ON SOME STATISTICAL TESTS FOR *M*TH ORDER MARKOV CHAINS[1]

By Leo A. Goodman

University of Chicago

1. Introduction and summary. Certain ψ^2 statistics were defined by Good in [8] and he stated that there is a "strong analogy" between certain functions of these statistics and some given likelihood ratio (LR) statistics appropriate for testing hypotheses concerning the order of a Markov chain. He also indicated that the analogy held when the hypothesis of "perfect randomness" is true. The present author has indicated in [12] that, for the ψ^2 statistics in [8], this analogy (i.e., the asymptotic equivalence of the corresponding statistics) does not hold under some more general conditions when the "perfect randomness" hypothesis is not true. It will be seen herein that certain functions of a modified form of the ψ^2 statistics are asymptotically equivalent to certain LR statistics in the more general case when the hypothesis $H(P_m)$ that the positively regular Markov chain (see [2]) is governed by a completely specified system P_m of mth order transition probabilities is true. Also, certain functions of a different modified form of the ψ^2 statistics will be seen to be asymptotically equivalent to certain LR statistics in the case when the hypothesis H_m that the positively regular Markov chain is of order m is true. These results are helpful in determining the asymptotic distributions of various statistics and the null hypotheses that can be tested with a given statistic. For example, if a given statistic G is asymptotically equivalent, under $H(P_1)$, to the LR statistic L for testing the null hypothesis $H(P_1)$ within the alternate hypothesis H_2, then the asymptotic distribution, under $H(P_1)$, of G will be χ^2 with a known number of degrees of freedom (i.e., with a known expectation); G can be used directly to test $H(P_1)$ within H_2 (if G is sensitive to these hypotheses), although the asymptotic distribution, under H_2, of G may differ, in a certain sense, from that of L (see Section 6 in [1]). However, if a given statistic ΔG is asymptotically equivalent, under $H(P_1)$, to the LR statistic ΔL for testing the null hypothesis H_1 within the alternate hypothesis H_2 (i.e., if there is an "ostensible analogy" between ΔG and ΔL), but this asymptotic equivalence does not hold under some more general conditions (e.g., under H_1), then ΔG can not be used to test H_1 within H_2; the asymptotic distribution, under $H(P_1)$, of ΔG will be χ^2, but the asymptotic disbution, under the null hypothesis H_1, will not be χ^2, and furthermore the expectation, under H_1, of G can approach infinity (see [15]).

The present author has indicated in [15] that certain functions of a modified

Received June 3, 1958; revised September 24, 1958.

[1] Research carried out at the Statistical Research Center, University of Chicago, under the sponsorship of the Statistics Branch, Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government.

I am indebted to I. J. Good for bringing [18] to my attention, and to S. Ghurye for helping with the Russian translation.

form of the ψ^2 statistics, which were investigated by Stepanow in [18] and which are computed for a specified P_1, are asymptotically equivalent, under $H(P_1)$, to certain LR statistics, but that they will not be equivalent under H_1. Although it is stated in [18] that the results presented there can be applied to the solution of the problem of testing the null hypothesis H_1, it is shown in [15] that none of the statistics in [18] can be used directly to test this composite hypothesis. In the present paper, it will be seen that a statistic based on a different modified form of ψ^2, as well as certain other statistics described herein, will be asymptotically equivalent, under H_1, to a certain LR statistic and can be used to test the null hypothesis H_1.

The ψ^2 statistics defined by Good in [10] are more general than those given in [3], [8], [18]. Besides studying the relation between these statistics and the LR statistics, we shall also discuss certain conjectures proposed in [10] concerning the asymptotic distributions of these statistics, which were investigated by Billingsley [4] for the cases H_0 and H_1 (the author mentions that a more general result for H_m ($m \geqq 0$) can be obtained using similar methods) and independently, using different methods, by the present author [14] for the case H_m ($m \geqq 0$) when the transition probabilities are all positive (this author also mentions that a more general result can be obtained by similar methods). The ψ^2 and LR statistics defined in [10] and [8], as well as some related statistics developed in the present paper, will be generalized further herein, and the asymptotic distributions of these generalized statistics will be investigated. This investigation leads to generalizations of the asymptotic distributions obtained by Good [8], Billingsley [3] [4], and the present author [13] [14], and it helps to clarify the relation between the various statistics.

The different asymptotically equivalent forms of various statistics presented here make it possible for the statistician to choose whichever form he finds preferable both from the computational point of view and also from some other viewpoints (see [1], [5], [11]).

2. The first order chain. Let $\{X_1, X_2, \cdots, X_n\}$ be an observed sequence from a stochastic process. It will be convenient to deal herein with a circularized sequence of observations obtained by regarding the first observation X_1 as immediately following the nth observations X_n (see [8], [12]). In this case, the frequency $f(\underline{u}_s)$ of the s consecutive observations (i.e., the s-tuple) $\underline{u}_s = (u_1, u_2, \cdots, u_s)$ in the circularized sequence will be such that $\sum_{u_1} f(\underline{u}_s) = \sum_{w_s} f(\underline{w}_s) = f(\underline{w}_{s-1})$, where $\underline{w}_s = (w_1, w_2, \cdots, w_s)$, $(w_1, w_2, \cdots, w_{s-1}) = (u_2, u_3, \cdots, u_s) = \underline{w}_{s-1}$, and $f(\underline{w}_{s-1})$ is the frequency of the $(s-1)$-tuple $\underline{w}_{s-1}$ in the circularized sequence. A method of modifying results obtained for circularized sequences so that they can be applied to noncircularized sequences has been given in [12]; results for circularized sequences can not in general be applied directly to noncircularized sequences (see [12] and Corrigenda to [8]).

The following result has been presented in [18]: Consider an observed sequence $\{X_1, X_2, \cdots, X_n\}$ from a positively regular Markov chain with constant transition probability matrix $P_1 = (p_{ij})$, where the possible states are $1, 2, \cdots, a$.

Let p_i' denote the stationary probabilities, and let k_s be the number of s-tuples that are possible given P_1; e.g., if all $p_{ij} > 0$, then $k_s = a^s$. Let $\psi_{1,s}^2 = \sum_{\underline{u}_s} [f(\underline{u}_s) - f_1(\underline{u}_s)]^2/f_1(\underline{u}_s)$, where $f_1(\underline{u}_s) = np'_{u_1}\prod_{i=1}^{s-1} p_{u_i u_{i+1}}$ is the expected value (asymptotically) of $f(\underline{u}_s)$ in the new sequence of length n given $H(P_1)$, and where the summation is taken over the k_s values of $\underline{u}_s$ where $f_1(\underline{u}_s) > 0$. (The $f_1(\underline{u}_s)$ given above is not the exact expected value, but is an asymptotic approximation; similar asymptotic approximations for expected values will be used throughout). Then (a) the statistics $\Delta\psi_{1,s}^2 = \psi_{1,s}^2 - \psi_{1,s-1}^2$ $(s \geqq 2)$ are asymptotically distributed $(n \to \infty)$ as χ^2 with $\Delta k_s = k_s - k_{s-1}$ degrees of freedom (d.f.), and (b) the $\Delta^2\psi_{1,s}^2 = \psi_{1,s}^2 - 2\psi_{1,s-1}^2 + \psi_{1,s-2}^2$ for $s \geqq 3$ are asymptotically independent and distributed as χ^2 with $\Delta^2 k_s = k_s - 2k_{s-1} + k_{s-2}$ d.f.

We shall now introduce ϕ^2 statistics, which are related to, but different from, the ψ^2 statistics. Let $\phi_{1,s}^2 = 2 \sum_{\underline{u}_s} f(\underline{u}_s) \log [f(\underline{u}_s)/f_1(\underline{u}_s)]$. Then, for $s \geqq 2$,

$$\begin{aligned}\Delta\phi_{1,s}^2 &= \phi_{1,s}^2 - \phi_{1,s-1}^2 \\ &= 2 \sum_{\underline{u}_s} f(\underline{u}_s) \log [f(\underline{u}_s)/f_1'(\underline{u}_s)] \\ &= L_{1,s},\end{aligned}$$

where $\underline{u}_{s-1} = (u_1, u_2, \cdots, u_{s-1})$, and $f_1'(\underline{u}_s) = f(\underline{u}_{s-1})p_{u_{s-1}u_s}$ is the expected value (asymptotically) of $f(\underline{u}_s)$ in a new sequence of length n given $f(\underline{u}_{s-1})$ and $H(P_1)$. For $s \geqq 3$,

$$\begin{aligned}\Delta^2\phi_{1,s}^2 &= \Delta\phi_{1,s}^2 - \Delta\phi_{1,s-1}^2 = \Delta L_{1,s} \\ &= 2 \sum_{\underline{u}_s} f(\underline{u}_s) \log [f(\underline{u}_s)/\bar{f}_{s-2}(\underline{u}_s)] \\ &= M_{s-2,s}\end{aligned}$$

where $\underline{w}_{s-2} = (w_1, w_2, \cdots, w_{s-2})$, $\bar{P}_{s-2}$ is the maximum likelihood estimate of the $(s-2)$th order transition probability matrix when H_{s-2} is true (see [14]), and $\bar{f}_{s-2}(\underline{u}_s) = f(\underline{u}_{s-1})f(\underline{w}_{s-1})/f(\underline{w}_{s-2})$ is the expected value (asymptotically) of $f(\underline{u}_s)$ in a new sequence of length n given $f(\underline{u}_{s-1})$ and $H(\bar{P}_{s-2})$. Let

$$K_s = 2 \sum_{\underline{u}_s} f(\underline{u}_s) \log f(\underline{u}_s),$$

$$K_{1,2} = 2 \sum_{\underline{u}_2} f(\underline{u}_2) \log p(\underline{u}_2),$$

and

$$K_{1,1} = 2 \sum_{\underline{u}_1} f(\underline{u}_1) \log np'(\underline{u}_1),$$

where $p(\underline{u}_2) = p_{u_1u_2}$ and $p'(\underline{u}_1) = p'_{u_1}$. Then $\phi_{1,s}^2 = K_s - (s-1)K_{1,2} - K_{1,1}$, $\Delta\phi_{1,s}^2 = \Delta K_s - K_{1,2}$, and $\Delta^2\phi_{1,s}^2 = \Delta^2 K_s$. It can be seen that, given $H(P_1)$, the statistics $\phi_{1,s}^2$ are asymptotically equivalent to the $\psi_{1,s}^2$, and that (a) the $\Delta\phi_{1,s}^2 = \Delta K_s - K_{1,2} = L_{1,s}$ $(s \geqq 2)$ are asymptotically χ^2 with Δk_s d.f., and (b) the $\Delta^2\phi_{1,s}^2 = \Delta L_{1,s} = \Delta^2 K_s = M_{s-2,s}$ are asymptotically independent and distributed as χ^2 with $\Delta^2 k_s$ d.f. (e.g., see methods in [1], [4], [14]).

Let

$$G_{1,s} = \sum_{\underline{u}_s} [f(\underline{u}_s) - f_1'(\underline{u}_s)]^2/f_1'(\underline{u}_s),$$

and

$$F_{s-2,s} = \sum_{\underline{u}_s} [f(\underline{u}_s) - \tilde{f}_{s-2}(\underline{u}_s)]^2/\tilde{f}_{s-2}(\underline{u}_s).$$

Then it can be seen (with methods in [1], [4]) that, given $H(P_1)$, the $G_{1,s}$ are asymptotically equivalent to $L_{1,s} = \Delta\phi_{1,s}^2$ and thus to $\Delta\psi_{1,s}^2$. Hence, the $\Delta G_{1,s} = G_{1,s} - G_{1,s-1}$ are asymptotically equivalent, given $H(P_1)$, to $\Delta^2\psi_{1,s}^2$. Also, the $F_{s-2,s}$ are asymptotically equivalent, given $H(P_1)$, to $M_{s-2,s} = \Delta^2\phi_{1,s}^2$ and thus to $\Delta^2\psi_{1,s}^2$.

The statistics $L_{1,s}$ are the LR statistics for testing the null hypothesis $H(P_1)$ within the alternate hypothesis H_{s-1} (see [2]). Although $G_{1,s}$ and $\Delta\psi_{1,s}^2$ are asymptotically equivalent, given $H(P_1)$, to $L_{1,s}$, it can be seen that, in the case where $H(P_1)$ is not true, $G_{1,s}$ and $\Delta\psi_{1,s}^2$ are asymptotically equivalent to $L_{1,s}$ only in the special sense that the usual χ^2 goodness of fit statistic for the standard test of a simple null hypothesis concerning a multinomial distribution is considered to be asymptotically equivalent (even when this null hypothesis is not true) to the usual LR statistic for this hypothesis (see Section 6 of [1]); i.e., if the null hypothesis $H(P_1)$ is not true, these statistics will be asymptotically equivalent only if the true hypothesis approaches, in a certain sense, $H(P_1)$ at a sufficiently fast rate as $n \to \infty$. The relative advantages and disadvantages of $L_{1,s}$, $\Delta\psi_{1,s}^2$, and $G_{1,s}$ as tests of $H(P_1)$ will not be discussed here, since such discussions for somewhat related problems appear in [1], [5], and [11].

The statistics $M_{s-2,s}$ are the LR statistics for testing the null hypothesis H_{s-2} within H_{s-1}, and their asymptotic distribution, under H_{s-2}, is χ^2 with $\Delta^2 k_s$ d.f. (see [8], [16]). Although $\Delta^2\psi_{1,s}^2$ and $\Delta G_{1,s}$ are asymptotically equivalent, given $H(P_1)$, to $M_{s-2,s}$, it can be seen that the former statistics are not asymptotically equivalent, given H_{s-2}, to $M_{s-2,s}$, and their asymptotic distribution will depend on P_1 (which is used in the computation of $\Delta^2\psi_{1,s}^2$ and $\Delta G_{1,s}$) and on the particular system P_{s-2}' of transition probabilities that is true when H_{s-2} is true (P_{s-2}' can be viewed as a particular $a^{s-2} \times a^{s-2}$ matrix of transition probabilities describing a given Markov chain of order $s - 2$); e.g., for some P_1 that differ from P_{s-2}', the statistics $\Delta^2\psi_{1,s}^2$ will converge in probability, given $H(P_{s-2}')$, to infinity even though the null hypothesis H_{s-2} is true (see [15]). Thus, the asymptotic distribution, given H_{s-2}, of the statistics $\Delta^2\psi_{1,s}^2$ and $\Delta G_{1,s}$ will depend on unknown values of the parameters; these statistics can not be used to test the null hypothesis H_{s-2} in the same simple manner as when the LR statistic $M_{s-2,s}$ is used. However, it can be seen that $F_{s-2,s}$ is equivalent, given H_{s-2}, to $M_{s-2,s}$; and thus can be used to test H_{s-2} within H_{s-1} (see [12], [13]).

3. The general case. Stepanow [18] mentions that, given $H(P_1)$, the asymptotic independence of the $\Delta^2\psi_{1,s}^2$ statistics leads to the fact that the statistics $\Delta\psi_{1,s}^2 -$

$\Delta\psi^2_{1,t}$ $(s > t \geqq 2)$ are asymptotically χ^2 with $\Delta k_s - \Delta k_t$ d.f. It can also be seen, given H_{s-2}, that the $\Delta^2 K_j$ statistics (for $j \geqq s$) are asymptotically independent (see [4], [14]); thus the $\Delta K_s - \Delta K_t$ statistics, given H_{t-1}, are asymptotically χ^2 with $\Delta k_s - \Delta k_t$ d.f. (see [8]). To test H_{t-1} within H_{s-1}, the $\Delta K_s - \Delta K_t$ are the LR statistics (see [8]). The $\Delta\psi^2_{1,s} - \Delta\psi^2_{1,t}$, given H_{t-1}, are not asymptotically equivalent to the LR statistics, and can not serve as a test of H_{t-1} (see [15]).

We have that $\Delta K_s - \Delta K_t = 2\sum_{\underline{u}_s} f(\underline{u}_s) \log (f(\underline{u}_s)/\tilde{f}_{t-1}(\underline{u}_s)) = M_{t-1,s}$, where $\underline{w}_t = (w_1, w_2, \cdots, w_t) = (u_{s-t+1}, u_{s-t+2}, \cdots, u_s)$, $\underline{w}_{t-1} = (w_1, w_2, \cdots, w_{t-1})$, and $\tilde{f}_{t-1}(\underline{u}_s) = f(\underline{u}_{s-1})f(\underline{w}_t)/f(\underline{w}_{t-1})$ is the expected value (asymptotically) of $f(\underline{u}_s)$ in a new sequence of length n given $f(\underline{u}_{s-1})$ and $H(\hat{P}_{t-1})$. Let $F_{t-1,s} = \sum_{\underline{u}_s} [f(\underline{u}_s) - \tilde{f}_{t-1}(\underline{u}_s)]^2/\tilde{f}_{t-1}(\underline{u}_s)$. Then, it can be seen that, given H_{t-1}, the statistics $F_{t-1,s}$ are asymptotically equivalent to $M_{t-1,s} = \Delta K_s - \Delta K_t = \Delta\phi^2_{1,s} - \Delta\phi^2_{1,t}$.

Let

$$\hat{\psi}^2_{t-1,s} = \sum_{\underline{u}_s} [f(\underline{u}_s) - \hat{f}_{t-1}(\underline{u}_s)]^2/\hat{f}_{t-1}(\underline{u}_s),$$

where (for $s > t \geqq 1$)

$$\underline{u}_t = (u_1, u_2, \cdots, u_t), \underline{u}_{t,i} = (u_i, u_{i+1}, \cdots, u_{i+t-1}),$$

$$\underline{u}_{t-1,i} = (u_i, u_{i+1}, \cdots, u_{i+t-2}),$$

and

$$\hat{f}_{t-1}(\underline{u}_s) = f(\underline{u}_{t-1}) \prod_{i=1}^{s-t+1} f(\underline{u}_{t,i})/f(\underline{u}_{t-1,i})$$

is the expected value (asymptotically) of $f(\underline{u}_s)$ in a new sequence of length n given $f(\underline{u}_{t-1})$ and $H(\hat{P}_{t-1})$. Then, given H_{t-1}, the statistics $\hat{\psi}^2_{t-1,s}$ are asymptotically equivalent to

$$\hat{\phi}^2_{t-1,s} = 2\sum_{\underline{u}_s} f(\underline{u}_s) \log [f(\underline{u}_s)/\hat{f}_{t-1}(\underline{u}_s)] = K_s - K_{t-1} - (s - t + 1)(\Delta K_t).$$

Since $\Delta\hat{\phi}^2_{t-1,s} = \Delta K_s - \Delta K_t$, the $\Delta\hat{\phi}^2_{t-1,s}$ or the $\Delta\hat{\psi}^2_{t-1,s}$, as well as the $F_{t-1,s}$, can be used to test H_{t-1} within H_{s-1}.

Consider now the hypothesis $H(P_t)$ that the positively regular Markov chain is of the tth order and governed by the system of transition probabilities P_t (see [2]). Let $\Pr\{u_{t+1} \mid u_1, u_2, \cdots, u_t\} = p_{u_1 u_2 \cdots u_{t+1}} = p(\underline{u}_{t+1})$ denote the transition probability that the j-th observation in the sequence will be u_{t+1}, given that the $(j-1)$-th, $(j-2)$-th, $\cdots$, $(j-t)$-th observations were $u_t, u_{t-1}, u_{t-2}, \cdots, u_1$ respectively, and let $p'(\underline{u}_t)$ denote the stationary (absolute) probability for $\underline{u}_t$. Let $\phi^2_{t,s} = 2\sum_{\underline{u}_s} f(\underline{u}_s) \log[f(\underline{u}_s)/f_t(\underline{u}_s)]$, where (for $s > t \geqq 1$) $f_t(\underline{u}_s) = np'(\underline{u}_t) \prod_{i=1}^{s-t} p_{u_i u_{i+1} \cdots u_{i+t}}$ is the expected value (asymptotically) of $f(\underline{u}_s)$ in a new sequence of length n given $H(P_t)$. Then

$$\Delta\phi^2_{t,s} = \phi^2_{t,s} - \phi^2_{t,s-1} = 2\sum_{\underline{u}_s} f(\underline{u}_s) \log (f(\underline{u}_s)/f'_t(\underline{u}_s)) = L_{t,s},$$

where $f_t'(\underline{u}_s) = f(\underline{u}_{s-1})p_{u_{s-t}u_{s-t+1}\cdots u_s}$ is the expected value (asymptotically) of $f(\underline{u}_s)$ in a new sequence of length n given $f(\underline{u}_{s-1})$ and $H(P_t)$; and

$$\Delta^2\phi_{t,s}^2 = \Delta\phi_{t,s}^2 - \Delta\phi_{t,s-1}^2 = \Delta L_{t,s}$$
$$= 2\sum_{\underline{u}_s} f(\underline{u}_s) \log [f(\underline{u}_s)/f_{s-2}(\underline{u}_s)] = M_{s-2,s} = \Delta^2\phi_{1,s}^2 = \Delta^2\phi_s^2.$$

Since

$$\phi_{t,s}^2 = K_s - (s-t)K_{t,t+1} - K_{t,t},$$

where

$$K_{t,t+1} = 2\sum_{\underline{u}_{t+1}} f(\underline{u}_{t+1}) \log p(\underline{u}_{t+1})$$
$$K_{t,t} = 2\sum_{\underline{u}_t} f(\underline{u}_t) \log np'(\underline{u}_t),$$

we see that

$$\Delta\phi_{t,s}^2 = \Delta K_s - K_{t,t+1}.$$

Let $\psi_{t,s}^2 = \sum_{\underline{u}_s} [f(\underline{u}_s) - f_t(\underline{u}_s)]^2/f_t(\underline{u}_s)$, and $G_{t,s} = \sum_{\underline{u}_s} [f(\underline{u}_s) - f_t'(\underline{u}_s)]^2/f_t'(\underline{u}_s)$. Then it can be seen that, given $H(P_t)$, the $\Delta\psi_{t,s}^2$, $\Delta\phi_{t,s}^2 = L_{t,s}$ and $G_{t,s}$ are all asymptotically equivalent, and each is asymptotically distributed as χ^2 with Δk_s d.f.

The reader will note that k_s and the statistics mentioned in the preceding paragraph depend on P_t. In the case where the null hypothesis tested is H_t, the particular statistics mentioned earlier herein appropriate for such a test do not depend on P_t, but their distribution does since k_s does. If the null hypothesis to be tested is H_t (and P_t is not specified), then the value of k_s to be used can be estimated consistently from the observed number of $\underline{u}_s$ where $f(\underline{u}_s) > 0$. (In the general case where the chain may contain some transient states, but where only the recurrent states are of interest and the transient states are not, the condition $f(\underline{u}_s) > 0$ should be applied to the sequence obtained by omitting all observations before the first one that is in a recurrent state (see [4]); the recurrent states can also be estimated consistently from the observed sequence.)

4. The distribution of the $\hat{\psi}^2$ and $\psi^{\mp 2}$ statistics. Conjectures concerning the asymptotic distribution of $\hat{\psi}_{t,s}^2$ and $\psi_{t,s}^2$ were proposed in [10] and modified forms of these conjectures were proved in [4] and [14]. We now present the following generalization of these results:

(A) The asymptotic distribution of $\hat{\psi}_{t,s}^2$, given H_t, is $*_{\lambda=1}^{s-t-1} K_{g(\lambda)}(x/\lambda)$, where $*$ denotes convolution, $g(\lambda) = \Delta^2 k_{(s+1-\lambda)}$, and $K_g(x)$ is the χ^2 distribution with g degrees of freedom.

(B) Let $\psi_{t,s}^{\mp 2} = \sum_{\underline{u}_s} [f(\underline{u}_s) - f_t^{\mp}(\underline{u}_s)]^2/f_t^{\mp}(\underline{u}_s)$, where

$$f_t^{\mp}(\underline{u}_s) = f(\underline{u}_t)\prod_{i=1}^{s-t} p_{u_i u_{i+1}\cdots u_{i+t}}$$

is the expected value (asymptotically) of $f(\underline{u}_s)$ in a new sequence of length n

given $H(P_t)$ and $f(\underline{u}_t)$. Then the asymptotic distribution $(n \to \infty)$ of $\psi_{t,s}^{\mp 2}$, given $H(P_t)$, is $*_{\lambda=1}^{s-t-1} K_{g(\lambda)}(x/\lambda) * K_{h(t)}[x/(s-t)]$, where $h(t) = \Delta k_{t+1}$.

Statement (A) can be seen to follow from the fact that $\hat{\psi}_{t,s}^2$ is asymptotically equivalent, given H_t, to

$$\hat{\phi}_{t,s}^2 = K_s - K_t - (s-t)\Delta K_{t+1} = \sum_{j=1}^{s-t-1} j\Delta^2 K_{(s+1-j)}, \tag{4.1}$$

where the statistics $\Delta^2 K_s$ (for $s > t + 1$) are asymptotically independent (see [4], [14]). Statement (B) can be seen to follow from the fact that $\psi_{t,s}^{\mp 2}$ is asymptotically equivalent, under $H(P_t)$, to

$$\begin{aligned} \phi_{t,s}^{\mp 2} &= 2 \sum_{\underline{u}_s} f(\underline{u}_s) \log [f(\underline{u}_s)/f_t^{\mp}(\underline{u}_s)] \\ &= K_s - K_t - (s-t)K_{t,t+1} = \hat{\phi}_{t,s}^2 + (s-t)\Delta\phi_{t,t+1}^2 \end{aligned} \tag{4.2}$$

where $\hat{\phi}_{t,s}^2$ and $\Delta\phi_{t,t+1}^2$ are asymptotically independent (see [4], [14]). We also note that the asymptotic distribution of $\psi_{t,s}^{\mp 2}$ (or $\phi_{t,s}^{\mp 2}$) is different from that of $\psi_{t,s}^2$; but the asymptotic distributions, given $H(P_t)$, of $\Delta\psi_{t,s}^2$ and $\Delta\psi_{t,s}^{\mp 2}$ are identical since $\Delta\phi_{t,s}^2 = \Delta\phi_{t,s}^{\mp 2}$.

The results presented here were for $t \geqq 1$. The case where $t = 0$ can be treated in a similar fashion (see [3], [9], [10], [14]).

5. Some generalized statistics and their distributions. From (4.1) we see that, for $s = t + 2$, $\hat{\phi}_{t,s}^2$ and $\hat{\psi}_{t,s}^2$ are asymptotically equivalent, given H_t, to $\Delta^2 K_s = M_{t,s}$, the LR statistic for testing the null hypothesis H_t within the alternate hypothesis H_{s-1} (i.e., $M_{t,s} = -2 \log \lambda_{t,s-1}$, where $\lambda_{t,s-1}$ is the ratio of the maximum likelihood given H_t to that given H_{s-1}). This relationship between $M_{t,s}$ and $\hat{\psi}_{t,s}^2$ (and $\hat{\phi}_{t,s}^2$) does not hold for $s > t + 2$. Also, for $s = t + 1$, $\phi_{t,s}^{\mp 2}$ and $\psi_{t,s}^{\mp 2}$ are asymptotically equivalent, under $H(P_t)$, to $\Delta\phi_{t,s}^2 = L_{t,s}$, the LR statistic for testing the null hypothesis $H(P_t)$ within H_{s-1}. This relationship between $L_{t,s}$ and $\psi_{t,s}^{\mp 2}$ (and $\phi_{t,s}^{\mp 2}$) does not hold for $s > t + 1$. We shall now present, for $s \geqq t + 2$, a generalized statistic $\hat{\phi}_{t,s;r}^2$ that will include both $\hat{\phi}_{t,s}^2$ and $M_{t,s}$ as special cases, and a statistic $\hat{\psi}_{t,s;r}^2$ that will be asymptotically equivalent, given H_t, to $\hat{\phi}_{t,s;r}^2$. Also, we shall present, for $s \geqq t + 1$, a generalized statistic $\phi_{t,s;r}^{\mp 2}$ that will include both $\phi_{t,s}^{\mp 2}$ and $L_{t,s}$ as special cases, and a statistic $\psi_{t,s;r}^{\mp 2}$ that will be asymptotically equivalent, given $H(P_t)$, to $\phi_{t,s;r}^{\mp 2}$. Finally, a generalized statistic $M_{t,s;r}$ (different from $\hat{\phi}_{t,s;r}^2$) will be presented that will include $M_{t,s}$ as a special case, and the asymptotic distribution of each of these generalized statistics will be investigated.

Let

$$\hat{\phi}_{t,s;r}^2 = 2 \sum_{\underline{u}_s} f(\underline{u}_s) \log [f(\underline{u}_s)/\hat{f}_{t;r}(\underline{u}_s)],$$

$$\hat{\psi}_{t,s;r}^2 = \sum_{\underline{u}_s} [f(\underline{u}_s) - \hat{f}_{t;r}(\underline{u}_s)]^2/\hat{f}_{t;r}(\underline{u}_s),$$

$$\phi_{t,s;r}^{\mp 2} = 2 \sum_{\underline{u}_s} f(\underline{u}_s) \log [f(\underline{u}_s)/f_{t;r}^{\mp}(\underline{u}_s)],$$

and

$$\psi^{\ddagger 2}_{t,s;r} = \sum_{\underline{u}_s} [f(\underline{u}_s) - f^{\ddagger}_{t;r}(\underline{u}_s)]^2 / f^{\ddagger}_{t;r}(\underline{u}_s),$$

where

$$\hat{f}_{t;r}(\underline{u}_s) = f(\underline{u}_{t+r}) \prod_{i=1}^{s-t-r} f(\underline{u}_{t+1,i+r}) / f(\underline{u}_{t,i+r}),$$

$$f^{\ddagger}_{t;r}(\underline{u}_s) = f(\underline{u}_{t+r}) \prod_{i=1}^{s-t-r} p_{u_{i+r}u_{i+r+1}\cdots u_{i+r+t}},$$

$$\underline{u}_{t+r} = (u_1, u_2, \cdots, u_{t+r}), \underline{u}_{t+1,i+r} = (u_{i+r}, u_{i+r+1}, \cdots, u_{i+r+t}),$$

and

$$\underline{u}_{t,i+r} = (u_{i+r}, u_{i+r+1}, \cdots, u_{i+r+t-1}) \quad \text{for } 0 \leqq r < s - t.$$

Then for $r = 0$,

$$\hat{\phi}^2_{t,s;r} = \hat{\phi}^2_{t,s}, \qquad \hat{\psi}^2_{t,s;r} = \hat{\psi}^2_{t,s}, \qquad \phi^{\ddagger 2}_{t,s;r} = \phi^{\ddagger 2}_{t,s}, \qquad \psi^{\ddagger 2}_{t,s;r} = \psi^{\ddagger 2}_{t,s};$$

for $r = s - t - 1$,

$$\hat{\phi}^2_{t,s;r} = M_{t,s}, \qquad \hat{\psi}^2_{t,s;r} = F_{t,s}, \qquad \phi^{\ddagger 2}_{t,s;r} = L_{t,s}, \qquad \psi^{\ddagger 2}_{t,s;r} = G_{t,s}.$$

Let

$$M_{t,s;r} = 2 \sum_{\underline{u}_s} f(\underline{u}_s) \log [f(\underline{u}_s)/\tilde{f}_{t;r}(\underline{u}_s)]$$

and

$$F_{t,s;r} = \sum_{\underline{u}_s} [f(\underline{u}_s) - \tilde{f}_{t;r}(\underline{u}_s)]^2 / \tilde{f}_{t;r}(\underline{u}_s),$$

where

$$\tilde{f}_{t;r}(\underline{u}_s) = f(\underline{u}_{s-1-r}) f(\underline{w}_{t+1+r}) / f(\underline{w}_t),$$

$$\underline{w}_{t+1+r} = (w_1, w_2, \cdots, w_{t+1+r}) = (u_{s-t-r}, u_{s-t-r+r}, \cdots, u_s),$$

and

$$\underline{w}_t = (w_1, w_2, \cdots, w_t) \qquad \text{for } 0 \leqq r \leqq s - t - 2.$$

Then for $r = 0$, $M_{t,s;r} = M_{t,s}$ and $F_{t,s;r} = F_{t,s}$; for $r = s - t - 2$, $M_{t,s;r} = \bar{M}_{t,s}$ and $F_{t,s;r} = \bar{F}_{t,s}$, where $\bar{F}_{t,s}$ is $F_{t,s}$ computed for the sequence $\{X_n, X_{n-1}, \cdots, X_1\}$ circularized rather than for the sequence $\{X_1, X_2, \cdots, X_n\}$. The statistic $M_{t,s;r}$ (or $F_{t,s;r}$) can be seen to be the sum of the LR statistics (or the goodness of fit statistics) for testing "independence" in each of k_t "contingency tables" (i.e., $M_{t,s;r}$ is the product of the ratios of the maximum likelihood given independence to that given "nonindependence" in each "contingency table" when normed ln the usual way; viz., -2 times the log of this product) obtained by "splitting" each s-tuple $\underline{u}_s$ into a $(s - t - 1 - r)$-tuple $(u_1, u_2, \cdots, u_{s-t-1-r})$, a t-tuple $(u_{s-t-r}, u_{s-t-r+1}, \cdots, u_{s-r-1})$, and a $(1 + r)$-tuple $(u_{s-r}, u_{s-r+1}, \cdots, u_s)$; for each t-tuple a "contingency table" can be formed where the "expected value" of the observed cell entry $f(\underline{u}_s)$ is $\bar{f}_{t;r}(\underline{u}_s)$ under the assumption of independence in the table (see [13], [14]).

We shall now present simple derivations for the asymptotic distributions of the generalized statistics using a similar approach to that described in Section 4. We have that

$$\hat{\phi}^2_{t,s;r} = K_s - K_{t+r} - (s - t - r)\Delta K_{t+1} = \sum_{j=1}^{s-t-1} d(j)\Delta^2 K_{(s+1-j)}$$

where

$$d(j) = \begin{cases} j & \text{for } 0 \leqq j \leqq s - t - r \\ (s - t - r) & \text{for } s - t - r \leqq j \leqq s - t - 1. \end{cases}$$

Therefore, the asymptotic distribution $(n \to \infty)$ of $\hat{\phi}^2_{t,s;r}$, given H_t, is

$$\mathop{*}_{\lambda=1}^{s-t-1} K_{g(\lambda)}[x/d(\lambda)] = \mathop{*}_{\lambda=1}^{s-t-r-1} K_{g(\lambda)}(x/\lambda) * K_{h(t+r)-h(t)}[x/(s - t - r)].$$

This will also be the asymptotic distribution of $\hat{\psi}^2_{t,s;r}$, under H_t, since $\hat{\psi}^2_{t,s;r}$ and $\hat{\phi}^2_{t,s;r}$ are asymptotically equivalent under H_t.

By a similar approach, we see that

$$M_{t,s;r} = K_s - K_{s-1-r} - (K_{t+1+r} - K_t) = \sum_{j=1}^{s-t-1} c(j)\Delta^2 K_{(s+1-j)},$$

where

$$c(j) = \begin{cases} j & \text{for } 0 \leqq j \leqq v \\ v & \text{for } v \leqq j \leqq s - t - v \\ (s - t - j) & \text{for } s - t - v \leqq j \leqq s - t - 1, \end{cases}$$

and $v = \min[r + 1, s - t - r - 1]$. Therefore, the asymptotic distribution $(n \to \infty)$ of $M_{t,s;r}$ (and $F_{t,s;r}$), given H_t, is

$$\mathop{*}_{\lambda=1}^{s-t-1} K_{g(\lambda)}[x/c(\lambda)]$$

(see [14]).

We also see that

$$\phi^{+2}_{t,s;r} = K_s - K_{t+r} - (s - t - r)K_{t,t+1} = \hat{\phi}^2_{t,s;r} + (s - t - r)\Delta\phi^2_{t,t+1}.$$

Therefore the asymptotic distribution of $\phi^{+2}_{t,s;r}$ (and $\psi^{+2}_{t,s;r}$), given $H(P_t)$, is

$$\mathop{*}_{\lambda=1}^{s-t-1} K_{g(\lambda)}[x/d(\lambda)] * K_{h(t)}[x/(s - t - r)]$$

$$= \mathop{*}_{\lambda=1}^{s-t-r-1} K_{g(\lambda)}(x/\lambda) * K_{h(t+r)}[x/(s - t - r)].$$

If P_t is not completely specified but is a function $P_t(\alpha)$ of a vector parameter $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_v)$ that ranges over an open subset of v-dimensional Euclidean space, and if this function satisfies certain regularity conditions (see [4]), then it can be seen that the statistic $\phi^{\mp 2}_{t,s;r}(\hat{\alpha})$ (or $\psi^{\mp 2}_{t,s;r}(\hat{\alpha})$), obtained by replacing P_t by its maximum likelihood estimate $P_t(\hat{\alpha})$ in the computation of $\phi^{\mp 2}_{t,s;r}$ (or $\psi^{\mp 2}_{t,s;r}$), will have the same asymptotic distribution as $\phi^{\mp 2}_{t,s;r}$ except that the degrees of freedom $h(t + r)$ should be replaced by $h(t + r) - v$. If P_t is the transition probability matrix for a completely unspecified positively regular tth order Markov chain, then $v = \Delta k_{t+1} = h(t)$, $\phi^{\mp 2}_{t,s;r}(\hat{\alpha}) = \hat{\phi}^2_{t,s;r}$, $\psi^{\mp 2}_{t,s;r}(\hat{\alpha}) = \hat{\psi}^2_{t,s;r}$, and the asymptotic distribution, under H_t, of these statistics was given earlier herein. This result is closely related to and generalizes the asymptotic distributions in [4] for $r = 0$ and $s - t - 1$ when $t = 0$ or 1.

This investigation of the asymptotic distributions of various generalized statistics, under particular null hypotheses, indicates that each of these statistics is asymptotically equivalent (under a particular hypothesis) to a weighted sum of the LR statistics $\Delta^2 K_s$, $\Delta^2 K_{s-1}$, $\cdots$, $\Delta^2 K_{t+2}$, $\Delta\phi^2_{t,t+1}$ (or $\Delta\phi^2_{t,t+1}(\hat{\alpha})$, where $\phi^2_{t,t+1}(\hat{\alpha})$ is defined as $\phi^2_{t,t+1}$ with P_t replaced by $P_t(\hat{\alpha})$). The particular generalized statistic that will be appropriate for a given problem will depend in part on the appropriate weighting of the LR statistics, which will in turn depend on the specific null and alternate hypotheses considered (see [4], [14]).

In closing, we point out that the asymptotic mean values of $\hat{\phi}^2_{t,s;r}$ and $M_{t,s;r}$ under H_t, and of $\phi^{\mp 2}_{t,s;r}$ under $H(P_t)$, can be computed directly by reference to the decomposition of the various statistics in terms of the K's. When all transition probabilities are positive, these mean values are $a^{t+r}(a^{s-t-r} - 1) - (s - t - r)a^t(a - 1)$, $a^t(a^{s-t-1-r} - 1)(a^{1+r} - 1)$, and $a^{t+r}(a^{s-t-r} - 1)$ for $\hat{\phi}^2_{t,s;r}$, $M_{t,s;r}$ and $\phi^{\mp 2}_{t,s;r}$, respectively, and they can be given some interpretation in terms of the asymptotic mean values of certain corresponding LR statistics computed from a set of a^t (and a^{t+r}) independent "contingency tables" (see [13], [14]).

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, "The frequency goodness of fit test for probability chains," *Proc. Cambridge Philos. Soc.*, Vol. 47 (1951), pp. 86–95.

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

[4] P. Billingsley, "On testing Markov chains," presented at the meetings of the Institute of Mathematical Statistics, Sept. 10–13, 1957.

[5] W. G. Cochran, "The χ^2-test of goodness of fit," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 315–45.

[6] R. Dawson and I. J. Good, "Exact Markov probabilities from oriented linear graphs," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 946–56.

[7] C. Derman, "Some asymptotic distribution theory for Markov chains with a denumerable number of states," *Biometrika*, Vol. 43 (1956), pp. 285–94.

[8] I. J. Good, "The likelihood ratio test for Markov chains," *Biometrika*, Vol. 42 (1955), pp. 531–3; "Corrigenda," Vol. 44 (1957), p. 301.

[9] I. J. Good, "On the weighted combination of significance tests," *J. Roy. Stat. Soc. Series B (Methodological)*, Vol. 17 (1955), pp. 264–5.

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

[11] I. J. GOOD, "Saddle point methods for the multinomial distribution," *Ann. Math. Stat.*, Vol. 28 (1957), pp. 861–81.

[12] LEO A. GOODMAN, "Simplified runs tests and likelihood ratio tests for Markov chains," *Biometrika*, Vol. 45 (1958), pp. 181–97.

[13] LEO A. GOODMAN, "Exact probabilities and asymptotic relationships for some statistics from mth order Markov chains," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 476–90.

[14] LEO A. GOODMAN, "Asymptotic distributions of 'psi-squared' goodness of fit criteria for mth order Markov chains," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 1123–33.

[15] LEO A. GOODMAN, "A note on Stepanow's tests for Markov chains," *Teoriya Veroyatnostei i ee Primereniya* (The Theory of Probability and Its Applications), Vol. 4 (1959), to be published.

[16] P. G. HOEL, "A test for Markov chains," *Biometrika*, Vol. 41 (1954), pp. 430–3.

[17] G. M. JENKINS, "Contribution to the discussion on 'Regression analysis of binary sequences' by D. R. Cox," *J. Roy. Stat. Soc. Series B (Methodological)*, Vol. 20 (1958), pp. 238–9.

[18] V. E. STEPANOW, "Some statistical tests for Markov chains," (in Russian), *Teoriya Veroyatnostei i ee Primereniya* (The Theory of Probability and Its Applications), Vol. 2 (1957), pp. 143–4.

[19] P. WHITTLE, "Some distribution and moment formulas for Markov chains," *J. Roy. Stat. Soc. Series B (Methodological)*, Vol. 17 (1955), pp. 235–42.