, c$, *be a sequence of hypotheses as follows:*

$$H_1 : p_{ij}(t) = p_{ij} \qquad i = 1, \cdots, m_1; j = 1, \cdots, m_i^{(1)}; t = 1, \cdots, t_i^{(1)}$$

*where* $1 \leqq m_1 \leqq m$, $1 \leqq m_i^{(1)} \leqq m$ *for* $i \leqq m_1$, $m_i^{(1)} = 0$ *for* $i > m_1$, $1 \leqq t_i^{(1)} \leqq T$ *for* $i \leqq m_1$, $t_i^{(1)} = 0$ *for* $i > m_1$ *and*

$$H_s : p_{ij}(t) = p_{ij} \qquad i = 1, \cdots, m_s; j = 1, \cdots, m_i^{(s)}; \quad t = 1, \cdots, t_i^{(s)}; s = 2, \cdots, c$$

*where* $m_{s-1} \leqq m_s \leqq m$, $m_i^{(s-1)} \leqq m_i^{(s)} \leqq m$ *for* $i \leqq m_{s-1}$, $1 \leqq m_i^{(s)} \leqq m$ *for* $m_{s-1} < i \leqq m_s$, $m_i^{(s)} = 0$ *for* $i > m_s$, $t_i^{(s-1)} \leqq t_i^{(s)} \leqq T$ *for* $i \leqq m_{s-1}$, $1 \leqq t_i^{(s)} \leqq T$ *for* $m_{s-1} < i \leqq m_s$ *and* $t_i^{(s)} = 0$ *for* $i > m_s$.

*Let*

$$\chi_1^2 = \sum_{i=1}^{m_1} \chi^2[i : m_i^{(1)}, t_i^{(1)}]$$

*and*

$$\chi_s^2 = \sum_{i=1}^{m_s} \chi^2[i : m_i^{(s)}, t_i^{(s)}] - \sum_{i=1}^{m_{s-1}} \chi^2[i : m_i^{(s-1)}, t_i^{(s-1)}] \qquad s = 2, \cdots, c.$$

*Let*

$$d_s = \sum_{i=1}^{m_s} \sum_{j=1}^{m_i^{(s)}-1} \sum_{t=1}^{t_i^{(s)}-1} I_{i,j,t}^{(s)} \qquad s = 1, \cdots, c$$

*where*

$$I_{i,j,t}^{(s)} = \begin{cases} 1 \textit{ for } 1 \leqq i \leqq m_s, 1 \leqq j \leqq m_i^{(s)} - 1, 1 \leqq t \leqq t_i^{(s)} - 1 \\ 0 \textit{ otherwise.} \end{cases}$$

*Then if $H_s$ ($s = 1, \cdots, c$) is true, the statistics $\chi_1^2, \chi_2^2, \cdots, \chi_s^2$ are independently distributed in the limit as $\chi^2$ with degrees of freedom $d_1, d_2 - d_1, \cdots, d_s - d_{s-1}$ respectively.*

**4. Simultaneous confidence intervals for all linear functions of multinomial probabilities.** Let us assume that we are sampling from $k$ independent multinomial populations. Let $p_{ij}$ ($i = 1, \cdots, k; j = 1, \cdots, c$) be the probability of the $j$th class in the $i$th sequence. The assumption that each multinomial sequence has the same number, $c$, of categories is made for convenience and is in no way necessary for what follows. We also assume that $p_{ij} > 0$ ($i = 1, \cdots, k; j = 1, \cdots, c$). Let $n_i$ be the number of observations from the $i$th multinomial population. Denote by $x_{ij}$ ($i = 1, \cdots, k; j = 1, \cdots, c$) the number of observations in the $j$th class of the $i$th sequence. We shall find asymptotic simultaneous confidence intervals for all linear functions

$$\theta = \sum_{i=1}^{k} \sum_{j=1}^{c} b_{ij} p_{ij}, \tag{4.1}$$

where the $b_{ij}$ may be any real numbers.

Let

$$\hat{\theta} = \sum_{i=1}^{k} \sum_{j=1}^{c} b_{ij} \hat{p}_{ij}, \tag{4.2}$$

where

$$\hat{p}_{ij} = x_{ij}/n_i . \tag{4.3}$$

Then the variance of $\hat{\theta}$ is readily seen to be

$$\sigma^2(\hat{\theta}) = \sum_{i=1}^{k} (1/n_i) \left[ \sum_{j=1}^{c} b_{ij}^2 p_{ij} - \left( \sum_{j=1}^{c} b_{ij} p_{ij} \right)^2 \right]. \tag{4.4}$$

Denote by $S^2(\hat{\theta})$ the estimate of $\sigma^2(\hat{\theta})$ given by

$$S^2(\hat{\theta}) = \sum_{i=1}^{k} (1/n_i) \left[ \sum_{j=1}^{c} b_{ij}^2 \hat{p}_{ij} - \left( \sum_{j=1}^{c} b_{ij} \hat{p}_{ij} \right)^2 \right]. \tag{4.5}$$

Since $\sum_{j=1}^{c} p_{ij} = 1$ ($i = 1, \cdots, k$), we may write

$$\theta = \theta' + \sum_{i=1}^{k} b_{ic}, \tag{4.6}$$

where

$$(4.7) \qquad \theta' = \sum_{i=1}^{k} \sum_{j=1}^{c-1} (b_{ij} - b_{ic}) p_{ij},$$

and putting

$$(4.8) \qquad a_{ij} = b_{ij} - b_{ic} \qquad j = 1, \cdots, c-1; i = 1, \cdots, k,$$

we write

$$(4.9) \qquad \theta' = \sum_{i=1}^{k} \sum_{j=1}^{c-1} a_{ij} p_{ij}.$$

We shall derive asymptotic simultaneous confidence intervals for all $\theta'$ and by adding $\sum_{i=1}^{k} b_{ic}$ to each of the endpoints, we shall then have the desired intervals for all $\theta$.

THEOREM 3. *As* $n_i \to \infty$ $(i = 1, \cdots, k)$ *the lower limit of the probability that the values of all functions* $\theta$ *simultaneously satisfy* $\hat{\theta} - S(\hat{\theta})L \leqq \theta \leqq \hat{\theta} + S(\hat{\theta})L$, *is at least* $1 - \alpha$. Here $L$ is the positive square root of the upper $(1 - \alpha)$th percentage point of the $\chi^2$ distribution with $k(c - 1)$ degrees of freedom.

PROOF. Consider for each $i$,

$$(4.10) \qquad Q_i = \sum_{j=1}^{c} \frac{(x_{ij} - n_i p_{ij})^2}{n_i p_{ij}} = \sum_{j=1}^{c-1} \frac{(x_{ij} - n_i p_{ij})^2}{n_i p_{ij}} + \frac{\left[\sum_{j=1}^{c-1} (x_{ij} - n_i p_{ij})\right]^2}{n_i p_{ic}}.$$

It is well known that $Q_i$ is asymptotically distributed as $\chi^2$ with $c - 1$ degrees of freedom, and because of the independence of the $k$ multinomial sequences, the limiting distribution of

$$(4.11) \qquad Q = \sum_{i=1}^{k} Q_i$$

is a $\chi^2$ distribution with $k(c - 1)$ degrees of freedom.

For each $i$, let

$$(4.12) \qquad y_{ij} = \left[\frac{p_{i1} + \cdots + p_{ij} + p_{ic}}{n_i p_{ij}(p_{i1} + \cdots + p_{i,j-1} + p_{ic})}\right]^{\frac{1}{2}} \left[(x_{ij} - n_i p_{ij}) + \frac{p_{ij}}{p_{i1} + \cdots + p_{ij} + p_{ic}} \sum_{h>j} (x_{ih} - n_i p_{ih})\right] \quad j = 1, \cdots, c-1.$$

It is easily verified that the $y_{ij}$'s are uncorrelated, asymptotically normally distributed variables each with 0 mean and unit variance. Also, since $\sum_{j=1}^{c-1} y_{ij}^2 = Q_i$,

$$(4.13) \qquad \sum_{i=1}^{k} \sum_{j=1}^{c-1} y_{ij}^2 = Q.$$

Now let

$$(4.14)\quad z_{ij} = \left[\frac{x_{i1} + \cdots + x_{ij} + x_{ic}}{x_{ij}(x_{i1} + \cdots + x_{i,j-1} + x_{ic})}\right]^{\frac{1}{2}} \left[(x_{ij} - n_i p_{ij}) + \frac{x_{ij}}{x_{i1} + \cdots + x_{ij} + x_{ic}} \sum_{h>j} (x_{ih} - n_i p_{ih})\right]$$

$$j = 1, \cdots, c-1; i = 1, \cdots, k.$$

Since $(n_i)^{\frac{1}{2}}(x_{ij}/n_i - p_{ij})$ tends to a limiting distribution and because of the consistency of the $x_{ij}/n_i$, it can be shown that

$$(4.15)\quad p \lim (y_{ij} - z_{ij}) = 0.$$

Since the $y_{ij}$ tend to a limiting distribution, it follows from (4.15) that

$$(4.16)\quad p \lim (y_{ij}^2 - z_{ij}^2) = 0,$$

and from this that

$$(4.17)\quad p \lim \left[\sum_{i=1}^{k}\sum_{j=1}^{c-1} y_{ij}^2 - \sum_{i=1}^{k}\sum_{j=1}^{c-1} z_{ij}^2\right] = 0.$$

Thus, the limiting distribution of $\sum_{i=1}^{k}\sum_{j=1}^{c-1} z_{ij}^2$ is the same as that of $Q$, i.e., a $\chi^2$ distribution with $k(c-1)$ degrees of freedom. It follows that

$$(4.18)\quad \text{Prob}\left\{\sum_{i=1}^{k}\sum_{j=1}^{c-1} z_{ij}^2 \leqq L^2\right\} \doteq 1 - \alpha,$$

where the symbol $\doteq$ indicates equality in the limit. Alternatively, we may write

$$(4.19)\quad \text{Prob}\left\{\left|\left(\sum_{i=1}^{k}\sum_{j=1}^{c-1} z_{ij}^2\right)^{\frac{1}{2}}\right| \leqq L\right\} \doteq 1 - \alpha.$$

Now let $\alpha_{i1}, \alpha_{i2}, \cdots, \alpha_{i,c-1}$ $(i = 1, \cdots, k)$ be any sets of real numbers such that $\alpha_{ij} \neq 0$ for at least one pair $(i, j)$ of indices. We have from Schwarz's inequality that

$$(4.20)\quad \left|\sum_{i=1}^{k}\sum_{j=1}^{c-1} \alpha_{ij} z_{ij}\right| \leqq \left|\left(\sum_{i=1}^{k}\sum_{j=1}^{c-1} \alpha_{ij}^2\right)^{\frac{1}{2}}\right| \cdot \left|\left(\sum_{i=1}^{k}\sum_{j=1}^{c-1} z_{ij}^2\right)^{\frac{1}{2}}\right|,$$

from which we may write

$$\frac{\left|\sum_{i=1}^{k}\sum_{j=1}^{c-1} \alpha_{ij} z_{ij}\right|}{\left|\left(\sum_{i=1}^{k}\sum_{j=1}^{c-1} \alpha_{ij}^2\right)^{\frac{1}{2}}\right|} \leqq \left|\left(\sum_{i=1}^{k}\sum_{j=1}^{c-1} z_{ij}^2\right)^{\frac{1}{2}}\right|, \qquad \text{for all } \alpha_{ij}$$

such that $\sum\sum\alpha_{ij}^2 > 0$. Therefore,

$$(4.22)\quad \text{Prob}\left\{\frac{\left|\sum_i\sum_j \alpha_{ij} z_{ij}\right|}{\left|\left(\sum_i\sum_j \alpha_{ij}^2\right)^{\frac{1}{2}}\right|} \leqq L, \quad \text{for all } \alpha_{ij} \text{ such that } \sum\sum \alpha_{ij}^2 > 0\right\}$$

$$\geqq \text{Prob}\left\{\left|\left(\sum_i\sum_j z_{ij}^2\right)^{\frac{1}{2}}\right| \leqq L\right\} \doteq 1 - \alpha.$$

More conveniently, we write

$$\underline{\lim}\ \text{Prob}\ \Big\{\Big|\sum_i \sum_j \alpha_{ij} z_{ij}\Big| \leqq \Big|\Big(\sum_i \sum_j \alpha_{ij}^2\Big)^{\frac{1}{2}}\Big|\, L, \tag{4.23}$$
$$\text{for all } \alpha_{ij} \text{ such that } \sum \sum \alpha_{ij}^2 > 0\Big\} \geqq 1 - \alpha.$$

It should be pointed out that the above use of Schwarz's inequality is analogous to that in the derivation of joint confidence intervals for linear functions of the population means in the analysis of variance.

Consider for each $i$,

$$\begin{aligned}\sum_{j=1}^{c-1} \alpha_{ij} z_{ij} &= \sum_{j=1}^{c-1} \alpha_{ij} \left[\frac{x_{i1} + \cdots + x_{ij} + x_{ic}}{x_{ij}(x_{i1} + \cdots + x_{i,j-1} + x_{ic})}\right]^{\frac{1}{2}} \\ &\cdot \left[(x_{ij} - n_i p_{ij}) + \frac{x_{ij}}{x_{i1} + \cdots + x_{ij} + x_{ic}} \sum_{h>j} (x_{ih} - n_i p_{ih})\right].\end{aligned} \tag{4.24}$$

Setting

$$\begin{aligned}a_{ij} &= \sum_{h=1}^{j-1} n_i \alpha_{ih} \left[\frac{x_{ih}}{(x_{i_1} + \cdots + x_{i,h-1} + x_{ic})(x_{i1} + \cdots + x_{ih} + x_{ic})}\right] \\ &+ n_i \alpha_{ij} \frac{x_{i_1} + \cdots + x_{ij} + x_{ic}}{x_{ij}(x_{i_1} + \cdots + x_{i,j-1} + x_{ic})} \qquad i = 1, \cdots, k; j = 1, \cdots, c - 1,\end{aligned} \tag{4.25}$$

gives

$$\sum_{i=1}^{k} \sum_{j=1}^{c-1} \alpha_{ij} z_{ij} = \sum_{i=1}^{k} \sum_{j=1}^{c-1} a_{ij}(\hat{p}_{ij} - p_{ij}) = \hat{\theta}' - \theta', \tag{4.26}$$

where $\hat{\theta}' = \sum_{i=1}^{k} \sum_{j=1}^{c-1} a_{ij}\hat{p}_{ij}$. The determinants of coefficients of the systems (4.25) are triangular with diagonal elements

$$n_i \left[\frac{x_{i1} + \cdots + x_{ij} + x_{ic}}{x_{ij}(x_{i1} + \cdots + x_{i,j-1} + x_{ic})}\right]^{\frac{1}{2}} \qquad j = 1, \cdots, c - 1.$$

These systems, for each $i$, will have a solution provided all the $x_{ij} > 0$. Because of the assumption that all the $p_{ij} > 0$ and the consistency of the $\hat{p}_{ij}$, this condition will be met with probability tending to 1 as the sample sizes increase indefinitely. The solution is

$$n_i \alpha_{ij} = \left[a_{ij} - \frac{\sum_{h=1}^{j-1} a_{ih} x_{ih}}{x_{i1} + \cdots + x_{i,j-1} + x_{ic}}\right] \left[\frac{x_{ij}(x_{i1} + \cdots + x_{i,j-1} + x_{ic})}{x_{i1} + \cdots + x_{ij} + x_{ic}}\right]^{\frac{1}{2}} \tag{4.27}$$
$$j = 1, \cdots, c - 1; i = 1, \cdots, k.$$

Direct calculation gives

$$\sum_{i-1}^{k} \sum_{j=1}^{c-1} \alpha_{ij}^2 = \sum_{i=1}^{k} 1/n_i \left[\sum_{j=1}^{c-1} a_{ij}^2 \hat{p}_{ij} - \left(\sum_{j=1}^{c-1} a_{ij} \hat{p}_{ij}\right)^2\right] = S^2(\hat{\theta}'), \tag{4.28}$$

where $S^2(\hat{\theta}')$ is the estimate of the variance of $\hat{\theta}'$ analogous to $S^2(\hat{\theta})$. We then have from (4.23), (4.26) and (4.28) that the lower limit of the probability is at least $1 - \alpha$, that all $\theta'$ simultaneously satisfy

$$(4.29) \qquad |\hat{\theta}' - \theta'| \leqq S(\hat{\theta}')L,$$

and by (4.6)

$$(4.30) \qquad |\hat{\theta} - \theta| \leqq S(\hat{\theta}')L.$$

Finally, substituting (4.8) into (4.28), we have that

$$(4.31) \qquad S^2(\hat{\theta}') = S^2(\hat{\theta}),$$

which establishes the theorem.

It is clear that if only categories $j = 1, 2, \cdots, r$, where $r < c - 1$, are of interest, it would be advantageous to group the remaining $c - r$ classes into one category, thereby decreasing the number of degrees of freedom in the relevant $\chi^2$ to $k \cdot r$ which would in turn shorten the lengths of the intervals. I am grateful to the referee for pointing out that if not all sequences are of interest, a similar reduction in the degrees of freedom with a corresponding reduction in the lengths of the intervals can be effected by applying the procedure only to those sequences which are of interest.

**5. Simultaneous confidence intervals for all linear functions of the transition probabilities in a finite Markov chain.** Suppose now that we take $N$(large) observations from a Markov chain as described in Section 2. It follows from the results of Anderson and Goodman [2] that

$$(5.1) \qquad Q = \sum_{i,t} n_i(t-1) \sum_j (\hat{p}_{ij}(t) - p_{ij}(t))^2 / p_{ij}(t)$$

is distributed in the limit as $\chi^2$ with $mT(m-1)$ degrees of freedom. Because of the asymptotic independence of the $\hat{p}_{ij}(t)$ for different $i$ and $t$, their asymptotic joint normality and consistency [2], we have here an analogue to the case of $k$ independent multinomial sequences, with $k$ replaced by $mT$ and the $n_i(t-1)$ assuming the role of the $n_i$. Proceeding in the same manner as in Section 4, we obtain Theorem 4. For an extension of the procedure to the case of simultaneous confidence intervals for a finite number of certain non-linear functions of the transition probabilities, see [3].

THEOREM 4. *Let* $\psi = \sum_{t=1}^{T} \sum_{i=1}^{m} \sum_{j=1}^{m} b_{ij}(t) p_{ij}(t)$, *where the* $b_{ij}(t)$ *are any real numbers. Let* $\hat{\psi} = \sum_{t,i,j} b_{ij}(t)\hat{p}_{ij}(t)$ *and let* $\hat{\sigma}^2(\hat{\psi})$ *be the estimate of the variance of the asymptotic distribution of* $N^{\frac{1}{2}}(\hat{\psi} - \psi)$ *given by*

$$\hat{\sigma}^2(\hat{\psi}) = \sum_{i,t} \frac{N}{n_i(t-1)} \left[ \sum_j b_{ij}^2(t)\hat{p}_{ij}(t) - \left(\sum_j b_{ij}(t)\hat{p}_{ij}(t)\right)^2 \right].$$

*Let* $S^2(\hat{\psi}) = \hat{\sigma}^2(\hat{\psi})/N$ *and denote by* $M$ *the positive square root of the upper* $(1 - \alpha)$th *percentage point of the* $\chi^2$ *distribution with* $mT(m-1)$ *degrees of freedom. Then as* $N \rightarrow \infty$, *the lower limit of the probability that the values of all* *functions* $\psi$ *simultaneously satisfy* $\hat{\psi} - S(\hat{\psi})M \leqq \psi \leqq \hat{\psi} + S(\hat{\psi})M$ *is at least* $1 - \alpha$.

**6. Acknowledgments.** I am deeply grateful to Professor T. W. Anderson for suggesting the problem and for his helpful guidance throughout the work. I am much indebted also to Professors J. Fertig, A. Berger and R. Sitgreaves for many fruitful discussions and suggestions.

## REFERENCES

[1] ANDERSON, T. W. (1951). Probability models for analyzing time changes in attitude. RAND Research Memorandum No. 455.

[2] ANDERSON, T. W. and GOODMAN, LEO A. (1957). Statistical inference about Markov chains. *Ann. Math. Statist.* **28** 89–110.

[3] BERGER, AGNES and GOLD, RUTH Z. (1961). On the comparison of survival times. *Proc. Fourth Berkeley Symp. Mathematical Statistics and Probability* **4** 67–76. University of California Press.

[4] BILLINGSLEY, PATRICK (1961). Statistical methods in Markov chains. *Ann. Math. Statist.* **32** 12–40.

[5] CHERNOFF, HERMAN (1956). Large sample theory: parametric case. *Ann. Math. Statist.* **27** 1–22.

[6] CRAMÉR, H. (1946). *Mathematical Methods of Statistics.* Princeton University Press.

[7] GOODMAN, LEO A. (1956). Some statistical methods for the analysis of certain kinds of social processes. SRC-61003 G 18 rev., dittoed MS, University of Chicago.

[8] GOODMAN, LEO A. (1958). Exact probabilities and asymptotic relationships for some statistics from *m*th order Markov chains. *Ann. Math. Statist.* **29** 476–490.

[9] GOODMAN, LEO A. (1958). Asymptotic distributions of "psi-squared" goodness of fit criteria for *m*th order Markov chains. *Ann. Math. Statist.* **29** 1123–1133.

[10] IRWIN, J. O. (1949). A note on the subdivision of $\chi^2$ into components. *Biometrika.* **36** 130–134.

[11] LANCASTER, H. O. (1949). The derivation and partition of $\chi^2$ in certain distributions. *Biometrika* **36** 117–129.

[12] SCHEFFÉ, HENRY (1953). A method for judging all contrasts in the analysis of variance. *Biometrika* **40** 87–104.

[13] SPENCER, MARILYN (1957). The two-way contingency table and approximate chi-square distributions. Master's Essay, Columbia University Library.

[14] TUKEY, JOHN M. (1953). The problem of multiple comparisons. Dittoed MS, Princeton University.