EXACT PROBABILITIES AND ASYMPTOTIC RELATIONSHIPS FOR SOME STATISTICS FROM m-th ORDER MARKOV CHAINS[1]

By Leo A. Goodman

University of Chicago

Summary. An exact formula is presented for the probability of a specified frequency count of m-tuples ($m \geqq 1$) in a sequence $X_1, X_2, \cdots, X_N$ from a Markov chain of order $m - 1$ having a denumerable number $a \leqq \infty$ of states. An exact expression is also obtained for the conditional probability of a specified m-tuple count, given the n-tuple count, when the chain is of order $n - 1$ ($n < m$). If $a < \infty$, then this conditional probability, when regarded as a statistic computed from the observed sequence, is shown to be asymptotically equivalent to the product of the probabilities (regarded as a statistic) associated with a corresponding set of a^{n-1} contingency tables with assigned marginals (each table having a^{m-n} row and a columns), where in each table the two attributes described by the table are independent. This fact leads to several simplified tests, related to standard tests of independence in contingency tables, for the null hypothesis H_{n-1} that the Markov chain is of order $n - 1$ within the alternate hypothesis H_{m-1}. Analogous results are also obtained for circular sequences.

1. Introduction. For a circular sequence, Reed Dawson and I. J. Good [4] have presented an exact expression for the conditional probability of a specified frequency count of m-tuples, given the n-tuple count, in the special case where the sequence is stationary and is of so-called zero Markovity; i.e., all $(N - 1)!$ circular permutations of a sequence of N characters are equally likely. It is also proved in [4] that this expression, obtained under the assumption of zero Markovity, is also valid for "negligible" Markovity; i.e., for a stationary chain of order $n - 1$ or less ($n < m$). (The term "Markovity of order m" means that the Markov chain, from which a (linear) sequence of observations is obtained, is of order m; a definition of a "chain of order m" is given in [10] and in Section 3 here. The circular sequence is defined in [4] as a linear stationary sequence with the ends joined.) For a (linear) sequence of N consecutive observations from a stationary chain of order $n - 1$, the conditional probability of a specified m-tuple count, given the n-tuple count, is presented in [4] as the value obtained by augmenting the linear sequence with a blank placed at the end of the sequence, circularizing the augmented sequence, and then applying the exact expression

Received May 21, 1957; revised October 18, 1957.

[1] Research carried out at the Statistical Research Center, University of Chicago, under 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 and L. J. Savage for some helpful comments, and to R. Dawson and I. J. Good for the opportunity to read [4] before it was published.

for circular sequences to it. The treatment of linear sequences presented in the present paper is more direct, and leads to some different results from those given n [4]. An exact expression is given here for the probability of a specified m-tuple count in a sequence from a chain in the more general case where it need not be stationary and can be of order $m - 1$ (a case of nonnegligible Markovity). An exact formula is also obtained for the conditional probability of a specified m-tuple count $f_{i_1 \cdots i_m}$ in a linear sequence, given the n-tuple counts

$$f_{i_1 \cdots i_n \cdot} = \sum_{i_{n+1}} \sum_{i_{n+2}} \cdots \sum_{i_m} f_{i_1 \cdots i_m} \quad \text{and}$$

$$f_{\cdot i_{m-n+1} \cdots i_m} = \sum_{i_1} \sum_{i_2} \cdots \sum_{i_{m-n}} f_{i_1 \cdots i_m}$$

when the chain need not be stationary and can be of order $n - 1$. Even in the case where the chain is stationary, the formula developed here refers to a different question and is numerically different from that presented in [4]. (In [4], the conditional probability of a specified m-tuple count $f_{i_1 \cdots i_m}$, given the n-tuple count $f_{i_1 \cdots i_n}$, is presented for the stationary chain.) We shall see that, for a (linear) sequence of observations from a chain, it appears to be more relevant to compute the conditional probability when the n-tuple counts $f_{i_1 \cdots i_n \cdot}$ and $f_{\cdot i_{m-n+1} \cdots i_m}$ are given, rather than when the n-tuple count $f_{i_1 \cdots i_n}$ is given.

For stationary circular sequences, it is proved in [4] that, when the chain is of zero order and has $a < \infty$ states, then the conditional probability of the observed m-tuple count $\bar{f}_{i_1 \cdots i_m}$, given the 1-tuple count $\bar{f}_i$ (when this probability is regarded as a statistic), is asymptotically equivalent (for N large and $\bar{f}_i/N \to k_i > 0$), to the probability of the cell entries $\bar{f}_{i_1 \cdots i_m}$ in a contingency table with assigned marginals $\bar{f}_{i_1 \cdots i_{m-1}}$ and $\bar{f}_i$, when the two attributes described by the table are independent. In the present paper, this result is generalized to show the asymptotic equivalence, when the chain is of order $n - 1$, between the conditional probability of the observed m-tuple count, given the n-tuple count, and the product of the probabilities of a corresponding set of cell entries in a^{n-1} contingency tables with assigned marginals. An analogous result is also obtained for linear sequences from a chain of order $n - 1$. (The result in [4] for stationary circular sequences of zero order cannot be applied directly to the conditional probability, presented in [4], for the circularized augmented linear sequence (since the 1-tuple count f_B for the augmented blank is 1 and $f_B/N \to 0$); the authors in [4] refer the reader to the present paper for results for linear sequences).

These results lead to the fact that any asymptotic test of contingency for the a^{n-1} independent contingency tables can be used to test the null hypothesis H_{n-1} that the Markov chain is of order $n - 1$ within the alternate hypothesis H_{m-1}. The likelihood ratio test of H_{n-1} within H_n, given by P. G. Hoel [10], can be seen to be of the same form as the joint likelihood ratio test of contingency computed for the a^{n-1} independent contingency tables. The test of H_{n-1} within H_{m-1}, presented by I. J. Good [8] for the circularized sequence, can be seen to be of the same form as the joint likelihood ratio test for the a^{n-1} contingency tables related to the frequency count for the circularized (but not augmented) sequence. (Good also deals with the linear sequence in [8], but he agrees that his paper contains some slips. In applying results obtained for circular sequences to linear sequences, there is a real possibility of errors. (See Corrigenda to [8] and Leo A. Goodman [9].)) For the linear sequence, the likelihood ratio test of H_{n-1} within H_{m-1}, and the χ^2-test of the form used in contingency tables (which is equivalent to the likelihood ratio test), were presented by T. W. Anderson and Leo A. Goodman [1]; but these authors were concerned mainly, in [1], with a large number v of sequences of N consecutive observations from a chain with a finite number of states, where $v \to \infty$ and N was fixed and could, in fact, be small. There was one brief section in [1] dealing with $v = 1$ and $N \to \infty$, and it was based on a long sequence (asymptotic) result, due to M. S. Bartlett [2], concerning the 2-tuple count. The results developed in the present paper are based directly on the exact formula for the distribution of the m-tuple count when $v = 1$ and the chain has denumerably many states.

The approach used here is related to, but different from, earlier work ([1], [2], [6], [13]), where the observed transition proportions were shown to have some properties similar to those of the observed proportions from a set of independent multinomial distributions.

The exact formula developed here for the distribution of the m-tuple count from a chain of order $m - 1$ is a generalization of a result, due to P. Whittle [13], for the special case of $m = 2$. A different, and perhaps simpler, proof of the result in [13] will be presented, and it will be related to the work in [4]. The generalization in the present paper is based directly on this result.

When indicating how many degrees of freedom certain statistics (which were asymptotically χ^2) had, most of the articles mentioned in this section assumed (either explicitly or implicitly) that all the transition probabilities in the Markov chain were positive; for the sake of simplicity, we shall do likewise here when indicating the size of certain contingency tables (and thus how many degrees of freedom the χ^2 statistics corresponding to these tables have). If some of these probabilities are zero, then the methods developed in the present paper can be modified in a straightforward manner to obtain analogous results (see [2]).

2. The 2-tuple and 1-tuple counts. Suppose that a sequence $X_1, X_2, \cdots, X_N$ is obtained from a first order Markov chain with constant transition probability matrix $P = [p_{ij}]$; i.e., the probability is p_{ij} that $X_t = j$, given that $X_{t-1} = i$. For the sake of simplicity, we first assume that the chain has a finite number $a < \infty$ of states. We write f_{ij} for the frequency in the sequence of the 2-tuple (i, j) $(i, j = 1, 2, \cdots, a)$; we also write $\sum_j f_{ij} = f_{i\cdot}$ and $\sum_j f_{ji} = f_{\cdot i}$. If the chain begins in state r and ends in state s $(X_1 = r$ and $X_N = s)$, then

$$f_{i\cdot} - f_{\cdot i} = \delta_{ir} - \delta_{is} \qquad (i = 1, 2, \cdots, a), \tag{1}$$

and

$$\sum_i f_{i\cdot} = \sum_i f_{\cdot i} = N - 1, \tag{2}$$

where δ_{ij} equals 1 or 0 according as i and j are equal or unequal. The following result, based on the work in [13], will be used here. Let $T_s(f_{ij})$ be the (sr)th cofactor of the $a \times a$ matrix $[\delta_{ij} - f_{ij}/f_{i\cdot}] = \hat{M}$ if the f_{ij} satisfy (1) and (2), and let it be zero otherwise. (It can be seen that $T_s(f_{ij})$ does not depend on r and is nonnegative.) Then the probability $\prod_r (f_{ij}, s)$ that the 2-tuple count will be $f_{ij}(i, j = 1, 2, \cdots, a)$ and that the sequence ends in s, given that it begins with r, is

$$T_s(f_{ij}) \frac{\prod_i f_{i\cdot}!}{\prod_i \prod_j f_{ij}!} \prod_i \prod_j p_{ij}^{f_{ij}}. \tag{3}$$

(Actually, it is stated in [13] that (3) is the probability $\prod_{rs} (f_{ij})$ that the 2-tuple count is $f_{ij}(i, j = 1, 2, \cdots, a)$, given that the sequence begins with r and ends with s; this is not quite correct, but can easily be corrected, as has been done here.)

Formula (3) will hold only if $N \geqq a$, and $f_{i\cdot} > 0$ and $f_{\cdot i} > 0 (i = 1, 2, \cdots, a)$. However, for $N < a$ or some $f_{i\cdot}$ or $f_{\cdot i}$ equal to 0, (3) still holds if calculated on the basis of a process including only those states that have been observed (see [13]).

A proof of (3), different from that given in [13], will now be presented, since it may increase the understanding of this formula and also since a somewhat different procedure for computing (3) is obtained. This proof uses an approach similar to that applied in [4] to circular sequences with negligible Markovity. It is based on the following combinatorial theorem, called the BEST theorem in [4] (due to N.G. de Bruijn, T. van Aardenne-Ehrenfest, C.A.B. Smith and W.T. Tutte [5]): Given any $a \times a$ matrix $M = [m_{ij}]$ of nonnegative integers, there corresponds an oriented linear graph, with vertices $1, 2, \cdots, a$, such that the number of oriented paths (edges) leading from vertex i to vertex j equals m_{ij}. The matrix, unique to within the same rearrangement of rows as of columns, is called the incidence matrix of the corresponding oriented linear graph. The graph is defined as simple if $m_i = \sum_j m_{ij} = \sum_j m_{ji}$. A circuit in such a graph is defined as a unicursal path passing exactly once through each edge (in the right direction). Let $M' = [m'_{ij}]$ be the $a' \times a'$ matrix formed from M by deleting every row and column consisting wholly of zeros. Then $\sum_j m'_{ij} = \sum_j m'_{ji} = m'_i > 0$, for $i = 1, 2, \cdots, a'$. Let $M^* = [m^*_{ij}]$, where $m^*_{ij} = m'_i\delta_{ij} - m'_{ij}$. Since M^* is a square matrix with each row and column summing to zero, the cofactors of its elements are all equal; let $\| M^* \|$ be the common value of these cofactors. Then the BEST theorem asserts that the number $C(M)$ of distinct circuits, when all the edges are distinguishable, in a simple oriented linear graph with incidence matrix M is

$$C(M) = \| M^* \| \cdot \prod_{i=1}^{a'} (m_i - 1)!.$$

Let $N_{rs}(M)$ be the number of circuits that begin at vertex r and end at vertex s; i.e., the number of paths that pass once through each edge, except for one of the edges leading from vertex s to vertex r. Then, when all the m_{ij} oriented edges from vertex i to j are distinguishable, we have that $N_{rs}(M) = C(M)m_{sr}$. If these edges are not distinguishable, then the number of circuits that begin at vertex r and end at vertex s is $U_{rs}(M) = N_{rs}(M)/\prod_{ij} m_{ij}! = C(M)/\prod_{ij} f_{ij}!$, where $f_{sr} = m_{sr} - 1$ and $f_{ij} = m_{ij}$ for $(i, j) \neq (s, r)$; $U_{rs}(M)$ is the number of paths that pass directly from vertex i to j in total f_{ij} times, and that begin at vertex r and end at s. If $\sum_i \sum_j f_{ij} = N - 1$, the probability of observing any given path (a sequence of vertices or states) that begins at r and ends at s in a sequence $X_1, X_2, \cdots X_N$ from a chain with transition probability matrix $P = [p_{ij}]$, given that $X_1 = r$, is $\prod_{ij} p_{ij}^{f_{ij}}$. Since the number of such paths is $U_{rs}(M)$, the probability of observing one of these paths is

$$\prod_r (f_{ij}, s) = U_{rs}(M) \prod_{ij} p_{ij}^{f_{ij}} = \left[\frac{C(M)}{\prod_{ij} f_{ij}!}\right] \prod_{ij} p_{ij}^{f_{ij}}$$

$$= \left[\frac{T_s(f_{ij}) \prod_{i=1}^{a'} (m_i' - 1)! \prod_{i \neq s} m_i'}{\prod_{ij} f_{ij}!}\right] \prod_{ij} p_{ij}^{f_{ij}}$$

$$= \left[\frac{T_s(f_{ij}) \prod_i f_{i\cdot}!}{\prod_{ij} f_{ij}!}\right] \prod_{ij} p_{ij}^{f_{ij}},$$

where $T_s(f_{ij})$ is the cofactor of an element in the sth row of the matrix $M^{**} = [m_{ij}^{**}]$, where $m_{ij}^{**} = m_{ij}^{*}/m_i'$. Q.E.D.

A similar proof was also independently obtained by Dawson and Good in an unpublished note. This proof indicates that (3) holds even when a is infinite, since it depends essentially on the $a' \times a'$ matrix M^*, where a' is finite when N is finite, rather than on the $a \times a$ matrix M. Thus, the exact formula, presented in [13] for the chain with a finite number of states, also holds for the chain with denumerably many states. (See [6] for some asymptotic distribution theory for first order chains with denumerably many states.[2]) This proof also indicates that $\prod_r (f_{ij}, s)$ can be computed from the expression $[C(M)/\prod_{ij} f_{ij}!] \prod_{ij} p_{ij}^{f_{ij}}$, where $m_{ij} = f_{ij}$ for $(i, j) \neq (s, r)$ and $m_{sr} = f_{sr} + 1$, which may sometimes be simpler to apply directly than (3).

If r is given, and the f_{ij} satisfy (2) and also (1) for some s, then that s is unique. Thus, s can be determined by (1) as a function of the f_{ij} when r is given. Since the probability $\prod_r (f_{ij})$ of the 2-tuple count f_{ij}, given that $X_1 = r$, is obtained by

$$\prod_r (f_{ij}) = \sum_{x=1}^{a} \prod_r (f_{ij}, x); \tag{4}$$

and since $\prod_r (f_{ij}, x)$ is 0 for all values of $x \neq s$, we have that $\prod_r (f_{ij})$ is equal to (3), if the f_{ij} satisfy (2) and also (1) for some s.

The probability $\prod_r (f_{ij} \mid s)$ of the 2-tuple count $f_{ij}(i, j = 1, 2, \cdots)$, given that the sequence begins with r and ends with s (which is the verbal (not quite correct) description that was given in [13] for (3)), can actually be obtained by

[2] I am indebted to K. L. Chung for bringing [6] to my attention.

dividing $\prod_r (f_{ij}, s)$ by the probability $p_{rs}^{[N-1]}$ that the sequence ends with s: $p_{rs}^{[N-1]}$ is the (rs)th element of the transition probability matrix $P^{N-1} = [p_{ij}^{[N-1]}]$.

If X_1 is a random variable with a probability p_r of being in the rth state, then the probability $\prod (f_{ij}, s, r)$ that the 2-tuple count is f_{ij}, that $X_N = s$ and $X_1 = r$, is simply $\prod_r (f_{ij}, s)p_r$. The general approach given here can also be used to obtain exact expressions for the probability $\prod (f_{ij}, \cdot, r)$ that the 2-tuple count is f_{ij} and that $X_1 = r$, for the probability $\prod (f_{ij}, s)$ that the 2-tuple count is f_{ij} and that $X_N = s$, for the probability $\prod (f_{ij})$ that the 2-tuple count is f_{ij}, and also for various conditional probabilities.

The distribution of the $f_{\cdot j} = \sum_i f_{ij}$ will now be studied, when the chain is of zero order; i.e., $p_{ij} = p_{\cdot j}$ for $i, j = 1, 2, \cdots$. The $f_{\cdot j}$ are the 1-tuple frequencies in the sequence $X_2, X_3, \cdots, X_N$; i.e., $f_{\cdot j}$ is the number of observations in state j among this sequence. The probability $\prod_r (f_{\cdot j}, s)$ that the 1-tuple count in this sequence is $f_{\cdot j}$ and that $X_N = s$, can be derived using the standard multinomial formula, and we obtain

$$\prod_r (f_{\cdot j}, s) = \left(\frac{f_{\cdot s}}{N-1}\right) \frac{(N-1)!}{\prod_j f_{\cdot j}!} \prod_j p_{\cdot j}^{f_{\cdot j}}. \tag{5}$$

Therefore, for a zero order chain, the conditional probability $\prod_r (f_{ij} \mid f_{\cdot j}, s)$ of the 2-tuple count f_{ij}, given the $f_{\cdot j}$ and s and r, can be obtained by dividing (3) by (5), when the f_{ij} satisfy (1), (2), and also $\sum_i f_{ij} = f_{\cdot j}$. (We can assume, without loss of generality, that $f_{\cdot j}$ and s are such that $\prod_r (f_{\cdot j}, s) > 0$.) Thus

$$\prod_r (f_{ij} \mid f_{\cdot j}, s) = \left[T_s(f_{ij}) \Big/ \left(\frac{f_{\cdot s}}{N-1}\right)\right]\left[\frac{\prod_i f_{i\cdot}! \prod_j f_{\cdot j}!}{\prod_i \prod_j f_{ij}!(N-1)!}\right]: \tag{6}$$

the second factor is the probability $P(f_{ij} \mid f_{\cdot j}, f_{i\cdot})$ of the cell entries f_{ij} in an ordinary contingency table with assigned marginals $f_{\cdot j}$ and $f_{i\cdot}$, where the two attributes described by the table are independent (see, e.g., [12], p. 278).

Since the $f_{i\cdot}$ can be determined by (1) from the $f_{\cdot j}$, r, and s, we have that (6) is also the conditional probability $\prod_r (f_{ij} \mid f_{\cdot j}, f_{i\cdot})$ of the f_{ij}, given the $f_{\cdot j}, f_{i\cdot}$, and r; and (5) is the conditional probability $\prod_r (f_{\cdot j}, f_{i\cdot})$ of the $f_{\cdot j}$ and $f_{i\cdot}$, given r. Furthermore, since the 1-tuple count f_j in the sequence $X_1, X_2, \cdots, X_N$ can be determined from $f_{\cdot j}$ and r by the relation $f_j = f_{\cdot j} + \delta_{jr}$, (6) is also the conditional probability $\prod_r (f_{ij} \mid f_j, s)$ of the 2-tuple count f_{ij}, given the 1-tuple count f_j, and r and s.

From (6), we obtain

$$T_s(f_{ij}) \Big/ \left(\frac{f_{\cdot s}}{N-1}\right) = \prod_r (f_{ij} \mid f_{\cdot j}, f_{i\cdot}) / P(f_{ij} \mid f_{\cdot j}, f_{i\cdot}). \tag{7}$$

We shall now prove that the statistic (7) converges in probability to unity (thus $\prod_r (f_{ij} \mid f_{\cdot j}, f_{i\cdot})$ and $P(f_{ij} \mid f_{\cdot j}, f_{i\cdot})$ are asymptotically equivalent), if the chain is of zero order and $p_{ij} = p_{\cdot j} > 0$. In this case, $f_{\cdot s}/(N-1)$ converges in probability to $p_{\cdot s}$, and it will be necessary to prove only that $T_s(f_{ij})$ converges in probability to $p_{\cdot s}$.

For the sake of simplicity, assume that the chain has a finite number of states and that $f_{i\cdot} > 0$ for $i = 1, 2, \cdots, a$. The $a \times a$ matrix $\hat{M}$ will converge in probability to $M = [\delta_{ij} - p_{\cdot j}]$, and $T_s(f_{ij})$, which is the (sr)th cofactor of $\hat{M}$, will therefore converge in probability to the (sr)th cofactor of M. Since the sum of the entries in each row of M is zero, the cofactors in row s are all equal to the (ss)th cofactor $| M_s |$, the determinant of the $(a - 1) \times (a - 1)$ matrix M_s obtained by deleting the sth row and column in M. By some elementary transformations of M_s, or by the identities between the cofactors and the elements of a matrix (see, e.g., p. 109 in [3]) and the relationship between the principal minors and the characteristic equation (see, e.g., p. 19 in [11]), we see that $| M_s | = p_{\cdot s}$. Hence, $T_s(f_{ij})$ converges in probability to $p_{\cdot s}$. Q.E.D.

This result, concerning the asymptotic equivalence (under the assumption H_0 of zero Markovity) of $\prod_r (f_{ij} | f_{\cdot j}, f_{i\cdot})$ and $P(f_{ij} | f_{\cdot j}, f_{i\cdot})$, implies that the null hypothesis H_0 can be tested, within H_1, by any asymptotic test of contingency in the contingency table with cell entries f_{ij} and with assigned marginals $f_{\cdot j}$ and $f_{i\cdot}$. This implication follows from an application of the following lemma proved in [4]: If (a) an experiment (with parameter N) has, for each value of N (positive integers tending to infinity) a finite set $F^N = \{F_i^N\}$ of possible outcomes, (b) P_N (or simply P) and P'_N (or simply P') are two probability measures over F^N such that $P'(F_i^N)/P(F_i^N)$ converges in the probability P to unity, where $P'(F_i^N)/P(F_i^N)$ is regarded as a statistic whose distribution is determined by P, and (c) $S(F_i^N)$ is a statistic whose cumulative distribution function Φ_N converges, as N becomes infinite, to a limiting distribution Φ under P, then the distribution function Φ'_N of $S(F_i^N)$ under P' also converges to the same limiting distribution Φ. This lemma can be applied, in order to obtain the desired implication, by taking $P(F_i^N) = [P(f_{ij} | f_{\cdot j}, f_{i\cdot}) \prod_r (f_{\cdot j}, f_{i\cdot})]$, and $P'(F_i^N) = \prod_r (f_{ij}, s)$. Since H_0 is assumed, $P'(F_i^N)/P(F_i^N) = \prod_r(f_{ij} | f_{\cdot j}, f_{i\cdot})/P(f_{ij} | f_{\cdot j}, f_{i\cdot})$ will converge in probability to unity. Since any asymptotic test of contingency in the contingency table with cell entries f_{ij} and with assigned marginals $f_{\cdot j}$ and $f_{i\cdot}$ will have the same asymptotic distribution under $P(f_{ij} | f_{\cdot j}, f_{i\cdot})$ (i.e., in the standard case) as under $P(F_i^N)$ (since the $f_{\cdot j}/N$ and $f_{i\cdot}/N$ converge in probability to $p_{\cdot j}$ and $p_{\cdot i}$ respectively), it follows from the lemma that the same standard asymptotic distribution will also hold under $P'(F_i^N)$ (i.e., when H_0 is true).

Since the sequence obtained from the chain is finite, it will not provide estimates of p_{ij}, for all i, j, if the chain has a denumerable infinity of states (see [6]). Thus, when $a = \infty$, select (independently of the data) a finite subset of, say, b states, and consider all states that are not included in this subset as belonging to a single state; i.e., reduce the original number of states to $b + 1 = a'$ in the modified sequence. The tests of H_0, suggested in this section for the case where $a < \infty$, can be applied to the modified sequence consisting of a' states, and the results presented will hold also for this case, A rejection of H_0 for the modified sequence would imply a rejection of this hypothesis for the original chain consisting of denumerably many states. This general method is applied to some different hypotheses relating to Markov chains on p. 293 in [6].

3. The $n + 1$-tuple and the n-tuple counts. Suppose that a sequence $X_1, X_2, \cdots, X_N$ is obtained from a Markov chain of order $n(n \geqq 1)$, where the probability is $p_{\mathfrak{i}j}$ that $X_t = j$, given that $(X_{t-n}, X_{t-n+1}, \cdots, X_{t-1}) = (i_1, i_2, \cdots, i_n) = \mathfrak{i}$. For the sake of simplicity, we assume that there are a states in this chain; i.e., X_t can take its value from among a possible values of j. We define a new sequence of random vectors $Z_1 = (X_1, X_2, \cdots, X_n)$, $Z_2 = (X_2, X_3, \cdots, X_{n+1})$, $\cdots$, $Z_{N-n+1} = (X_{N-n+1}, X_{N-n+2}, \cdots, X_N)$ where each vector can take its value from among the a^n possible values of $\mathfrak{i}$. The probability $p_{\mathfrak{i}\mathfrak{f}}$ that $Z_t = \mathfrak{f}$, given that $Z_{t-1} = \mathfrak{i}$, is equal to $p_{\mathfrak{i}j_n}$ for $i^* = j'$, where $i^* = (i_2, i_3, \cdots, i_n)$, $j^t = (j_1, j_2, \cdots, j_{n-1})$, $\mathfrak{i} = (i_1, i^*)$, $\mathfrak{f} = (j', j_n)$, and $p_{\mathfrak{i}\mathfrak{f}}$ is zero otherwise. The sequence $Z_1, Z_2, \cdots, Z_{N-n+1}$, is from a first order chain with constant transition probability matrix $P_n = [p_{\mathfrak{i}\mathfrak{f}}]$. This chain has a^n states; P_n is an $a^n \times a^n$ matrix (see [2]).

The frequency $f_{\mathfrak{i}\mathfrak{f}}$ of the 2-tuple $(\mathfrak{i}, \mathfrak{f})$ in the sequence of $(N - n + 1)$ observed Z's gives the $(n + 1)$-tuple frequency $f_{\mathfrak{i}j_n}$ in the sequence of N observed X's for all $(\mathfrak{i}, \mathfrak{f})$ where $\mathfrak{f} = (i^*, j_r)$, and $f_{\mathfrak{i}\mathfrak{f}}$ will be zero otherwise. In other words, the frequency $f_{\mathfrak{i}i_{n+1}}$ in the sequence of X's of the $(n + 1)$-tuple $(i_1, i_2, \cdots, i_n, i_{n+1})$ is the number of values of t for which $(X_t, X_{t+1}, X_{t+2}, \cdots, X_{t+n}) = (\mathfrak{i}, i_{n+i})$, *i.e.* the number $f_{\mathfrak{i}\mathfrak{f}}$ of values of t for which $Z_t = (i_1, i_2, \cdots, i_n) = \mathfrak{i}$ and $Z_{t+1} = \mathfrak{f}$, for $\mathfrak{f} = (i^*, i_{n+1})$. Since $f_{\mathfrak{i}\mathfrak{f}}$ is the 2-tuple count in a sequence from a first order chain, (3) can be applied to obtain the probability $\prod_{\mathfrak{r}}(f_{\mathfrak{i}\mathfrak{f}}, \mathfrak{s})$ that the $(n + 1)$-tuple count in the observed sequence of X's will be $f_{\mathfrak{i}j_n}$ and that $Z_{N-n+1} = \mathfrak{s}$, given that $Z_1 = \mathfrak{r}$. We obtain

$$\prod\nolimits_{\mathfrak{r}} (f_{\mathfrak{i}\mathfrak{f}}, \mathfrak{s}) = T_{\mathfrak{s}}(f_{\mathfrak{i}\mathfrak{f}}) \frac{\prod_{\mathfrak{i}} f_{\mathfrak{i}\cdot}!}{\prod_{\mathfrak{i}} \prod_{\mathfrak{f}} f_{\mathfrak{i}\mathfrak{f}}!} \prod_{\mathfrak{i}} \prod_{\mathfrak{f}} p_{\mathfrak{i}\mathfrak{f}}^{f_{\mathfrak{i}\mathfrak{f}}} \tag{8}$$

$$= \prod\nolimits_{\mathfrak{r}} (f_{\mathfrak{i}j}, \mathfrak{s}) = T_{\mathfrak{s}}(f_{\mathfrak{i}\mathfrak{f}}) \frac{\prod_{\mathfrak{i}} f_{\mathfrak{i}\cdot}!}{\prod_{\mathfrak{i}} \prod_j f_{\mathfrak{i}j}!} \prod_{\mathfrak{i}} \prod_j p_{\mathfrak{i}j}^{f_{\mathfrak{i}j}}, \tag{9}$$

where $T_{\mathfrak{s}}(f_{\mathfrak{i}\mathfrak{f}})$ is the $(\mathfrak{s}\mathfrak{r})$th cofactor of the $a^n \times a^n$ matrix $[\delta_{\mathfrak{i}\mathfrak{f}} - f_{\mathfrak{i}\mathfrak{f}}/f_{\mathfrak{i}\cdot}] = \hat{M}_n$. This result could also be obtained by applying the BEST theorem to the vertices $\mathfrak{i}$.

The probability $\prod_{\mathfrak{r}}(f_{\mathfrak{i}\mathfrak{f}}) = \prod_{\mathfrak{r}}(f_{\mathfrak{i}j})$ of the $(n + 1)$-tuple count $f_{\mathfrak{i}j}(j = 1, 2, \cdots, a$, and $\mathfrak{i} = 1, 2, \cdots, a^n)$, given that $Z_1 = \mathfrak{r}$, can be obtained from (9), by applying (1) and (2) to the sequence of Z's. Also, the probability $\prod_{\mathfrak{r}}(f_{\mathfrak{i}\mathfrak{f}} \mid \mathfrak{s}) = \prod_{\mathfrak{r}} (f_{\mathfrak{i}j} \mid \mathfrak{s})$ of the $(n + 1)$-tuple count $f_{\mathfrak{i}j}$, given that $Z_1 = \mathfrak{r}$ and $Z_{N-n+1} = \mathfrak{s}$, can be determined with the aid of (9) and the $(N - n)$th power of P_n.

The distribution of the $f_{\cdot \mathfrak{j}}$ will now be studied, when the sequence of X's is from a chain of order $n - 1(n > 1)$. If the chain is of order $n - 1$ (within the hypothesis H_n), then $p_{\mathfrak{i}j} = p_{\cdot i^* j}$ for $i_1, j = 1, 2, \cdots, a$ and for all a^{n-1} values of i^*.

We define a new sequence of random vectors $W_1 = (X_2, X_3, \cdots, X_n)$, $W_2 = (X_3, X_4, \cdots, X_{n+1})$, $\cdots$, $W_{N-n+1} = (X_{N-n+2}, X_{N-n+3}, \cdots, X_N)$, where each vector can take its value from among the a^{n-1} possible vectors i^*.

The probability $p_{i^*j^*}$ that $W_t = j^*$, given that $W_{t-1} = i^*$, is equal to $p._{i^*j_n}$ for $\mathfrak{I} = \hat{j}$, where $\mathfrak{I} = (i_3, i_4, \cdots, i_r)$, $\hat{j} = (j_2, j_3, \cdots, j_{n-1})$, $i^* = (i_2, \mathfrak{I})$, $j^* = (\hat{j}, j_n)$, and $p_{i^*j^*}$ is zero otherwise. We have that $p_{i^*j^*} = p_{\mathfrak{i}\mathfrak{f}}$ for $j_1 = i_2$ and for all values of i_1, where $\mathfrak{i} = (i_1, i^*)$ and $\mathfrak{j} = (j_1, j^*)$. The sequence W_1, $W_2, \cdots, W_{N-n+1}$ is from a first order chain with transition probability matrix $P_{n-1} = [p_{i^*j^*}]$. This chain has a^{n-1} states.

The n-tuple count $f._{\mathfrak{f}} = g_{\mathfrak{f}}$ in the sequence $X_2, X_3, \cdots, X_N$ can be determined by the 2-tuple count $g_{i^*j^*}$ in the sequence of W's. Also, $g_{i^*j^*} = g_{\mathfrak{f}}$ for $\mathfrak{f} = (j_1, j^*) = (i^*, j_n)$. For the n-tuple count $h_{\mathfrak{f}}$ in the sequence $X_2, X_3, \cdots, X_{N-1}$, we have that $h_{\mathfrak{f}} = g_{\mathfrak{f}} - \delta_{\mathfrak{f}\mathfrak{s}}$. Since $h_{\mathfrak{f}}$ can be determined by the 2-tuple count $h_{i^*j^*}$ in the sequence $W_1, W_2, \cdots, W_{N-n}$ from a first order chain, (3) can be applied to obtain the probability $\prod_{r^*}(h_{i^*j^*}, s')$ that the n-tuple count in the sequence $X_2, X_3, \cdots, X_{N-1}$ will be $h_{\mathfrak{f}}$ and that $W_{N-n} = s'$, given that $W_1 = r^*$. The probability $\prod_{r^*}(g_{\mathfrak{f}}, \mathfrak{s})$ that the n-tuple count in the sequence $X_2, X_3, \cdots, X_N$ will be $g_{\mathfrak{f}}$ and that $Z_{N-n+1} = \mathfrak{s}$, given that $W_1 = r^*$, is simply $\prod_{r^*}(h_{i^*j^*}, s')p._{s's_n}$. Thus,

$$(10) \qquad \prod_{r^*}(g_{\mathfrak{f}}, \mathfrak{s}) = \left[T_{s'}(h_{i^*j^*})\left(\frac{g_{\mathfrak{s}}}{g_{s'}.}\right)\right]\left[\frac{\prod_{i^*} g_{i^*}.!}{\prod_{\mathfrak{f}} g_{\mathfrak{f}}!}\prod_{i^*}\prod_{j} p_{\cdot i^*j}^{g_{i^*j}}\right],$$

where $g_{\mathfrak{f}} = h_{\mathfrak{f}} + \delta_{\mathfrak{f}\mathfrak{s}}$.

The $(n + 1)$-tuple count in the sequence $X_1, X_2, \cdots, X_N$ can be denoted by f_{ii^*j} or $f_{ij'j}$, where $i^* = j'$. Also, $\sum_i f_{ii^*j} = f._{i^*j} = f._{\mathfrak{f}} = g_{\mathfrak{f}} = g_{i^*j}$ and

$$\sum_j g_{i^*j} = g_{i^*}. = \sum_j f._{i^*j} = \sum_j \sum_i f_{ii^*j} = f._{i^*}.,$$

where $\mathfrak{i} = (i, i^*)$ and $\mathfrak{f} = (j', j)$. Thus the probability $\prod_{\mathfrak{r}}(f._{i^*j}, \mathfrak{s})$ that

$$\sum_i f_{ii^*j} = f._{i^*j}$$

and that $Z_{N-n+1} = \mathfrak{s}$ given that $Z_1 = \mathfrak{r}$, is

$$(11) \qquad \prod_{\mathfrak{r}}(f._{i^*j}, \mathfrak{s}) = \left[T_{s'}(h_{i^*j^*})\left(\frac{f._{s's}}{f._{s'}.}\right)\right]\prod_{i^*}\left[\frac{f._{i^*}.!}{\prod_j f._{i^*j}!}\prod_j p_{\cdot i^*j}^{f._{i^*j}}\right]$$

where $\mathfrak{s} = (s', s)$. Therefore, if the chain is of order $n - 1$, the conditional probability of the $(n + 1)$-tuple count $f_{\mathfrak{i}j} = f_{ii^*j}$, given the $f._{i^*j}$ and $\mathfrak{s}$ and $\mathfrak{r}$, is obtained by dividing (9) by (11); i.e., $\prod_{\mathfrak{r}}(f_{ii^*j} \mid f._{i^*j}, \mathfrak{s})$ is equal to

$$(12) \qquad \left[T_{\mathfrak{s}}(f_{\mathfrak{i}\mathfrak{f}})/T_{s'}(h_{i^*j^*})\left(\frac{f._{s's}}{f._{s'}.}\right)\right]\prod_{i^*}\left[\frac{\prod_i f_{ii^*}.!\prod_j f._{i^*j}!}{\prod_i\prod_j f_{ii^*j}!\,f._{i^*}.!}\right]$$

(We can assume, without loss of generality, that the $f._{i^*j}$ and $\mathfrak{s}$ are such that $\prod_{\mathfrak{r}}(f._{i^*j}, \mathfrak{s}) > 0$.) The second factor in (12) is

$$\prod_{i^*} P_{i^*}(f_{ii^*j} \mid f._{i^*j}, f_{ii^*}.) = P(f_{ii^*j} \mid f._{i^*j}, f_{ii^*}.),$$

the product of the probabilities of the cell entries f_{ii^*j} in an $a \times a$ contingency table (for a given $(n - 1)$-tuple i^*), with assigned marginals $f._{i^*j}$ and $f_{ii^*}.$, where in each table the two attributes described by the table are independent; i.e., the joint probability of the cell entries f_{ii^*j} for all a^{n-1} independent contingency tables.

It can be seen that (12) is also the conditional probability

$$\prod\nolimits_{\mathfrak{r}} (f_{ii^*j} \mid f_{\cdot i^*j}, f_{ii^*\cdot})$$

of the f_{ii^*j}, given the $f_{\cdot i^*j}$, $f_{ii^*\cdot}$, and $\mathfrak{r}$; it is also the conditional probability $\prod_{\mathfrak{r}} (f_{ii^*j} \mid f_{i^*j}, \mathfrak{s})$ of the $(n + 1)$-tuple count f_{ii^*j}, given the n-tuple count f_{i^*j}, and $\mathfrak{r}$ and $\mathfrak{s}$.

From (12), we have that

$$\left[T_{\mathfrak{s}}(f_{\mathfrak{i}\mathfrak{t}})/T_{s'}(h_{i^*j^*}) \left(\frac{f_{\cdot s's}}{f_{\cdot s'\cdot}}\right)\right] = \frac{\prod_{\mathfrak{r}}(f_{ii^*j} \mid f_{\cdot i^*j}, f_{ii^*\cdot})}{P(f_{ii^*j} \mid f_{\cdot i^*j}, f_{ii^*\cdot})}. \tag{13}$$

We shall now prove that the statistic (13) converges in probability to unity (thus $T_{\mathfrak{s}}(f_{\mathfrak{i}\mathfrak{t}})$ and $T_{s'}(h_{i^*j^*})(f_{\cdot s's}/f_{\cdot s'\cdot})$ are asymptotically equivalent, and

$$\prod\nolimits_{\mathfrak{r}} (f_{ii^*j} \mid f_{\cdot i^*j}, f_{ii^*\cdot})$$

and $P(f_{ii^*j} \mid f_{\cdot i^*j}, f_{ii^*\cdot})$ are also asymptotically equivalent), if the chain is of order $n - 1$.

If the chain is of order n (a chain of order $n - 1$ is also of order n), we saw earlier that a first order chain could be defined with transition probability matrix $P_n = [p_{\mathfrak{i}\mathfrak{t}}]$, and we shall assume that the asymptotic occupation probabilities $p_{\mathfrak{t}}$ for this first order chain are all positive; i.e., $p_{\mathfrak{t}} > 0$, where $p_{\mathfrak{t}}$ is such that $\sum_{\mathfrak{i}} p_{\mathfrak{i}} p_{\mathfrak{i}\mathfrak{t}} = p_{\mathfrak{t}}$ for all $\mathfrak{t}$. This will be the case if the chain described by P_n is irreducible, (positive) recurrent and aperiodic (see, e.g., [6] and [7]). (If $p_{\mathfrak{t}} = 0$ for some $\mathfrak{t}$, the methods developed in the present paper can be modified in a straightforward manner to obtain analogous results (see [2]).) If the observed sequence is from a chain of order $n - 1$, then the occupation probability $p_{\mathfrak{t}} = p_{j'}p_{\cdot j'j}$, where $\mathfrak{t} = (j', j)$, and $p_{j'}$ is the asymptotic occupation probability for the first order chain with transition probability matrix $P_{n-1} = [p_{i^*j^*}]$. (Lemma 1 in [6] gives a somewhat different, but related, result for chains with denumerably many states.) Since $\prod_{\mathfrak{r}} (f_{\cdot i^*j}, \mathfrak{s}) > 0$, then $p_{\cdot s's} > 0$ where $\mathfrak{s} = (s', s)$, and $f_{\cdot s's}/f_{\cdot s'\cdot}$ will converge in probability to $p_{\cdot s's}$. Thus, it will be necessary to prove only that $T_{\mathfrak{s}}(f_{\mathfrak{i}\mathfrak{t}})/T_{s'}(h_{i^*j^*})$ also converges in probability to $p_{\cdot s's}$.

We have that $T_{\mathfrak{s}}(f_{\mathfrak{i}\mathfrak{t}})$ is the $(\mathfrak{s}\mathfrak{r})$-th cofactor of the matrix $\hat{M}_n$, and $T_{s'}(h_{i^*j^*})$ is the $(s'r^*)$th cofactor of the matrix $[\delta_{i^*j^*} - h_{i^*j^*}/h_{i^*\cdot}] = \hat{M}_{n-1}$. These matrices will converge in probability to the $a^n \times a^n$ matrix $M_n = [\delta_{\mathfrak{i}\mathfrak{t}} - p_{\mathfrak{i}\mathfrak{t}}]$ and the $a^{n-1} \times a^{n-1}$ matrix $M_{n-1} = [\delta_{i^*j^*} - p_{i^*j^*}]$ respectively, and $T_{\mathfrak{s}}(f_{\mathfrak{i}\mathfrak{t}})$ and

$$T_{s'}(h_{i^*j^*})$$

will converge in probability to the $(\mathfrak{s}\mathfrak{r})$th cofactor and the $(s'r^*)$th cofactor of M_n and M_{n-1} respectively. Since in each matrix the sum of the entries in each row is zero, all the cofactors in row $\mathfrak{s}$ of M_n are all equal to the $(\mathfrak{s}\mathfrak{s})$th cofactor $| M_{n,\mathfrak{s}} |$ of M_n, and the cofactors in row s' of M_{n-1} are all equal to the $(s's')$th cofactor $|M_{n-1,s'}|$ of M_{n-1}. Therefore, it will be necessary to prove only that $|M_{n,\mathfrak{s}}| = |M_{n-1,s'}| p_{\cdot s's}$.

We have that $M_n = I_n - P_n$, where I_n is the $a^n \times a^n$ identity matrix and P_n is the transition probability matrix for the first order chain for the Z's. We order the states $\mathfrak{i} = (i, i^*)$ of this chain recursively as follows: $(1, i^*) = 1i^*$, $(2, i^*) = 2i^*, \cdots, (a, i^*) = ai^*$, for $i^* = 1, 2, \cdots, a^{n-1}$, obtaining a numbering of $\mathfrak{i}$ from 1 to a^n. Since $p_{\mathfrak{i}\mathfrak{j}} = p_{\cdot i^*j}$ for $j' = i^*$, and it is zero otherwise, we have that

$$P_n = \begin{bmatrix} P_{\cdot 1} & P_{\cdot 2} & \cdots & P_{\cdot a} \\ P_{\cdot 1} & P_{\cdot 2} & \cdots & P_{\cdot a} \\ \vdots & \vdots & \vdots & \vdots \\ P_{\cdot 1} & P_{\cdot 2} & \cdots & P_{\cdot a} \end{bmatrix}, \tag{14}$$

where $P_{\cdot j}$ is the $a^{n-1} \times a^{n-1}$ matrix $[{}_jp_{i^*j^*}]$ with ${}_jp_{i^*j^*} = p_{i^*j^*}$ for $i_2 = j$, and ${}_jp_{i^*j^*} = 0$ for $i_2 \neq j$, for all j^* and j. Hence P_n consists of a block columns

$$[P_{\cdot j}, P_{\cdot j}, \cdots, P_{\cdot j}]' \qquad \text{for } j = 1, 2, \cdots, a;$$

it also consists of a block rows $[P_{\cdot 1}, P_{\cdot 2}, \cdots, P_{\cdot a}]$. We have that $|M_{n,\mathfrak{s}}|$ is the determinant of the $(a^n - 1) \times (a^n - 1)$ matrix $M_{n,\mathfrak{s}}$ obtained by deleting the row and column relating to state $\mathfrak{s} = (s_1, s^*)$, the s_1s^*th state in M_n; i.e., column s^* within the s_1th block column $[P_{\cdot s_1}, P_{\cdot s_1}, \cdots, P_{\cdot s_1}]'$ and also row s^* within the s_1th block row $[P_{\cdot 1}, P_{\cdot 2}, \cdots, P_{\cdot a}]$, and the corresponding column and row in the identity matrix. Let $P''_{\cdot s_1}$ be the $a^{n-1} \times a^{n-1}$ matrix obtained by replacing the s^*th column in $P_{\cdot s_1}$ by a column of zeros. By some elementary transformations of the matrix $M_{n,\mathfrak{s}}$, we find that $|M_{n,\mathfrak{s}}|$ is equal to the determinant $|\tilde{M}|$ of the $a^{n-1} \times a^{n-1}$ matrix $\tilde{M} = I_{n-1} - [\sum_{j \neq s_1} P_{\cdot j} + P''_{\cdot s_1}]$. Thus, it is necessary to prove only that $|\tilde{M}| = |M_{n-1,s'}| p_{\cdot s's}$. It can be seen that the only distinction between $\tilde{M}$ and M_{r-1} is that the term $p_{\cdot s's}$ appearing in row s' and column s^* of M_{n-1} is replaced by a zero in $\tilde{M}$. (If $s^* = s'$, then the term $1 - p_{\cdot s's}$ is replaced by 1.) Thus, each cofactor in the s'th row of $\tilde{M}$ is equal to $|M_{n-1,s'}|$. Since the sum of the entries in row s' of $\tilde{M}$ is

$$1 - \sum_{j^* \neq s^*} p_{s'j^*} = p_{\cdot s's},$$

we have that $p_{\cdot s's} |M_{n-1,s'}| = |\tilde{M}|$. Q.E.D.

This result, concerning the asymptotic equivalence, under the assumption H_{n-1}, of $\prod_{\mathfrak{r}} (f_{ii^*j} | f_{\cdot i^*j}, f_{ii^*\cdot})$ and $P(f_{ii^*j} | f_{\cdot i^*j}, f_{ii^*\cdot})$, implies that the null hypothesis H_{n-1} can be tested, within H_n, by any asymptotic test of contingency in the a^{n-1} ordinary $a \times a$ contingency tables with cell entries f_{ii^*j} and with assigned marginals $f_{\cdot i^*j}$ and $f_{ii^*\cdot}$.

4. The m-tuple and the n-tuple counts ($m > n$). Let ι be the $(m - n)$-tuple $(i_1, i_2, \cdots, i_{m-n})$ and I the $(n - 1)$-tuple $(i_{m-n+1}, i_{m-n+2}, \cdots, i_{m-1})$. Denote the m-tuple count in the sequence $X_1, X_2, \cdots, X_N$ by $f_{\iota \mathrm{I} j}$. Then $\sum_j f_{\iota \mathrm{I} j} = f_{\iota \mathrm{I} \cdot}$ is the $(m - 1)$-tuple count for the sequence $X_1, X_2, \cdots, X_{N-1}$, and $\sum_\iota f_{\iota \mathrm{I} j} =$

$f_{\cdot \mathrm{I} j}$ is the n-tuple count for the sequence $X_{m-n+1}, X_{m-n+2}, \cdots, X_N$. Let

$$\prod\nolimits_{\mathfrak{r}} (f_{\iota \mathrm{I} j} \mid f_{\cdot \mathrm{I} j}, \mathfrak{s})$$

be the probability that the m-tuple count will be $f_{\iota \mathrm{I} j}$, given that $\sum_\iota f_{\iota \mathrm{I} j} = f_{\cdot \mathrm{I} j}$ and $(X_1, X_2, \cdots, X_{m-1}) = \mathfrak{r}$ and $(X_{N-m+2}, X_{N-m+3}, \cdots, X_N) = \mathfrak{s}$. If $m = n + 1$, the results in Section 3 give the formula for this probability. If $m = n + 2$, then $\prod_{\mathfrak{r}} (f_{\iota \mathrm{I} j} \mid f_{\cdot \mathrm{I} j}, \mathfrak{s})$ is equal to

$$[\prod\nolimits_{\mathfrak{r}} (f_{i \mathrm{I} j} \mid f_{\cdot i_2 \mathrm{I} j}, \mathfrak{s}) \prod\nolimits_{\mathfrak{r}} (f_{\cdot i_2 \mathrm{I} j} \mid f_{\cdot \mathrm{I} j}, \mathfrak{s})]:$$

the first factor is the probability that the m-tuple count will be $f_{\iota \mathrm{I} j}$, given that $\sum_{i_1} f_{i \mathrm{I} j} = f_{\cdot i_2 \mathrm{I} j}$ and $(X_1, X_2, \cdots, X_{m-1}) = \mathfrak{r}$ and $(X_{N-m+2}, \cdots, X_N) = \mathfrak{s}$; the second factor is the probability that the $(m - 1)$-tuple count in the sequence $X_2, X_3, \cdots, X_N$ will be $f_{\cdot i_2 \mathrm{I} j}$, given that $\sum_{i_2} f_{\cdot i_2 \mathrm{I} j} = f_{\cdot \mathrm{I} j}$ and

$$(X_1, X_2, \cdots, X_{m-1}) = \mathfrak{r}$$

and $(X_{N-m+2}, \cdots, X_N) = \mathfrak{s}$. If the chain is of order n, the results in Section 3 indicate that the first factor is asymptotically equivalent to

$$\prod_{i_2 \mathrm{I}} \{\prod_{i_1} f_{i \mathrm{I} \cdot}! \prod_j f_{\cdot i_2 \mathrm{I} j}! / \prod_{i_1} \prod_j f_{i \mathrm{I} j}! f_{\cdot i_2 \mathrm{I} \cdot}!\}; \tag{15}$$

if the chain is of order $n - 1$, the second factor is asymptotically equivalent to

$$\prod_{\mathrm{I}} \{\prod_{i_2} f_{\cdot i_2 \mathrm{I}}! \prod_j f_{\cdot \mathrm{I} j}! / \prod_{i_2} \prod_j f_{\cdot i_2 \mathrm{I} j}! f_{\cdot \mathrm{I} \cdot}!\}, \tag{16}$$

since it can be shown from the derivation of (12) that $\prod_{\mathfrak{r}} (f_{\cdot i_2 \mathrm{I} j} \mid f_{\cdot \mathrm{I} j}, \mathfrak{s})$ is asymptotically equivalent to $\prod_{r^*} (f_{\cdot i_2 \mathrm{I} j} \mid f_{\cdot \mathrm{I} j}, s^*)$, where $\mathfrak{s} = (s_1, s^*)$ and $\mathfrak{r} = (r_1, r^*)$. Thus, for chains of order $n - 1$, $\prod_{\mathfrak{r}} (f_{i \mathrm{I} j} \mid f_{\cdot \mathrm{I} j}, \mathfrak{s})$ for $m = n + 2$ is asymptotically equivalent to the product of (15) and (16); viz.

$$\prod_{\mathrm{I}} \{\prod_{\iota} f_{\iota \mathrm{I} \cdot}! \prod_j f_{\cdot \mathrm{I} j}! / \prod_{\iota j} f_{\iota \mathrm{I} j}! f_{\cdot \mathrm{I} \cdot}!\}. \tag{17}$$

In the general case where $m > n$, by repeated application of the preceding results for $m = n + 1$ and $n + 2$, we find that, for chains of order $n - 1$,

$$\prod\nolimits_{\mathfrak{r}} (f_{\iota \mathrm{I} j} \mid f_{\cdot \mathrm{I} j}, \mathfrak{s})$$

is asymptotically equivalent to (17), the product of the probabilities

$$P_{\mathrm{I}}(f_{\iota \mathrm{I} j} \mid f_{\cdot \mathrm{I} j}, f_{\iota \mathrm{I} \cdot})$$

of the cell entries $f_{\iota \mathrm{I} j}$ in a contingency table (for a given $(n - 1)$-tuple I) consisting of a columns ($j = 1, 2, \cdots, a$) and a^{m-n} rows (the a^{m-n} values of ι), with assigned marginals $f_{\cdot \mathrm{I} j}$ and $f_{\iota \mathrm{I} \cdot}$, when in each table the attributes described by the table (there are a^{n-1} tables) are independent. This result implies that the null hypothesis H_{n-1} can be tested, within the hypothesis H_{m-1}, by any asymptotic test of contingency in the a^{n-1} ordinary $a \times a^{m-n}$ contingency table with cell entries $f_{\iota \mathrm{I} j}$ and with assigned marginals $f_{\cdot \mathrm{I} j}$ and $f_{\iota \mathrm{I} \cdot}$. These tests will have $a^{n-1}(a - 1)(a^{m-n} - 1) = (a^m - a^n)(a - 1)/a$ degrees of freedom (see [1] and [8]).

5. The circular counts. It was shown in [4] that, for zero order chains, the probability $\prod (\bar{f}_{i_1 \cdots i_m} \mid \bar{f}_{i_1 \cdots i_n})$ of a specified m-tuple count $\bar{f}_{i_1 \cdots i_m}$ in a circular sequence, given the n-tuple count $\bar{f}_{i_1 \cdots i_n} (n < m)$, is

$$(18) \qquad C([\bar{f}_{i_1 \cdots i_m}]) \prod \bar{f}_{i_1 \cdots i_n} ! / C([\bar{f}_{i_1 \cdots i_n}]) \prod \bar{f}_{i_1 \cdots i_m} !$$

if $n > 1$, or

$$C([\bar{f}_{i_1 \cdots i_m}]) \prod \bar{f}_i ! / (N - 1)! \prod \bar{f}_{i_1 \cdots i_m} ! \qquad \text{if } n = 1,$$

where $F^* = [\bar{f}_{i_1 \cdots i_n}]$ is the incidence matrix of the graph (see [4]), and $C(M)$ is defined in Section 2 here; (18) is valid for chains of order $n - 1$ or less. In the special case $n = 1$ it was proved in [4] that the statistic (18) is asymptotically equivalent to $\prod \bar{f}_{i_1 \cdots i_{m-1}} ! \prod \bar{f}_i ! / N! \prod \bar{f}_{i_1 \cdots i_m} !$, the probability

$$P(\bar{f}_{i_1 \cdots i_m} \mid \bar{f}_{i_1 \cdots i_{m-1}}, \bar{f}_i)$$

of the cell entries $\bar{f}_{i_1 \cdots i_m}$ in a contingency table with assigned marginals $\bar{f}_{i_1 \cdots i_{m-1}}$ and $\bar{f}_i$. This result for the special case $n = 1$ will now be generalized to the case $n \geqq 1$.

Let us first consider the case where $m = n + 1$. We can write

$$(19) \qquad F_n^* = D(\bar{f}_{i_1 \cdots i_n}) \left[\delta_{\mathfrak{i}\mathfrak{k}} - \frac{\bar{f}_{\mathfrak{i}\mathfrak{k}}}{\bar{f}_{\mathfrak{i}\cdot}} \right],$$

where $\mathfrak{i} = (i_1, i_2, \cdots, i_n)$, $\bar{f}_{\mathfrak{i}\mathfrak{k}}$ is defined for circular sequences in the same way as $f_{\mathfrak{i}\mathfrak{k}}$ was defined in Section 3, and $D(\bar{f}_{i_1 \cdots i_n})$ is the $a^n \times a^n$ diagonal matrix where the entry in row $\mathfrak{i}$ is $\bar{f}_{\mathfrak{i}}$. We shall assume that no row or column consists wholly of zeros. The common value $|F^*_{n,\mathfrak{i}}|$ of the cofactors of F_n^* can be obtained by determining the $(\mathfrak{i}\mathfrak{i})$th cofactor of $D(\bar{f}_{\mathfrak{i}})$, which is $\prod_{\mathfrak{k} \neq \mathfrak{i}} \bar{f}_{\mathfrak{k}}$, and also the $(\mathfrak{i}\mathfrak{i})$th cofactor of $[\delta_{\mathfrak{i}\mathfrak{k}} - \bar{f}_{\mathfrak{i}\mathfrak{k}}/\bar{f}_{\mathfrak{i}\cdot}]$, which converges in probability to the $(\mathfrak{i}\mathfrak{i})$th cofactor $|M_{n,\mathfrak{i}}|$ of M_n. From the results in Section 3, for the case where the chain is of order $n - 1$, we see that $|M_{n,\mathfrak{i}}| = |M_{n-1,i'}| p_{\cdot i' i}$, where $\mathfrak{i} = (i', i)$. Thus $|F^*_{n,\mathfrak{i}}|$ is asymptotically equivalent to $\prod_{\mathfrak{k} \neq \mathfrak{i}} \bar{f}_{\mathfrak{k}} |M_{n-1,i'}| \bar{f}_{i'i}/\bar{f}_{i'}$. Also, the $(i'i')$th cofactor $|F^*_{n-1,i'}|$ of F^*_{n-1} is asymptotically equivalent to $\prod_{j^* \neq i'} \bar{f}_{j^*} |M_{n-1,i'}|$. Hence, $|F^*_{n,\mathfrak{i}}|/|F^*_{n-1,i'}|$ is asymptotically equivalent to $\prod_{\mathfrak{k}} \bar{f}_{\mathfrak{k}} / \prod_{j^*} \bar{f}_{j^*}$, and $C(F_n^*)/C(F^*_{n-1})$ is asymptotically equivalent to

$$\prod_{\mathfrak{k}} \bar{f}_{\mathfrak{k}}! / \prod_{j^*} \bar{f}_{j^*}!.$$

Therefore, if the chain is of order $n - 1$, (18) for $m = n + 1$ is asymptotically equivalent to

$$(20) \qquad \frac{\prod_{\mathfrak{k}} \bar{f}_{\mathfrak{k}} ! \prod_{\mathfrak{k}} \bar{f}_{\mathfrak{k}} !}{\prod_{j^*} \bar{f}_{j^*} ! \prod_{\mathfrak{k}} \prod_j \bar{f}_{\mathfrak{k}j} !} = \prod_{j^*} \left[\frac{\prod_{j_1} \bar{f}_{j_1 j^*} ! \prod_j \bar{f}_{j^* j} !}{\prod_{j_1} \prod_j \bar{f}_{j_1 j^* j} ! \bar{f}_{j^*} !} \right],$$

the product of the probabilities $P_{j^*}(\bar{f}_{j_1 j^* j} \mid \bar{f}_{j_1 j^*}, \bar{f}_{j^* j})$ of the observed cell entries $\bar{f}_{j_1 j^* j}$ in an ordinary $a \times a$ contingency table (for a given $(n - 1)$-tuple

j^*), with assigned marginals $\bar{f}_{j_1 j^*}$ and $\bar{f}_{j^* j}$ (we have that $\sum_j \bar{f}_{j_1 j^* j} = \bar{f}_{j_1 j^* \cdot} = \sum_j \bar{f}_{j j_1 j^*} = \bar{f}_{\cdot j_1 j^*} = \bar{f}_{j_1 j^*}$), where in each table the two attributes described by the table are independent.

This result, concerning asymptotic equivalence in the special case $m = n + 1$, can be applied repeatedly to obtain a general result for the case $m > n$, as was done in Section 4. Thus, if the chain is of order $n - 1$, the statistic

$$\overline{\prod} (\bar{f}_{i_1 \cdots i_m} \mid \bar{f}_{i_1 \cdots i_n})$$

is asymptotically equivalent to

$$\frac{\prod \bar{f}_{i_1 \cdots i_{m-1}}! \prod \bar{f}_{i_1 \cdots i_n}!}{\prod \bar{f}_{i_1 \cdots i_m} \prod \bar{f}_{i_1 \cdots i_{n-1}}!} = \prod_{\mathrm{I}} \left[\frac{\prod_\iota \bar{f}_{\iota \mathrm{I}}! \prod_i \bar{f}_{\mathrm{I} i}!}{\prod_\iota \prod_i \bar{f}_{\iota \mathrm{I} i}! \bar{f}_{\mathrm{I}}!} \right],$$

where $\iota = (i_1, i_2, \cdots, i_{m-n})$, $\mathrm{I} = (i_{m-n+1}, i_{m-n+2}, \cdots, i_{m-1})$, $i = i_m$. Hence, any asymptotic test of contingency in the a^{n-1} ordinary $a \times a^{m-n}$ contingency tables (a table for each $(n - 1)$-tuple I) with cell entries $\bar{f}_{i_1 \cdots i_m}$, and with assigned marginals $\bar{f}_{i_1 \cdots i_{m-1}}$ and $f_{i_{m-n+1} \cdots i_m}$ (i.e., $\bar{f}_{i_1 \cdots i_n}$), can be used to test the null hypothesis H_{n-1} within H_{m-1}. The degrees of freedom are as in Section 4.

The reader will note that, in this and the preceding sections, each m-tuple was "split" into an $(m - n)$-tuple ι, an $(n - 1)$-tuple I, and a 1-tuple i; thus obtaining a^{n-1} contingency tables, each $a \times a^{m-n}$. It is possible to split each m-tuple into an $(m - n - r)$-tuple, an $(n - 1)$-tuple, and a $(1 + r)$-tuple $(0 \leqq r \leqq m - n - 1)$; thus obtaining a^{n-1} contingency tables, each $a^{1+r} \times a^{m-n-r}$. For $r = m - n - 1$, the m-tuple is split into a 1-tuple, an $(n - 1)$-tuple, and a $(m - n)$-tuple; the a^{n-1} contingency tables obtained will differ in general from the a^{n-1} tables obtained for $r = 0$. However, for circular sequences, the product of the likelihood ratios (for testing independence in each table) for the a^{n-1} tables obtained when $r = m - n - 1$ will be equal to the corresponding product for the tables obtained when $r = 0$. For linear sequences, the corresponding products when $r = m - n - 1$ and $r = 0$ will be asymptotically equivalent, under H_{n-1}. Both these products are asymptotically equivalent to the likelihood ratio for testing H_{n-1} within H_{m-1}. Similar remarks could be made about other statistics (e.g., the χ^2 statistic) used to test independence in each table. If the a^{n-1} separate tables were of interest, the choice between $r = 0$ or $m - n - 1$ would depend on the alternate hypotheses within H_{m-1} that were in mind.

For $0 \leqq r \leqq m - n - 1$, it can be shown that the asymptotic mean value of the product of the likelihood ratios (when normed in the usual way) is equal to $a^{n-1}(a^{m-n-r} - 1)(a^{1+r} - 1)$, under H_{n-1}. This statistic is not equivalent to the likelihood ratio for testing H_{n-1} within H_{m-1}, unless $r = 0$ or $m - n - 1$. Also, the asymptotic distribution, given H_{n-1}, of this statistic is not χ^2 unless $r = 0$ or $m - n - 1$. For $0 < r < m - n - 1$, the asymptotic distribution, given H_{n-1}, of this statistic is that of a weighted sum (with unequal weights) of χ^2 variates; in this case, the analysis of the a^{n-1} separate contingency tables is not in general as simple and straightforward as when $r = 0$ or $m - n - 1$, since the usual methods of analysis of contingency tables cannot be applied to this case. This case will be discussed more fully in a later publication by the present author.

6. The exact probability formulas. An illustration will now be presented to indicate the difference between (12), (18), and the formula suggested in [4] for the probability of the specified m-tuple count, given the n-tuple count, in a linear sequence. Consider the special case $a = 2$, $n = 1$, $m = 2$, $N = 5$, and $f_{11} = 0, f_{12} = 2, f_{21} = 1, f_{22} = 1$. Thus $f_{1\cdot} = 2, f_{2\cdot} = 2, f_{\cdot 1} = 1, f_{\cdot 2} = 3$. From (1), we see that $r = 1$, $s = 2$. From (6), the probability of the specified 2-tuple count f_{ij}, given the 1-tuple count $f_{\cdot j}$ and $f_{i\cdot}$ and r, is 2/3. The circularized 2-tuple frequencies are $\bar{f}_{11} = 0$, $\bar{f}_{12} = 2$, $\bar{f}_{21} = 2$, $\bar{f}_{22} = 1$. From (18), the probability of the $\bar{f}_{ij}$, given the $\bar{f}_i$, is 1/2. Using the approach suggested in [4] of applying (18) to the augmented circularized sequence, the probability of the specified 2-tuple count f_{ij}, given the 1-tuple count f_i, is 1/5; this approach yields a correct answer only if the chain is stationary. By listing all possible sequences for $a = 2$ and $N = 5$, the reader will see why different numerical results are obtained for the different probabilities.

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] Harald Cramér, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, 1946.

[4] R. Dawson, and I. J. Good, "Exact Markov probabilities from oriented linear graphs," *Ann. Math. Stat.* Vol. 28 (1957), pp. 838–946.

[5] N. G. de Bruijn and T. van Aardenne-Ehrenfest, "Circuits and trees in oriented linear graphs," *Simon Stevin*, Vol. 4 (1950–1), pp. 203–17.

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

[7] William Feller, *An Introduction to Probability Theory and its Applications*, John Wiley and Sons, Inc., New York, 1950.

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

[9] Leo A. Goodman, "Simplified runs tests and likelihood ratio tests for Markoff chains," *Biometrika*, Vol. 45 (1958).

[10] P. G. Hoel, "A test for Markoff chains," *Biometrika*, Vol. 41 (1954), pp. 430–3.

[11] C. C. MacDuffee, *The Theory of Matrices*, Chelsea Publishing Co., New York, 1946.

[12] A. M. Mood, *Introduction to the Theory of Statistics*, McGraw-Hill Book Company, New York, 1950.

[13] P. Whittle, "Some distribution and moment formulae for the Markov chain," *J. Roy. Stat. Soc., Series B*, Vol. 17 (1955), pp. 235–42.