EXACT MARKOV PROBABILITIES FROM ORIENTED LINEAR GRAPHS

By Reed Dawson and I. J. Good

Silver Spring, Maryland and Cheltenham, England

0. Summary. Using a theorem due to de Bruijn, van Aardenne-Ehrenfest, C. A. B. Smith and Tutte concerning the number of circuits in oriented linear graphs, an expression is found for the probability of a specified frequency count of m-tuples in a circular sequence where the n-tuple ($n < m$) count is given. The corresponding result for linear sequences can be deduced—see [14]. The result is valid for stationary Markovity of any order up to and including the ($n - 1$)-st. A method of deriving asymptotic distributions is indicated, and a few additional observations made concerning the distribution of pairs in a circular array.

1. Introduction. In studying runs, W. L. Stevens [10] considered the distribution of pairs of successive digits when n zeros and $N - n$ ones are randomly permuted about an oriented circle. He found the probability of A occurrences of 0 followed by 0, B occurrences of 01, C of 10 and D of 11, for any A, B, C, D subject to

$$A + B = A + C = n \quad \text{and} \quad C + D = B + D = N - n.$$

(Stevens' result has been generalized by Mood [9] and Whittle [13].) By using a combinatorial theorem due to de Bruijn, van Aardenne-Ehrenfest, C. A. B. Smith and Tutte (hereinafter known by initials as the BEST theorem), Stevens' result can be generalized to an expression for the probability of a specified m-tuple frequency count in a (circular) sequence of given n-tuple count, where the alphabet may be of any finite size. The BEST theorem was first stated, implicitly, as a "note added in proof" on page 217 of de Bruijn and Ehrenfest [2], and is largely based on Tutte [11] and Tutte and Smith [12].

2. The BEST theorem. Given any $u \times u$ matrix of nonnegative integers there corresponds an oriented linear graph, with vertices 1, 2, . . . , u, such that the number of oriented paths (edges) leading from vertex r to vertex s equals the matrix element in row r and column s. 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 "simple" (in the sense of Tutte [11] or a "T-graph" in the notation of de Bruijn and Ehrenfest [2]) if the number of edges leading into each vertex equals the number leading out, or, in terms of the incidence matrix, if each row has the same sum as the corresponding column. A (complete) circuit in such a graph is defined as a unicursal path passing exactly once through each edge (in the right direction). The BEST theorem gives the number of distinct circuits when all edges are regarded as distinguishable.

Received January 9, 1956; revised January 10, 1957.

Let $M = [m_{ij}]$ be the incidence matrix of some simple oriented linear graph and let

$$m_i = \sum_j m_{ij} = \sum_j m_{ji}$$

be the sum of the ith row (and of the ith column). Let $M' = [m'_{ij}]$ be the $u' \times u'$ matrix formed from M by deleting every row and column consisting wholly of zeros (in effect eliminating vertices lying on no edges). Then

$$\sum_j m'_{ij} = \sum_j m'_{ji} = m'_i, \quad \text{say, where} \quad m'_i > 0, \qquad i = 1, \cdots, u'.$$

Let $M^* = [m'_i \delta_{ij}] - M'$, i.e., the matrix of entries $m^*_{ij} = -m'_{ij}$ for $i \neq j$ and $m^*_{ii} = m'_i - m'_{ii}$. 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 of circuits, $C(M)$, in a simple oriented linear graph with incidence matrix M is

$$C(M) = \| M^* \| \cdot \prod_{i=1}^{u'} (m'_i - 1)!. \tag{1}$$

3. Distribution of pairs. In applying graph theory to circular arrangements of letters the several occurrences of one letter will be regarded as distinguishable, so that there will be $(N - 1)!$ possible circular sequences having a given frequency count of letters; for the present these sequences will be assumed equally probable. Let the frequencies of the individual letters be $f_1, f_2, \cdots, f_t$ with $\sum f_i = N$; the probability $P(F)$ that pairs of successive letters (i, j) will have the matrix of frequencies $F = [f_{ij}]$, where

$$\sum_i f_{ij} = f_j, \qquad \sum_j f_{ij} = f_i, \qquad \sum_{i,j} f_{ij} = N,$$

may be determined as follows. Imagine an oriented linear graph consisting of one vertex for each of the t letters in the alphabet, together with f_{ij} distinguishable oriented edges from vertex i to vertex j $(i, j = 1, \cdots, t)$; then the number of circuits is $C(F)$ (Eq. (1)). Although each circuit corresponds to a circular sequence of letters with the pair-frequencies F, the enumeration of the circular sequences requires distinguishing the f_i uses of the vertex i and then identifying the f_{ij} edges leading from vertex i to vertex j. Hence the total number of circular sequences with the pair-frequencies F is

$$(\prod f_i! / \prod f_{ij}!) C(F),$$

and the probability of F is

$$P(F) = \frac{\prod f_i!}{(N - 1)! \prod f_{ij}!} C(F). \tag{2}$$

(See [14] for a proof that Whittle's formula (8) in [13] is essentially equivalent to ours and can therefore also be derived from the BEST theorem. We independently noticed this fact, but would not have done so had Goodman not first drawn our attention to the existence of Whittle's paper. We had prepared a second appendix, in which Whittle's formula was deduced from the BEST theorem, but decided not to include it so as to avoid overlap with [14].)

4. Extension to n-tuples. The general formula for the number of circular sequences of N given letters having prescribed n-tuple frequencies is obtained by considering the oriented linear graph with vertices corresponding to $(n-1)$-tuples and edges corresponding to n-tuples. Let $f_{i_1 \cdots i_n}$ edges (in accordance with the prescription) run from the vertex $(i_1 \cdots i_{n-1})$ to the vertex $(i_2 \cdots i_n)$. By the BEST theorem the number of circuits in this graph is $C(F)$ where $F = [f_{i_1 \cdots i_n}]$ is the incidence matrix of the graph. (For an example of this notation, combined with the asterisk and cofactor notation of Sec. 2, see the Appendix.) Each circuit corresponds to $\prod f_i!$ circular sequences with the correct n-tuple frequencies except that the $f_{i_1 \cdots i_n}$ n-tuples $(i_1 \ldots i_n)$ are given a separate identity. Hence the total number of circular permutations (of N given letters) having prescribed n-tuple frequencies is

$$(3) \quad C(F) \cdot \prod f_i! / \prod (f_{i_1 \cdots i_n})!, \qquad \text{if } n > 1, \qquad \text{or } (N-1)!, \qquad \text{if } n = 1.$$

If all $(N-1)!$ circular sequences are equally likely, the probability of specified m-tuple frequencies, given the n-tuple frequencies $(n < m)$, is simply the ratio of the corresponding numbers of circular sequences satisfying the requirement, viz.,

$$(4) \quad \begin{aligned} &C([f_{i_1 \cdots i_m}]) \cdot \prod f_{i_1 \cdots i_n}! / C([f_{i_1 \cdots i_n}]) \cdot \prod f_{i_1 \cdots i_m}!, && \text{if } n > 1; \\ &C([f_{i_1 \cdots i_m}]) \cdot \prod f_i! / (N-1)! \prod f_{i_1 \cdots i_m}!, && \text{if } n = 1. \end{aligned}$$

It should be noted that the m-tuple frequencies imply unique n-tuple frequencies; for any other given n-tuple frequencies the probabilities (4) must be replaced by zero. The logarithm of the ratio (4) may be compared with the statistic $\nabla L_m - \nabla L_n$ in paragraph 8 of Good [4].

5. Linear sequences. In most applications, such as the analysis of Markov processes, the sequence is linear rather than circular; but the circular model is mathematically simpler. For example, in a linear sequence it is not always true that the n-tuple frequencies of the linear sequence determine the $(n-1)$-tuple frequencies; nor do the n-tuple frequencies necessarily determine the n-tuple frequencies of the circular sequence obtained by regarding the first letter of the linear sequence as the successor of the last letter. The linear sequences ABCAB and BCABC share the same triples (ABC, BCA, CAB) but differ in pair frequencies, single-letter frequencies, and in the triple frequencies of the corresponding circular sequences. [However, the n-tuple frequencies of a linear sequence do determine the $(n-1)$-tuple frequencies unless

$$(L) \ \sum_{i_n} f_{i_1 \cdots i_{n-1} i_n} = \sum_{i_n} f_{i_n i_1 \cdots i_{n-1}}$$

for all $i_1, \cdots, i_{n-1}$. If condition (L) is not satisfied, then the first and last $(n-1)$-tuples (and hence the complete $(n-1)$-tuple frequencies as well as the circular n-tuple frequencies) can be determined.] One way of treating linear sequences (for another see Whittle [13]) is to regard a linear sequence of N letters as consisting of $N + 1$ characters, the new character being a blank placed at the end. Then the new n-tuple frequencies (including the n-tuple ending with the blank) will determine uniquely the n-tuple frequencies of the corresponding circular sequence; and, conversely, the n-tuple frequencies in any circular sequence containing a blank will determine uniquely the n-tuple frequencies of the corresponding linear sequence formed by cutting the circular sequence right after the blank. Hence, in a linear sequence ending with a blank, we may define the probability of specified m-tuple frequencies, given the n-tuple frequencies, as the value found by circularizing the sequence (retaining the blank) and applying formula (4).

6. Negligible Markovity. The probability $P(a_1a_2 \cdots a_N)$ of a specified linear sequence $a_1, a_2, \cdots, a_N$ of N distinguishable letters under a Markov process of order $n - 1$ or less is

$$\prod f_i\,!\,P(a_1\cdots a_n)P(a_{n+1} \mid a_2\cdots a_n)\cdots P(a_N \mid a_{N-n+1}\cdots a_{N-1})$$

$$= \prod f_1\,!\,\frac{P(a_1\cdots a_n)P(a_2\cdots a_{n+1})\cdots P(a_{N-n+1}\cdots a_N)}{P(a_2\cdots a_n)P(a_3\cdots a_{n+1})\cdots P(a_{N-n+1}\cdots a_{N-1})}$$

(to be interpreted as zero if any factor in the numerator is zero), or

$$P(a_1\cdots a_N) = \frac{\prod f_i\,!\,\prod_{i_1\cdots i_n} P(i_1\cdots i_n)^{f_{i_1\cdots i_n}}}{\prod_{i_1\cdots i_{n-1}} P(i_1\cdots i_{n-1})^{f_{i_1\cdots i_{n-1}}}}. \tag{5}$$

where 0^0 is interpreted as 1. Now if the n-tuple probabilities of the alphabet and the n-tuple frequencies of the augmented sequence $a_1, \cdots, a_N, b$ (where b is the terminal blank) are given, then the $(n - 1)$-tuple frequencies and probabilities, and also the single-letter frequencies, are determined. Under these conditions the probability (5), being also determined, is mathematically independent of any further knowledge of the $(n + 1)$-tuple frequencies, so that the probability of a specified $(n + 1)$-tuple frequency-count is proportional to the number of ways in which this count can occur. It follows that, in applying formula (4) to linear sequences, the assumption that all permutations are equally likely may be replaced by the more general assumption of Markovity of order $n - 1$ or less without affecting the probability of the specified m-tuple frequencies. If a circular sequence is defined as a linear sequence with the ends joined (the most natural definition in interpreting the circular sequence as a Markov process), the total probability of the circular sequence $(a_1, a_2, \cdots, a_N)$ is the sum of the probability (5) over all N cyclic permutations of the linear sequence $a_1, a_2, \cdots, a_N$. Then by the same argument the probability of a circular m-tuple frequency-count, as given by formula (4), is valid for all orders of Markovity up to and including the $(n - 1)$-st.

7. Asymptotic relationships. The arguments are based on the following lemma.

LEMMA.

Hypotheses. (a) *An experiment (with a parameter* N*) has, for each value of* N *(positive integers tending to infinity) a finite set* $F^N = \{F_i^N\}$ *of possible outcomes.* P_N *(or simply* P*) and* P'_N *(or simply* P'*) are two probability measures over* F^N.

(b) $P'(F_i^N)/P(F_i^N)$ *converges in the probability* P *to unity in the sense that for all* $\eta > 0$ *and* $\delta > 0$ *there exists an* N_0 *such that, for all* $N > N_0$,

$$\text{Prob}_P \ \{P'(F_i^N)/P(F_i^N) \ \varepsilon \ I_\delta\} > 1 - \eta, \tag{6}$$

where I_δ *is the interval* $(1 - \delta, 1 + \delta)$, *and* $P'(F_i^N)/P(F_i^N)$ *is regarded as a statistic whose distribution is determined by* P.

(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.

CONCLUSION. *The distribution function* Φ'_N *of* $S(F_i^N)$ *under* P' *converges to the same limiting distribution* Φ.

PROOF. *Let* $\Lambda_{N,\lambda}$ (or simply Λ) be the set of all indices i for which $S(F_i^N) \leqq \lambda$, where λ is any real number. Let Δ be the set of all indices i for which $P'(F_i^N/P(F_i^N) \ \varepsilon \ I_\delta$, for any fixed arbitrary δ. Let Δ' be the complement of Δ, and let $\Lambda\Delta$ be the intersection of Λ and Δ. Suppose an arbitrary η given, and that $N > N_0(\eta, \delta)$. For all i in Δ, $| P'(F_i^N) - P(F_i^N) | \leqq \delta P(F_i^N)$. Hence $\sum_\Delta P'(F_i^N) \geqq \sum_\Delta P(F_i^N) - \delta$. But, by hypothesis (b), $\sum_\Delta P(F_i^N) > 1 - \eta$ (a restatement of (6) above). Therefore $\sum_\Delta P'(F_i^N) > 1 - \eta - \delta$, or $\sum_{\Delta'} P'(F_i^N) < \eta + \delta$. It follows that $| \sum_{\Lambda\Delta} P'(F_i^N) - \sum_{\Lambda\Delta} P(F_i^N) | \leqq \delta$ and $| \sum_{\Lambda\Delta'} P'(F_i^N) - \sum_{\Lambda\Delta'} P(F_i^N) | < 2\eta + \delta$. Therefore $| \sum_\Lambda P'(F_i^N) - \sum_\Lambda P(F_i^N) | = | \Phi'_N(\lambda) - \Phi_N(\lambda) | < 2(\eta + \delta)$. But η and δ are arbitrary, and so $\Phi_N(\lambda) \to \Phi(\lambda)$ implies $\Phi'_N(\lambda) \to \Phi(\lambda)$, and the conclusion follows.

If all the f_i are strictly positive, formula (4) for the probability of an m-tuple frequency count, given the single-letter frequency count $\{f_i\}$, may be written in the form

$$P(\{f_{i_1 \cdots i_m}\}) = \frac{N \, \| \, [f_{i_1 \cdots i_m}]^* \, \|}{\prod f_{i_1 \cdots i_{m-1}}} \, \frac{\prod f_{i_1 \cdots i_{m-1}}! \prod f_i!}{N! \prod f_{i_1 \cdots i_m}!}. \tag{7}$$

The second of these two factors will be recognized [7] as the probability $P'([f_{i_i \cdots i_m}])$ of the cell entries $[f_{i_1 \cdots i_m}]$ in an ordinary contingency table with fixed marginal totals $\{f_{i_1 \cdots i_{m-1}}\}$ and $\{f_m\}$, under the hypothesis that the two attributes are independent. Suppose N approaches infinity while each f_i/N converges to a strictly positive constant k_i, with all $(N - 1)!$ circular permutations always equally probable. Then, for any fixed m, the relative frequencies of the m-tuples converge in probability to the corresponding product of k_i's:

$$f_{i_1 \cdots i_m}/N \to_p k_{i_1} \cdots k_{i_m},$$

implying

$$\| \, [f_{i_1 \cdots i_m}]^* \, \| / N^{t^{m-1}-1} \to_p \| \, [k_{i_1} \cdots k_{i_m}]^* \, \| .$$

The latter can be evaluated in various ways including factoring out the common factor in each row and diagonalizing; its value is $(k_1k_2 \cdots k_t)^{(m-1)t^{m-2}}$. (For another proof see the Appendix.) But

$$\prod f_{i_1 \cdots i_{m-1}} \to_p N^{t^{m-1}}(k_1k_2 \cdots k_t)^{(m-1)t^{m-2}},$$

and so the first factor in (7) converges in probability to unity. (The test may be applied to a linear sequence by applying the test to the corresponding circularized sequence without the blank (or see [14]).) Therefore, in view of the lemma, the hypothesis (H_0) of independence may be tested (N large) within the hypotheses (H_{m-1}) of Markovity of order $m - 1$ (see Good [4]) by any asymptotic test of contingency in the ordinary contingency table described above. Indeed, the statistic $-2 \log \lambda_{0,m-1}$ of [4] for testing H_0 within H_{m-1} is identical to the log likelihood-ratio statistic in the corresponding contingency table. (The asymptotic validity of the usual χ^2 tests on the contingency table corresponding to a Markov chain has already been indicated by Goodman [6], but the relationship between the exact probabilities for the contingency table and for the Markov chain is interesting.)

8. Distribution of pairs. In accordance with Sec. 6, the probability $P(F)$ of the specified pair-frequencies $F = [f_{ij}]$ arising in a circular permutation (see formula (2)) is asymptotic (in the stochastic sense) to the probability $P'(F)$ of the same entries F appearing in a contingency table of fixed marginal totals $\{f_i\}$. By formulas (1) and (2),

$$P(F)/P'(F) = NC(F)/\prod f_i!.$$

It follows that the expectation $E'g$ of a function $g(F)$ with respect to P' may be converted to an expectation Eg with respect to P by the relation

$$Eg = (N/\prod f_i!)E'(g \cdot C). \tag{8}$$

For example, the expectation of a product of factorial powers of the cell entries $\{f_{ij}\}$ of a contingency table with fixed marginal totals $f_{i.} = \sum_j f_{ij}$ and $f_{.j} = \sum_i f_{ij}$ is (cf. Haldane [7]), under the hypothesis of independence,

$$E' \prod f_{ij}^{(\alpha_{ij})} = \prod_i f_{i.}^{(\alpha_{i.})} \prod_j f_{.j}^{(\alpha_{.j})}/N^{(\alpha)} \tag{9}$$

where

$$\alpha_{i.} = \sum_j \alpha_{ij}, \alpha_{.j} = \sum_i \alpha_{ij} \quad \text{and} \quad \alpha = \sum_{i,j} \alpha_{ij}.$$

The α_{ij} are nonnegative integers.

[Proof of (9):

$$E' \prod f_{ij}^{(\alpha_{ij})} = \sum_{\{f_{ij}\}} \prod f_{ij}^{(\alpha_{ij})} P'(F)$$

$$= \sum_{\{f_{ij}\}} \prod f_{ij}^{(\alpha_{ij})} \frac{\prod f_{i.}! \prod f_{.j}!}{N! \prod f_{ij}!}$$

$$= \frac{\prod f_{i.}^{(\alpha_{i.})} \prod f_{.j}^{(\alpha_{.j})}}{N^{(\alpha)}} \sum_{\{f_{ij}\}} \frac{\prod (f_{i.} - \alpha_{i.})! \prod (f_{.j} - \alpha_{.j})!}{(N - \alpha)! \prod (f_{ij} - \alpha_{ij})!}.$$

But the terms of the latter sum are zero except when each $f_{ij} \geqq \alpha_{ij}$. Hence the sum may be taken over nonnegative values of $\{f_{ij} - \alpha_{ij}\}$, and so be recognized as a sum of probabilities in a (new) contingency table, and hence unity. This expectation is undoubtedly known, but the article by Haldane, who uses the same method for less general expectancies, is the only reference known to the authors.] If $f_{i.} = f_{.i} = f_i$, then

$$E' \prod f_{ij}^{(\alpha_{ij})} = \prod_i f_i^{(\alpha_{i.})} f_i^{(\alpha_{.i})} / N^{(\alpha)}.$$

Now the relation (8) and some algebraic manipulations give the corresponding expectation with respect to P:

$$E \prod f_{ij}^{(\alpha_{ij})} = \frac{\prod f_i^{(\alpha_{.i})} f_i^{(\alpha_{i.})}}{(N-1)^{(\alpha)}} \cdot \frac{|[f_i \delta_{ij} - \alpha_{ij}]|}{\prod f_i}. \tag{10}$$

where the vertical bars indicate the determinant of the enclosed matrix.

[Outline of steps necessary to obtain (10): The essence of the problem may be described as the evaluation of the factor X for which

$$E' \prod f_{ij}^{(\alpha_{ij})} \begin{vmatrix} f_2 - f_{22} & \cdots & -f_{2t} \\ \vdots & & \\ -f_{t2} & \cdots & -f_{tt} \end{vmatrix} = X \cdot E' \prod f_{ij}^{(\alpha_{ij})}.$$

Since $f_{ij}^{(\alpha_{ij})} \cdot f_{ij} = \alpha_{ij} f_{ij}^{(\alpha_{ij})} + f_{ij}^{(\alpha_{ij}+1)}$, and since the f_i and the α_{ij} commute with E', part of X (neglecting for the moment the expectation of terms containing the "higher powers" $\alpha_{ij} + 1$) is

$$m = \begin{vmatrix} f_2 - \alpha_{22} & \cdots & -\alpha_{2t} \\ \vdots & & \\ -\alpha_{t2} & \cdots & f_t - \alpha_{tt} \end{vmatrix}.$$

The next terms to consider are those which arise from selecting just one higher power factor. The effect of increasing one α_{ij} to $\alpha_{ij} + 1$ in $E' \prod f_{ij}^{(\alpha_{ij})}$ is to multiply the expectation by

$$(f_i - \alpha_{i.})(f_j - \alpha_{.j})/(N - \alpha).$$

Adding in these new terms gives

$$X = m - \sum_{i,j=2}^{t} (f_i - \alpha_{i.})(f_j - f_{.j}) \mu_{ij}/(N - \alpha) + R,$$

where μ_{ij} is the cofactor, in m, of the entry containing α_{ij}, and R, the remainder, consists of the expectation of terms containing two or more of the higher powers. But for each such term there is a term of equal and opposite expectation arising from using a different path through the determinant, namely the altered path which replaces the first (in the sense of going from the top down, say) two elements by the other two corners of the rectangle they span. Hence $R = 0$, and all that remains to be shown is that

$$(N - \alpha)m - \sum_{i,j=2}^{t} (f_i - \alpha_{i.})(f_j - \alpha_{.j}) \mu_{ij} = |[f_i \delta_{ij} - \alpha_{ij}]|.$$

Working backwards one finds, by elementary row and column operations,

$$\begin{vmatrix} f_1 - \alpha_{11} & -\alpha_{12} & \cdots & -\alpha_{1t} \\ -\alpha_{21} & f_2 - \alpha_{22} & \cdots & -\alpha_{2t} \\ \vdots & & & \\ -\alpha_{t1} & -\alpha_{t2} & \cdots & f_t - \alpha_{tt} \end{vmatrix} = \begin{vmatrix} f_1 - \alpha_{.1} & f_2 - \alpha_{.2} & \cdots & f_t - \alpha_{.t} \\ -\alpha_{21} & f_2 - \alpha_{22} & \cdots & -\alpha_{2t} \\ \vdots & & & \\ -\alpha_{t1} & -\alpha_{t2} & \cdots & f_t - \alpha_{tt} \end{vmatrix}$$

$$= \begin{vmatrix} N - \alpha & f_2 - \alpha_{.2} & \cdots & f_t - \alpha_{.t} \\ f_2 - \alpha_{2.} & f_2 - \alpha_{22} & \cdots & -\alpha_{2t} \\ \vdots & & & \\ f_t - \alpha_{t.} & -\alpha_{t2} & \cdots & f_t - \alpha_{tt} \end{vmatrix} = (N - \alpha)m - \sum_{i,j=2}^{t} (f_i - \alpha_{i.})(f_j - \alpha_{.j})\mu_{ij}].$$

In particular,

$$Ef_{ij}^{(\alpha)} = f_i^{(\alpha)}(f_j - \delta_{ij})^{(\alpha)}/(N-1)^{(\alpha)}, \tag{11}$$

whence the distribution of the frequency of a single pair (i, j) is found to be

$$P(f_{ij} = r) = \binom{f_i}{r}\binom{N - 1 - f_i}{f_j - \delta_{ij} - r} \bigg/ \binom{N-1}{f_j - \delta_{ij}}. \tag{12}$$

(The simplest verification is to compute the factorial moments of (12), but it is also possible to proceed directly with the help of Good and Toulmin [5], p. 46, together with Vandermonde's theorem.) One way of testing the null hypothesis of independence in a process against any alternative altering the probabilities of the consecutive pairs would be to use the familiar-looking statistic

$$\sum (f_{ij} - Ef_{ij})^2/Ef_{ij}$$

(where Ef_{ij} is given by (11) with $\alpha = 1$) on the circularized sample; the distribution is asymptotically gamma-variate with $(t - 1)^2$ degrees of freedom. Goodman [6] has found the same statistic except for the use here of the exact mean; and the analogous likelihood-ratio statistic is given by Hoel [8].

9. Acknowledgment. The authors are indebted to L. A. Goodman for valuable comments and suggestions.

APPENDIX

Evaluation of $\| [k_{i_1} \cdots k_{i_m}]^* \|$. We shall prove, without insisting on the condition

(i) $k_1 + k_2 + \cdots + k_t = 1$,

that

(ii) $\| [k_{i_1} \cdots k_{i_m}]^* \| = (k_1 k_2 \cdots k_t)^{(m-1)t^{m-2}} \cdot (k_1 + \cdots + k_t)^{t^{m-1}-m}$.

Proof. Let $M = [k_{i_1} \cdots k_{i_m}]$. Then $M = UG$ where U is the diagonal matrix $\{k_{i_1} \cdots k_{i_{m-1}}\}$, and G is the matrix that has zero elements at all places except in rows and columns with labels of the form

$$(i_1, \cdots, i_{m-1}) \quad \text{and} \quad (i_2, \cdots, i_m),$$

while at these places G has the element k_{i_m}. For example, with $t = m = 3$, and with $k_1 = a$, $k_2 = b$, $k_3 = c$, we have

$$U = \begin{bmatrix} aa & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & ab & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & ac & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & ba & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & bb & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & bc & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & ca & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & cb & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & cc \end{bmatrix},$$

$$G = \begin{bmatrix} a & b & c & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & a & b & c & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & a & b & c \\ a & b & c & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & a & b & c & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & a & b & c \\ a & b & c & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & a & b & c & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & a & b & c \end{bmatrix}.$$

The matrix whose cofactors we want is

$$M^* = sU - M = U(sI - G) = U \cdot K, \quad \text{say},$$

where

$$s = k_1 + k_2 + \cdots + k_t .$$

The rows of K each add up to zero while the rows and the columns of M^* each add up to zero. Thus the cofactors of all the elements in a fixed row of K are equal, while the cofactors of all the elements of M^* are equal. Denote their common value by

$$\kappa = \| [k_{i_1} \cdots k_{i_m}]^* \| .$$

κ may be obtained from the cofactor of any element in the νth row of K by multiplying this cofactor by the product of all the elements of U except its νth one. Therefore the sum of the cofactors of the diagonal elements of K is equal to

$$\text{(iii)} \quad \kappa \frac{\sum k_{i_1} \cdots k_{i_{m-1}}}{\prod k_{i_1} \cdots k_{i_{m-1}}} = \kappa s^{m-1} / (k_1 \cdots k_t)^{(m-1)t^{m-2}}.$$

But it is also equal to plus or minus the coefficient of λ in the characteristic polynomial of K, $| \lambda I - K |$. This coefficient can be found by considering the eigenvalues of G.

In the above example with $t = m = 3$, by assuming that the components of a column eigenvector are $x_1, x_2, \cdots, x_9$, and by multiplying this eigenvector by G, we easily see that either the corresponding eigenvalue is zero, or else it is s and $x_1 = x_2 = \cdots = x_9$. These two eigenvalues both occur at least once, so the characteristic polynomial of G is of the form

(iv) $|\lambda I - G| = \lambda^\alpha(\lambda - s)^\beta \quad (\alpha > 0, \beta > 0, \alpha + \beta = t^{m-1})$. (A more explicit proof of (iv) is given below.) Hence the characteristic polynomial of K is of the form

$$\lambda^\beta(\lambda - s)^\alpha,$$

and, since κ is positive, Eq. (iii) shows that $\beta = 1$ and that (ii) is true.

The following checks of equation (ii) will help to clarify its relationship to previous literature. If we put $k_1 = k_2 = \cdots = k_t = 1$ and apply the BEST theorem, then we find that the number of circular arrays that contain each m-tuple precisely once is

$$(t!)^{t^{m-1}}/t^m,$$

and this agrees with p. 203 of de Bruijn and Ehrenfest [2]. If instead we put $k_1 = k_2 = \cdots = k_t = k^{1/m}$, then we find that if each m-tuple is to appear exactly k times the number of arrays is

$$[(tk)!]^{t^{m-1}}/(kt^m),$$

and this agrees with Theorem 3 of the same paper.

The following proof of (iv) was kindly provided by Mr. O. S. Rothaus. If we insist on condition (i), and it is easily seen that there is no real loss of generality in doing so, then G gives the $(m - 1)$-tuple transition probabilities in an independent and stationary process. Now G^{m-1} has constant columns; in fact every entry in the column labelled $i_2, \cdots, i_m$ is $k_{i_2} \cdots k_{i_m}$. The stochastic matrix G^{m-1} is of rank 1. It has the eigenvalue 1 and all other eigenvalues are zero. Since G is stochastic, it too has the eigenvalue 1; but the eigenvalues of G^{m-1} are the $(m - 1)$-st powers of those of G, and so the other eigenvalues of G must all be zero. This proves (iv) and also that $\beta = 1$.

REFERENCES

1. M. S. Bartlett, "The frequency goodness of fit test for probability chains," *Proc. Cambridge Philos. Soc.*, Vol. 47 (1951), pp. 86–95.
2. N. G. de Bruijn and T. van Aardenne-Ehrenfest, "Circuits and trees in oriented linear graphs," Simon Stevin, Vol. 4 (1950–51), pp. 203–217.
3. I. J. Good, "The serial test for sampling numbers and other tests for randomness," *Proc. Cambridge Philos. Soc.*, Vol. 49 (1953), pp. 276–284.
4. I. J. Good, "The likelihood ratio test for Markoff chains," *Biometrika*, Vol. 42 (1955), pp. 531–533; corrigenda, Vol. 44 (1957), p. 301.
5. I. J. Good and G. H. Toulmin, "The number of new species, and the increase in population coverage, when a sample is increased," *Biometrika*, Vol. 43 (1956), pp. 45–63.
6. L. A. Goodman, "On the statistical analysis of Markov chains," *Ann. Math. Stat.*, Vol. 26 (1955), p. 771.
7. J. B. S. Haldane, "The mean and variance of χ^2, when used as a test of homogeneity, when expectations are small," *Biometrika*, Vol. 31 (1940), pp. 346–355.

8. P. G. Hoel, "A test for Markoff chains," *Biometrika*, Vol. 41 (1954), pp. 430–433.
9. A. M. Mood, "The distribution theory of runs," *Ann. Math. Stat.*, Vol. 11 (1940), pp. 367–392.
10. W. L. Stevens, "Distribution of groups in a sequence of alternatives," Ann. Engenics, Vol. 9 (1939), pp. 10–17.
11. W. T. Tutte, "The dissection of equilateral triangles into equilateral triangles," *Proc. Cambridge Philos. Soc.*, Vol. 44 (1948), pp. 463–482.
12. W. T. Tutte and C. A. B. Smith, "On unicursal paths in a network of degree 4," *Amer. Math. Monthly*, Vol. 48 (1941), pp. 233–237.
13. P. Whittle, "Some distribution and moment formulae for the Markov chain," *J. Roy. Stat. Soc.*, Series B, Vol. 17 (1955), pp. 235–242.
14. L. A. Goodman, "Exact probabilities and asymptotic relationships for some statistics from m-th order Markov chains," to appear.