A BAYESIAN STUDY OF THE MULTINOMIAL DISTRIBUTION[1]

By Daniel A. Bloch and Geoffrey S. Watson

The Johns Hopkins University

1. Introduction and summary. Lindley [6] studies the topic in our title. By using Fisher's conditional-Poisson approach to the multinomial and the logarithmic transformation of gamma variables to normality, he showed that linear contrasts in the logarithms of the cell probabilities θ_i are asymptotically jointly normal and suggested that the approximation can be improved by applying a "correction" to the sample. By studying the asymptotic series for the joint distribution in Section 2 an improved correction procedure is found below. A more detailed expansion is given in Section 3 for the distribution of a single contrast in the log θ_i. In many problems a linear function of the θ_i is of interest. The exact distribution is obtained and is of a form familiar in the theory of serial correlation coefficients. A beta approximation is given. For three cells, a numerical example is given to show the merit of this approximation. A genetic linkage example is considered which requires the joint distribution of two linear functions of the θ_i. The exact joint distribution is found but is too involved for practical use. A normal approximation leads to Lindley's results [7].

2. Relations between the cell probabilities. By using Fisher's conditional-Poisson approach to the multinomial and the logarithmic transformation of Type III variables to normality, Lindley [6] shows that contrasts between the log θ_i's ($A_p = \sum_{i=1}^k a_{pi} \log \theta_i$, $p = 1, \cdots, m$; $\sum_{i=1}^k a_{pi} = 0$; $m < k$) may be easily handled. If the joint prior density of $\theta_1, \cdots, \theta_k$ is proportional to $(\prod_{i=1}^k \theta_i)^{-1}$, he indicates that, as $n \to \infty$, $A_1, \cdots, A_m$ will be approximately normally distributed with means $\sum_{i=1}^k a_{pi} \log n_i$ and covariances $(A_p, A_q) = \sum_{i=1}^k a_{pi} a_{qi} n_i^{-1}$ $(p, q = 1, \cdots, m)$. He suggests that the approximation will be improved by replacing n_i by $n_i - \frac{1}{2}$. The "corrections" derived below vary from cell to cell but have one-half as their leading term.

With a uniform prior, the posterior density of $\theta_1, \cdots, \theta_k$ is given by

$$\text{Post}(\theta_1, \cdots, \theta_k) = [(n + k - 1)!/\prod_{i=1}^k (n_i)!] \prod_{i=1}^k \theta_i^{n_i}. \tag{2.1}$$

The joint moment generating function of $A_1, \cdots, A_m$ is defined by $\varphi(t_1, \cdots, t_m) = E(\exp(\sum_{p=1}^m t_p A_p))$, i.e.

$$\varphi = E(\prod_{i=1}^k \theta_i^{\sum_{p=1}^m t_p a_{pi}}). \tag{2.2}$$

From the normalizing constant in (2.1), it is clear that, since $\sum_{i=1}^k a_{pi} = 0$,

$$\varphi = \prod_{i=1}^k \{\Gamma(n_i + \sum_{p=1}^m t_p a_{pi} + 1)/\Gamma(n_i + 1)\}. \tag{2.3}$$

Received 1 November 1966; revised 31 May 1967.

[1] Research supported by the Office of Naval Research under contract NONR 4010(09) awarded to the Department of Statistics, The Johns Hopkins University. This paper in whole or in part may be reproduced for any purpose of the United States Government.

The joint normality of $A_1, \cdots, A_m$ with means $\sum_{i=1}^k a_{pi} \log n_i$ and covariances $\sum_{i=1}^k a_{pi}a_{qi}n_i^{-1}$ then follows by using the first term of Stirlings's series for each factorial in (2.3), as all $n_i \to \infty$. The difference between the uniform prior and $(\prod_{i=1}^k \theta_i)^{-1}$ is lost as all the $n_i \to \infty$ since the effect is only to decrease all the n_i by unity. Lindley suggests that the special case $A_i = \log \theta_i$ ($i = 1, 2, \cdots, k$) be used to yield $\sum_{i=1}^k n_i(u_i - \bar{u})^2 \approx X_{k-1}^2$ where $u_i = \log \theta_i - \log n_i$. Other methods of getting approximate or exact joint (Bayesian) confidence regions for the θ_i are given in Watson [12].

To improve the approximation it is clearly necessary to use more terms in the asymptotic expansions for the means and covariances of the A_p's which may be obtained by differentiating their joint cumulative generating function $K = \log \varphi$ which from (2.3) is given by

$$K(t_1, \cdots, t_m) = \sum_{i=1}^k \{\log \Gamma(n_i + \sum_{p=1}^m t_p a_{pi} + 1) - \log \Gamma(n_i + 1)\}. \tag{2.4}$$

Routine calculations lead to the formulas

$$E(A_q) \sim \sum_{i=1}^k a_{qi}\{\log n_i + (2n_i)^{-1} - (12n_i^2)^{-1} + (120n_i^4)^{-1} - \cdots\}, \tag{2.5}$$

$$\operatorname{Cov}(A_q, A_l) \sim \sum_{i=1}^k a_{qi}a_{li}\{n_i^{-1} - (2n_i^2)^{-1} + (6n_i^3)^{-1} - (30n_i^5)^{-1} + \cdots\}. \tag{2.6}$$

These formulas apply to the case of a uniform prior and the sample $(n_1, n_2, \cdots, n_k)$. The corrections $c_1, c_2, \cdots, c_k$ need to be chosen so that $\sum_{i=1}^k a_{qi} \log (n_i + c_i)$ is closer to $E(A_q)$ than $\sum_{i=1}^k a_{qi} \log n_i$ and so that $\sum_{i=1}^k a_{qi}a_{li}(n_i + c_i)^{-1}$ is closer to $\operatorname{Cov}(A_q, A_l)$ than $\sum_{i=1}^k a_{qi}a_{li}n_i^{-1}$. Comparing the expansion

$$\log (n_i + c_i) \sim \log n_i + c_i/n_i - c_i^2/2n_i^2 + c_i^3/3n_i^3 - \cdots$$

with that in (2.5), the choice $c_i = \frac{1}{2}$ equalizes the $O(n_i^{-1})$ terms. However, the choice

$$c_i = \tfrac{1}{2} + 1/24n_i \tag{2.7}$$

makes the series equal to $O(n_i^{-2})$. Comparing the series for $(n_i + c_i)^{-1}$ with the series in (2.6), the choice (2.7) keeps them equal to $O(n_i^{-2})$ since the difference is $5(24n_i^3)^{-1} + O(n_i^{-4})$.

Had the prior density of $\theta_1, \cdots, \theta_k$ been proportional to $\prod_{i=1}^k \theta_i^{m_i}$, the n_i above would be replaced by $n_i + m_i$. In this case our results suggest that the joint distribution of $A_1, \cdots, A_m$ be approximated by the m-variate normal distribution with

$$E(A_q) = \sum_{i=1}^k a_{qi} \log (n_i + m_i + c_i), \tag{2.8}$$
$$\operatorname{Cov}(A_q, A_l) = \sum_{i=1}^k a_{qi}a_{li}(n_i + m_i + c_i)^{-1}$$

where $c_i = \frac{1}{2} + 1/24(n_i + m_i)$ ($i = 1, 2, \cdots, k$). For the particular case considered by Lindley, $m_1 = m_2 = \cdots = m_k = -1$ so that $n_i + m_i + c_i = n_i - \frac{1}{2} + 1/24(n_i - 1)$, which is his result except for the $O(n_i^{-1})$ term.

3. Expansion for the distribution of $A = \sum_{i=1}^k a_i \log \theta_i$**.** For a single contrast A it seems more sensible to give an expansion for its distribution rather than simply to "correct" to improve the normal approximation with mean $\sum_{i=1}^k a_i \log n_i$ and variance $\sum_{i=1}^k a_i^2 n_i^{-1}$. However to make the first term more effective, the n_i should be increased by $\frac{1}{2}$ for a uniform prior by the argument of Section 2. If $N_i = n_i + \frac{1}{2}$, then we can effect this by taking as the characteristic function $\varphi_A(t)$ of A,

$$\varphi_A(t) = \prod_{j=1}^k \Gamma(N_j + ita_j + 1)/\Gamma(N_j + 1). \tag{3.1}$$

Setting $\mu = \sum_{i=1}^k a_i \log N_i$, $\sigma^2 = \sum_{i=1}^k a_i^2 N_i^{-1}$, $z = (A - \mu)\sigma^{-1}$, and $\varphi(z) = (2\pi)^{-\frac{1}{2}} \exp(-z^2/2)$, we require a series with first term $\varphi(z)$. This may be obtained by the method shown in Cramér [2]. Denoting the derivatives of $\varphi(z)$ by $\varphi^{(1)}(z)$, $\varphi^{(2)}(z)$, etc., we find that the density $f(z)$ of $(A - \mu)\sigma^{-1}$ is given by the asymptotic series

$$\begin{aligned} f(z) = \varphi(z) - c_1\sigma^{-1}\varphi^{(1)}(z) + c_2\sigma^{-2}\varphi^{(2)}(z) - c_3\sigma^{-3}\varphi^{(3)}(z) \\ + (c_4 + 4c_1c_3)\sigma^{-4}\varphi^{(4)}(z) + O(n^{-5/2}), \end{aligned} \tag{3.2}$$

where

$$\begin{aligned} c_1 &= \textstyle\sum_{i=1}^k a_i\{3/8N_i^2 + 1/3N_i^3 + 27/64N_i^4\}, \\ c_2 &= -\textstyle\sum_{i=1}^k \frac{1}{4}a_i^2\{1/N_i^3 + 5/2N_i^4\}, \\ c_3 &= -\textstyle\sum_{i=1}^k a_i^3\{1/N_i^2 - 3/2N_i^4\}, \\ c_4 &= 2\textstyle\sum_{i=1}^k a_i^4/N_i^3. \end{aligned}$$

4. The distribution of an arbitrary linear combination of the cell probabilities. Let $n_1, n_2, \cdots, n_k$ be a sample from the multinomial distribution with cell probabilities $\theta_1, \theta_2, \cdots, \theta_k$ respectively $(\sum_{i=1}^k \theta_i = 1)$. Let $l = \sum_{i=1}^k \lambda_i\theta_i$, where $\lambda_1, \lambda_2, \cdots, \lambda_k$ are k distinct real numbers. We will first find the exact distribution of l and then consider approximations to it. Without loss of generality we may assume that $\lambda_1 < \lambda_2 < \cdots < \lambda_k$. Let the prior distribution of the θ_i's be uniform. Then the posterior distribution of the θ_i's, given by formula (2.1), may be rewritten as

$$\text{Post}\ (\theta_1, \cdots, \theta_k) = [\Gamma(\textstyle\sum_{i=1}^k \nu_i)/\prod_{i=1}^k \Gamma(\nu_i)]\theta_1^{\nu_1-1}\theta_2^{\nu_2-1} \cdots \theta_k^{\nu_k-1} \tag{4.1}$$

where $\nu_i = n_i + 1$, $i = 1, 2, \cdots, k$.

It is known (see e.g. [13]) that the θ_i's can be expressed as

$$\theta_i = y_i/\textstyle\sum_{i=1}^k y_i \tag{4.2}$$

where the y_i's are independent random variables having gamma distributions γ_{ν_i}, $i = 1, 2, \cdots, k$. Thus

$$l = \textstyle\sum_{i=1}^k \lambda_i\theta_i = \sum_{i=1}^k \lambda_i y_i/\sum_{i=1}^k y_i.$$

The distribution problem in this form was first discussed by McCarthy [8] and Koopmans [5]. Box [1] introduced a Fisherian trick with ratios to lead directly to the distribution function of l. Thus

$$\Pr(l > \rho) = \Pr\{\textstyle\sum_{i=1}^k \lambda_i y_i > \rho \sum_{i=1}^k y_i\} = \Pr\{X > 0\},$$

where $X = \sum_{i=1}^k w_i y_i$, $w_i = \lambda_i - \rho$, $i = 1, 2, \cdots, k$. Since it is assumed that the ν_i are integral, the characteristic function of X, $\varphi_X(t)$, is equal to $\prod_{j=1}^k (1 - itw_j)^{-\nu_i}$. $\varphi_X(t)$ may be inverted directly after being put in partial fraction form. The density of X on $(0, \infty)$ is then a linear combination of gamma densities associated with positive w_j's. After integrating from 0 to ∞, Box finds the formula

$$\Pr\{l > \rho\} = \textstyle\sum_{\{j:\lambda_j>\rho\}} \prod_{j'\neq j}^k \{(\lambda_j - \rho)/(\lambda_j - \lambda_{j'})\}^{\nu_{j'}} \sum_{s=1}^{\nu_j} \mu'_{j,\nu_j-s}/(\nu_j - s)!$$

where μ'_{j,ν_j-s} is related to

$$K_{j,\nu_j-s} = (\nu_j - s - 1)! \textstyle\sum_{j'\neq j}^k \{\nu_{j'}[(\lambda_{j'} - \rho)/(\lambda_{j'} - \lambda_j)]^{\nu_j-s}\}$$

in the same way as are the $(\nu_j - s)$th moment about the origin and the $(\nu_j - s)$th cummulant (we define $\mu'_{j,0} = 1$).

Similarily,

$$(4.3)\quad \Pr\{l < \rho\} = \textstyle\sum_{\{j:\lambda_j<\rho\}} \prod_{j'\neq j}^k \{(\rho - \lambda_j)/(\lambda_{j'} - \lambda_j)\}^{\nu_{j'}} \cdot \sum_{s=1}^{\nu_j} \mu'_{j,\nu_j-s}/(\nu_j - s)!.$$

Assuming that $\lambda_{p-1} \leqq \rho < \lambda_p$ and differenting (4.3) with respect to ρ (at $\rho = l$) we find that the exact density function of l is given by

$$(4.4)\quad g(l) = \sum_{j=1}^{p=1} \left\{ \frac{\left(\sum_{j'\neq j}^k \nu_{j'}\right)(l - \lambda_j)^{(\Sigma_{j'\neq j}^k \nu_{j'} - 1)}}{\prod_{j'\neq j}^k (\lambda_{j'} - \lambda_j)^{\nu'_j}} \sum_{s=1}^{\nu_j} \frac{\mu'_{j,\nu_j-s}}{(\nu_j - s)!} + \frac{(l - \lambda_j)^{(\Sigma_{j'\neq j}^k \nu_{j'})}}{\prod_{j'\neq j}^k (\lambda_{j'} - \lambda_j)^{\nu_{j'}}} \sum_{s=1}^{\nu_j} \frac{d}{dl} \frac{\mu'_{j,\nu_j-s}}{(\nu_j - s)!} \right\}, \quad \lambda_{p-1} \leqq l < \lambda_p$$

$$= 0, \qquad \text{otherwise.}$$

Since the exact result is too cumbersome for practical use, it is necessary to derive approximations. Since k is in general small and the λ_i arbitrary, we must utilize the fact that the ν_i can be large. This leads immediately to the suggestion that the distribution of $X = \sum_{i=1}^k w_i y_i$ be approximated by a normal distribution with mean $\sum_{i=1}^k \nu_i w_i$ and variance $\sum_{i=1}^k \nu_i w_i^2$ with the implication that

$$\Pr(l > \rho) \approx 1 - \Phi(-\Sigma\nu_j(\lambda_j - \rho)[\Sigma\nu_j(\lambda_j - \rho)^2]^{-\frac{1}{2}}).$$

It is easy to use $\varphi_x(t)$ to obtain an expansion for the density $g(l)$ with the leading term a normal density with mean $E(l)$ and variance Var (l). The finite range of l suggests that a beta distribution would be a more accurate approximation.

This too can be made the first term of an expansion by using Jacobi polynomials as in Durbin and Watson [3] since all the moments of l can be easily obtained from

$$E(l^p) = E(\sum_{i=1}^k \lambda_i y_i)^p / E(\sum_{i=1}^k y_i)^p.$$

Instead we merely suggest the beta approximation given below and compare it to the normal approximation and the exact density for a numerical example.

Let $X = (l - \lambda_1)/(\lambda_k - \lambda_1)$ be distributed as Be (p, q), the constants p and q being chosen so that the distribution has the same first two moments as l. Then we have that $l = \sum_{i=1}^k \lambda_i \theta_i$ is distributed approximately as

$$[\Gamma(p+q)/(\lambda_k - \lambda_1)\Gamma(p)\Gamma(q)]((l - \lambda_1)/(\lambda_k - \lambda_1))^{p-1}((\lambda_k - l)/(\lambda_k - \lambda_1))^{q-1}$$

where

$$p = ((E(l) - \lambda_1)/(\lambda_k - \lambda_1))[\{(E(l) - \lambda_1)(\lambda_k - E(l))/\mathrm{Var}\,(l)\} - 1],$$
$$q = ((\lambda_k - E(l))/(\lambda_k - \lambda_1))[\{(E(l) - \lambda_1)(\lambda_k - E(l))/\mathrm{Var}\,(l)\} - 1],$$
$$E(l) = \sum_{i=1}^k \lambda_i \nu_i/\nu \quad \text{and}$$
$$\mathrm{Var}\,(l) = [\nu \sum_{i=1}^k \lambda_i^2 \nu_i - (\sum_{i=1}^k \lambda_i \nu_i)^2]/\nu^2(\nu + 1), \quad \sum_{i=1}^k \nu_i = \nu.$$

Since the distribution of $F = (q/p)((l - \lambda_1)/(\lambda_k - l))$ is $F_{2p,2q}$ (see [7]), $F_{\frac{1}{2}\alpha}(2p, 2q) = [F_{1-\frac{1}{2}\alpha}(2q, 2p)]^{-1} < F < F_{1-\frac{1}{2}\alpha}(2p, 2q)$ is a $1 - \alpha$ confidence interval for F. Hence

$$(4.5) \quad [\lambda_1 + (p/q\lambda_k)(F_{1-\frac{1}{2}\alpha}(2q, 2p))^{-1}]\{1 + (p/q)[F_{1-\frac{1}{2}\alpha}(2q, 2p)]^{-1}\}^{-1}$$
$$< l < [\lambda_1 + ((p/q)\lambda_k)F_{1-\frac{1}{2}\alpha}(2p, 2q)][1 + (p/q)F_{1-\frac{1}{2}\alpha}(2p, 2q)]^{-1}$$

is an approximate $1 - \alpha$ confidence interval for l.

EXAMPLE. $k = 3$. $\nu_1 = 2, \nu_2 = 3, \nu_3 = 6$. $\lambda_1 = -1, \lambda_2 = 1, \lambda_3 = 2$ $(l = -\theta_1 + \theta_2 + 2\theta_3)$,

(a) *Exact.*

$$g(l) = 5(l+1)^8(11 - 7l)/2^3 3^6, \qquad -1 \leqq l \leqq 1,$$
$$= 5(l+1)^8(11 - 7l)/2^3 3^6 + ((l-1)^7/2^2)(967.5 - 765l + 157.5l^2), \qquad 1 \leqq l \leqq 2,$$
$$= 0, \qquad \text{otherwise.}$$

(b) *Beta approximation.*

$$E(l) = 1.18, \quad \mathrm{Var}\,(l) = 0.103, \quad p = 11.89, \quad q = 4.47.$$
$$g(l) \doteq [\Gamma(16.36)/3\Gamma(11.89)\Gamma(4.47)]((l+1)/3)^{10.89}((2 - l)/3)^{3.47}, \qquad -1 \leqq l \leqq 2.$$

(c) *Normal approximation.*

$$g(l) \doteq (2\pi(0.103))^{-\frac{1}{2}} \exp\{-2(0.103)^{-1}(l - 1.18)^2\}, \quad -\infty < l < \infty.$$

Using formula (4.5), an approximate 95% confidence interval for l is given by (.49, 1.72), i.e. Pr $\{.49 < l < 1.72\} \doteq .95$.

Figure I is a graph of the probability densities (calculated from formulas (a), (b) and (c)) for the above example. This example suggests that the simple beta approximation is quite a good one, especially for the tail probabilities, and appears to be considerably better than the normal approximation.

5. The joint distribution of two linear functions in the cell probabilities. This joint distribution will often be required. For example the genetic linkage problem described by Fisher [4] may be treated in the Bayesian manner (Lindley [7] considers simpler problems of this kind). For our purposes it suffices to consider the 4-nomial

$$(5.1)\quad \theta_1(\theta) = \tfrac{1}{4}(2+\theta),\quad \theta_2(\theta) = \tfrac{1}{4}(1-\theta),\quad \theta_3(\theta) = \tfrac{1}{4}(1-\theta),\quad \theta_4(\theta) = \tfrac{1}{4}\theta,$$

where θ is known only to be in $(0,1)$. $\theta = 0$ corresponds to no linkage. The problem is to check whether the data fits the model (5.1) for some θ.

To obtain a Bayesian version of this goodness-of-fit test, it is necessary to introduce parameters that will be zero when the linkage model is followed. From (5.1) it is natural, but not of course necessary, to consider linear combinations of the cell probabilities, e.g. for cell frequencies (n_1, n_2, n_3, n_4) let

$$(5.2)\quad \begin{aligned} \phi_1 &= \tfrac{1}{2}(\theta_2 - \theta_3), \quad \phi_2 = \tfrac{1}{4}\theta_1 - \tfrac{1}{4}\theta_2 - \tfrac{1}{4}\theta_3 - \tfrac{3}{4}\theta_4, \\ &\qquad\qquad \phi_3 = \tfrac{1}{4}(\theta_1 - \theta_2 - \theta_3 + \theta_4). \end{aligned}$$

When $\theta_1, \theta_2, \theta_2, \theta_3, \theta_4$ satisfy (5.1), ϕ_1 and ϕ_2 are zero and $\phi_3 = \frac{1}{4}\theta$ (ϕ_3 is related to the inefficient estimator $n^{-1} \times (n_1 - n_2 - n_3 + n_4)$ for θ). We need the posterior confidence region for the linear forms ϕ_1, ϕ_2 and the goodness-of-fit test will be made by checking whether or not this covers the point $\phi_1 = \phi_2 = 0$.

Generally, consider $l_1 = \sum_{i=1}^k \lambda_i\theta_i$ and $l_2 = \sum_{i=1}^k \mu_i\theta_i$. We wish to find the joint distribution of l_1 and l_2, $g(l_1, l_2)$ say. As before we can express l_1 and l_2 as $l_1 = \sum_{i=1}^k \lambda_i y_i / \sum_{i=1}^k y_i$ and $l_2 = \sum_{i=1}^k \mu_i y_i / \sum_{i=1}^k y_i$ where the y_i's are as in Section 4. We now find the joint distribution of

$$r_1 = \textstyle\sum_{i=1}^k \lambda_i y_i, \quad r_2 = \sum_{i=1}^k \mu_i y_i \quad \text{and} \quad r = \sum_{i=1}^k y_i.$$

The joint characteristic function of r_1, r_2 and r is given by

$$(5.3)\qquad \varphi_{r_1,r_2,r}(t_1, t_2, t_3) = [\textstyle\prod_{j=1}^k (1 - it_1\lambda_j - it_2\mu_j - it_3)^{\nu_j}]^{-1}.$$

Watson (see [11]) showed that

$$\Psi(t_1, t_2, t_3) = [\textstyle\prod_{j=1}^k (1 - it_1\lambda_j - it_2\mu_j - it_3)]^{-1}$$

admits a partial fraction decomposition

$$(5.4)\quad \begin{aligned} \Psi(t_1, t_2, t_3) = \textstyle\sum_{j_1<j_2}^k B_{j_1j_2}[(1 - it_2)^{k-2}(1 - it_1\lambda_{j_1} \\ - it_2\mu_{j_1} - it_3)(1 - it_1\lambda_{j_2} - it_2\mu_{j_2} - it_3]^{-1} \end{aligned}$$

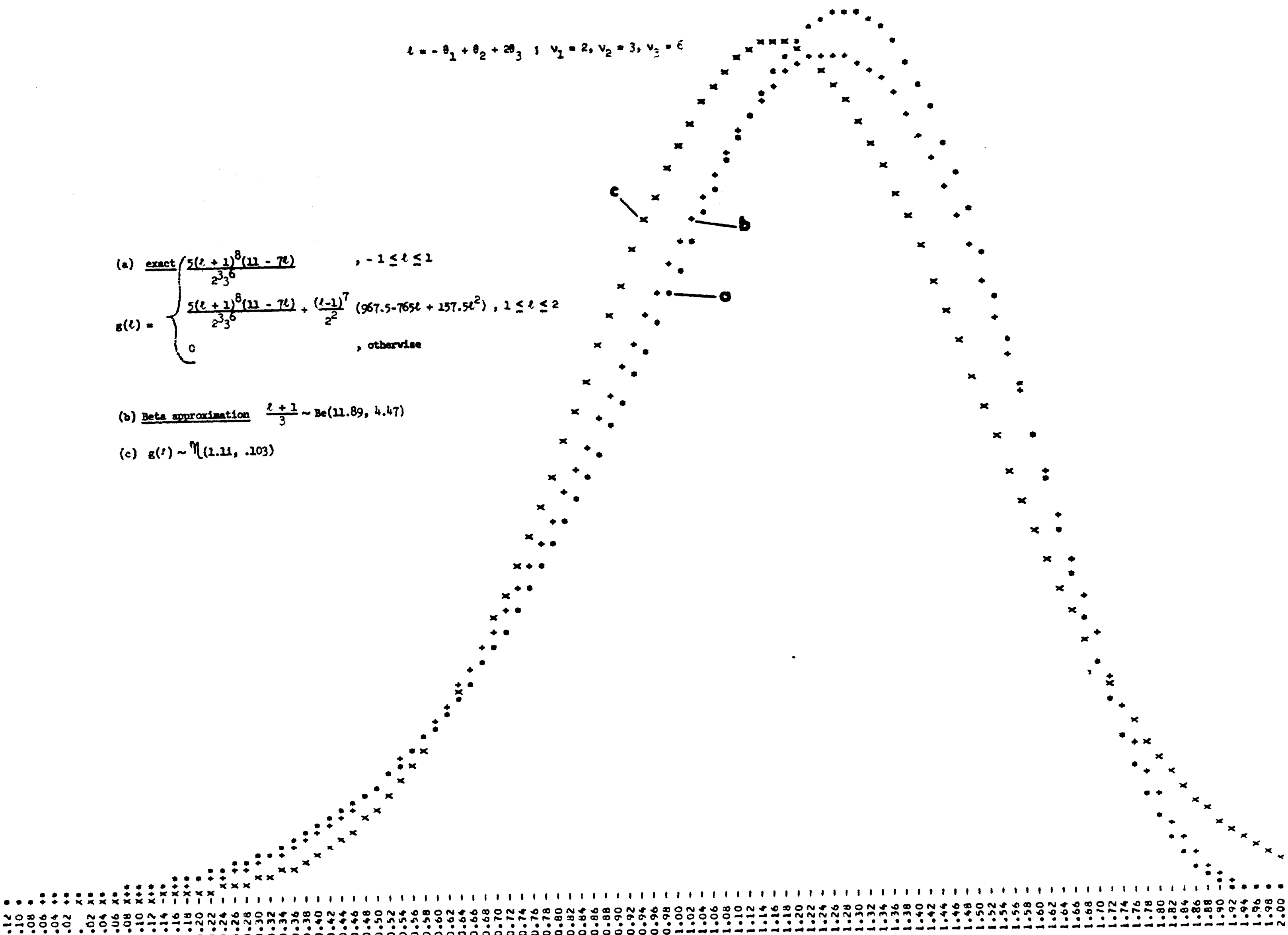
$\ell = -\theta_1 + \theta_2 + 2\theta_3$; $\nu_1 = 2, \nu_2 = 3, \nu_3 = 6$
(a) exact $g(\ell) = \begin{cases} \frac{5(\ell+1)^8(11-7\ell)}{2^3 3^6} & , -1 \le \ell \le 1 \\ \frac{5(\ell+1)^8(11-7\ell)}{2^3 3^6} + \frac{(\ell-1)^7}{2^2}(967.5-765\ell+157.5\ell^2) & , 1 \le \ell \le 2 \\ 0 & , \text{otherwise} \end{cases}$
(b) Beta approximation $\frac{\ell+1}{3} \sim Be(11.89, 4.47)$
(c) $g(\ell) \sim \eta(1.11, .103)$
a
b
c

FIG. 1

where

$$B_{j_1j_2} = \begin{vmatrix} \lambda_{j_1} & \mu_{j_1} \\ \lambda_{j_2} & \mu_{j_2} \end{vmatrix}^{k-2} \Big/ \prod_{j\neq j_1,j_2} \begin{vmatrix} 1 & \lambda_j & \mu_j \\ 1 & \lambda_{j_1} & \mu_{j_1} \\ 1 & \lambda_{j_2} & \mu_{j_2} \end{vmatrix}$$

provided the λ's and μ's are such that all the B's are finite. From formula (5.4) we have that

$$(5.5)\quad (y_m y_l y_k)^{-1} = B_{ml}/(1 - it_3)y_m y_l + B_{mk}/(1 - it_3)y_m y_k + B_{lk}/(1 - it_3)y_l y_k$$

where

$$y_m = 1 - it_1\lambda_m - it_2\mu_m - it_3 ,$$
$$y_l = 1 - it_1\lambda_l - it_2\mu_l - it_3 ,$$
$$y_k = 1 - it_1\lambda_k - it_2\mu_k - it_3 .$$

Hence

$$(5.6)\quad (y_m^s y_l^t y_k^{\nu_k})^{-1} = B_{ml}/(1 - it_3)y_m^s y_l^t y_k^{\nu_k-1} + B_{mk}/(1 - it_3)y_m^s y_l^{t-1} y_k^{\nu_k} + B_{lk}/(1 - it_3)y_m^{s-1} y_l^t y_k^{\nu_k}.$$

Formula (5.6) provides a recursive relationship which (upon repeatedly applying formula (5.5) proves that $\varphi_{r_1 r_2, r}(t_1 , t_2 , t_3)$ admits the partial fraction decomposition

$$(5.7)\quad \varphi_{r_1 ,r_2, r}(t_1 , t_2 , t_3) = \textstyle\sum_{m<l}^{k} \sum_{s=1}^{\nu_m} \sum_{t=1}^{\nu_l} A_{mlst}/D_{mlst}$$

where

$$\nu = \textstyle\sum_{i=1}^{k} \nu_i ,$$
$$D_{mlst} = (1 - it_1\lambda_m - it_2\mu_m - it_3)^s(1 - it_1\lambda_l - it_2\mu_l - it_3)^t(1 - it_3)^{\nu-s-t}.$$

It is easily verified that the joint characteristic function of

$$l_1' = \lambda_m\tilde{y}_m + \lambda_l\tilde{y}_l ,$$
$$l_2' = \mu_m\tilde{y}_m + \mu_l\tilde{y}_l ,$$
$$r' = \tilde{y}_m + \tilde{y}_l + \tilde{y}_u ,$$

where $\tilde{y}_m$, $\tilde{y}_l$, $\tilde{y}_u$ are independent gamma random variables with parameter s, t and $\nu - s - t$ respectively is given by

$$[(1 - it_1\lambda_m - it_2\mu_m - it_3)^s(1 - it_1\lambda_l - it_2\mu_l - it_3)^t(1 - it_3)^{\nu-s-t}]^{-1}.$$

Hence we wish to find the joint density of (l_1', l_2', r'). Let $r_1' = l_1'/r'$ and $r_2' = l_2'/r'$. r_1' and r_2' are distributed independently of r' by Pitman's theorem (see [9]). Without loss of generality, take $m = 1$ and $l = 2$. Then

$$r_1' = \lambda_1 c_1 + \lambda_2 c_2 ,$$
$$r_2' = \mu_1 c_1 + \mu_2 c_2 ,$$

where $c_1 = \tilde{y}_1/r'$, $c_2 = \tilde{y}_2/r'$. Hence the joint density of c_1 , c_2 , r' (see (9)) is given by

$$j\,df(c_1, c_2, r') = \Gamma(\nu)c_1^{s-1}c_2^{t-1}(1 - c_1 - c_2)^{\nu-s-t-1}(r')^{\nu-1}e^{-r'} \cdot [\Gamma(s)\Gamma(t)\Gamma(\nu - s - t)\Gamma(\nu)]^{-1} \tag{5.8}$$

where $c_1, c_2 \geqq 0$ and $c_1 + c_2 \leqq 1$.

Assuming $\begin{vmatrix} \lambda_1 & \mu_1 \\ \lambda_2 & \mu_2 \end{vmatrix} \neq 0$,

$$\begin{aligned} c_1 &= \frac{(\mu_2 r_1' - \lambda_2 r_2')}{\begin{vmatrix} \lambda_1 & \mu_1 \\ \lambda_2 & \mu_2 \end{vmatrix}} \geqq 0, \\ c_2 &= \frac{\lambda_1 r_2' - \mu_1 r_1'}{\begin{vmatrix} \lambda_1 & \mu_1 \\ \lambda_2 & \mu_2 \end{vmatrix}} \geqq 0, \\ 1 \geqq 1 - c_1 - c_2 &= \frac{\begin{vmatrix} 1 & r_1' & r_2' \\ 1 & \lambda_1 & \mu_1 \\ 1 & \lambda_2 & \mu_2 \end{vmatrix}}{\begin{vmatrix} \lambda_1 & \mu_1 \\ \lambda_2 & \mu_2 \end{vmatrix}} \geqq 0. \end{aligned} \tag{5.9}$$

Writing "sign of X" = sgn X, and making the change of variables from (c_1, c_2, r') to (l_1', l_2', r') we have that

$$\begin{aligned} j\;df(l_1', l_2', r')\,dl_1'\,dl_2'\,dr' &= \frac{\Gamma(\nu)(\mu_2 l_1'/r' - \lambda_2 l_2'/r')^{s-1}(\lambda_1 l_2'/r' - \mu_1 l_1'/r')^{t-1}}{\Gamma(s)\Gamma(t)\Gamma(\nu - s - t)\begin{vmatrix} \lambda_1 & \mu_1 \\ \lambda_2 & \mu_2 \end{vmatrix}^{\nu-2}} \\ &\cdot \begin{vmatrix} 1 & \dfrac{l_1'}{r'} & \dfrac{l_2'}{r'} \\ 1 & \lambda_1 & \mu_1 \\ 1 & \lambda_2 & \mu_2 \end{vmatrix}^{\nu-s-t-1} \operatorname{sgn}\begin{vmatrix} \lambda_1 & \mu_1 \\ \lambda_2 & \mu_2 \end{vmatrix} [r'^{\nu-1}e^{-r'}/\Gamma(\nu)]\,dl_1'/r'\,dl_2'/r'\,dr'. \end{aligned} \tag{5.10}$$

Transforming from (l_1', l_2', r') to (l_1, l_2, r), integrating out r, and using the expression for A_{mlst} derived in the appendix we conclude that the exact joint density of l_1 and l_2 is given by

$$\begin{aligned} g(l_1, l_2) = \sum_{m,l \in T_{ml}}^{k} \sum_{s=1}^{\nu_m} \sum_{t=1}^{\nu_l} \Bigg\{ & \frac{[\mu'_{m,l;\nu_m-s,\nu_l-t}/(\nu_m - s)!(\nu_l - t)!]\Gamma(\nu)(\mu_l l_1 - \lambda_l l_2)^{s-1}}{\Gamma(s)\Gamma(t)\Gamma(\nu - s - t)\begin{vmatrix} \lambda_m & \mu_m \\ \lambda_l & \mu_l \end{vmatrix}^{\nu_m+\nu_l-2}} \\ & \cdot \frac{(\lambda_m l_2 - \mu_m l_1)^{t-1}\operatorname{sgn}\begin{vmatrix} \lambda_m & \mu_m \\ \lambda_l & \mu_l \end{vmatrix} \cdot \begin{vmatrix} 1 & l_1 & l_2 \\ 1 & \lambda_m & \mu_m \\ 1 & \lambda_l & \mu_l \end{vmatrix}^{\nu-s-t-1}}{\prod_{j \neq m,l}^{k} \begin{vmatrix} 1 & \lambda_j & \mu_j \\ 1 & \lambda_m & \mu_m \\ 1 & \lambda_l & \mu_l \end{vmatrix}^{\nu_j}} \Bigg\} \end{aligned} \tag{5.11}$$

where T_{ml} is the triangular region defined by

$$(5.12)\quad \frac{(\mu_l l_1 - \lambda_l l_2)}{\begin{vmatrix}\lambda_m & \mu_m\\ \lambda_l & \mu_l\end{vmatrix}} \geqq 0, \qquad \frac{(\lambda_m l_2 - \mu_m l_1)}{\begin{vmatrix}\lambda_m & \mu_m\\ \lambda_l & \mu_l\end{vmatrix}} \geqq 0, \quad \text{and } 1 \geqq \frac{\begin{vmatrix}1 & l_1 & l_2\\ 1 & \lambda_m & \mu_m\\ 1 & \lambda_l & \mu_l\end{vmatrix}}{\begin{vmatrix}\lambda_m & \mu_m\\ \lambda_l & \mu_l\end{vmatrix}} \geqq 0$$

provided that $\begin{vmatrix}\lambda_m & \mu_m\\ \lambda_l & \mu_l\end{vmatrix} \neq 0.$

In order to apply (5.12) to the genetic linkage problem, rewrite ϕ_1 and ϕ_2 (defined by (5.2)) as

$$\phi_1 = \lambda_1\theta_1 + \lambda_2\theta_2 + \lambda_3\theta_3 + \lambda_4\theta_4,$$
$$\phi_2 = \mu_1\theta_1 + \mu_2\theta_2 + \mu_3\theta_3 + \mu_4\theta_4,$$

where $(\lambda_1, \lambda_2, \lambda_3, \lambda_4) = (0, \frac{1}{2}, -\frac{1}{2}, 0)$ and $(\mu_1, \mu_2, \mu_3, \mu_4) = (\frac{1}{4}, -\frac{1}{4}, -\frac{1}{4}, -\frac{3}{4})$. For this example, the ratios λ_i/μ_i are not distinct ($\lambda_1/\mu_1 = \lambda_4/\mu_4 = 0$). It is easily verified that if $\lambda_i/\mu_i = \lambda_j/\mu_j$ ($i \neq j$), then the first summation in (5.12) is only over the set of all m, l such that $(m, l) \neq (i, j)$ and $m, l \varepsilon T_{ml}$. Hence from (5.12) we have the exact posterior density of ϕ_1 and ϕ_2. However, it is too involved for practical use.

We can use Watson's [11] generalization of a result due to von Neumann [10] to obtain an integral equation for the joint density. Specifically,

$$(5.13)\quad \textstyle\prod_{j=1}^{k} (1 - it_1\lambda_j - it_2\mu_j)^{-\nu_j} = \iint_R g(l_1, l_2)\, dl_1\, dl_2/(1 - it_1 l_1 - it_2 l_2)^{\nu}$$

where $\nu = \sum_{i=1}^{k} \nu_i$ and R is the convex closure of $(\lambda_1, \mu_1), \cdots, (\lambda_k, \mu_k)$. If the coefficients are such that R can be well approximated by a simple region, then (5.13) may be useful for deriving approximations. The integral equation can also be used to show that for arbitrary λ's and μ's the joint density tends to a bivariate normal distribution as all the $\nu_j \to \infty$. This leads to Lindley's results [7]. Closer approximations can be obtained by expanding $g(l_1, l_2)$ in a bivariate Edgeworth series.

APPENDIX.

To find the values of the constants A_{mlst}, we let

$$y_1 = 1 - it_1\lambda_m - it_2\mu_m - it_3,$$
$$y_2 = 1 - it_1\lambda_l - it_2\mu_l - it_3,$$
$$y_3 = 1 - it_3.$$

Then

$$\begin{aligned}
&(1 - it_1\lambda_j - it_2\mu_j - it_3)^{-\nu_j}\\
&= [y_3[(\lambda_l\mu_m - \mu_l\lambda_m) - (\mu_m - \mu_l)\lambda_j - (\lambda_l - \lambda_m)\mu_j]/[\lambda_l\mu_m - \mu_l\lambda_m]\\
(A1)\quad &+ y_1(\lambda_l\mu_j - \mu_l\lambda_j)/(\lambda_l\mu_m - \mu_l\lambda_m)\\
&+ y_2(\mu_m\lambda_j - \lambda_m\mu_j)/(\lambda_l\mu_m - \mu_l\lambda_m)]^{-\nu_j}\\
&= (y_3\alpha_j(m, l) + y_1\beta_j(m, l) + y_2\gamma_j(m, l))^{-\nu_j}
\end{aligned}$$

where

$$\alpha_j(m, l) = \begin{vmatrix} 1 & \lambda_j & \mu_j \\ 1 & \lambda_m & \mu_m \\ 1 & \lambda_l & \mu_l \end{vmatrix} \cdot \begin{vmatrix} \lambda_m & \mu_m \\ \lambda_l & \mu_l \end{vmatrix}^{-1},$$

$$\text{(A2)} \qquad \beta_j(m, l) = \begin{vmatrix} \lambda_j & \mu_j \\ \lambda_l & \mu_l \end{vmatrix} \cdot \begin{vmatrix} \lambda_m & \mu_m \\ \lambda_l & \mu_l \end{vmatrix}^{-1},$$

$$\gamma_j(m, l) = \begin{vmatrix} \mu_j & \lambda_j \\ \mu_m & \lambda_m \end{vmatrix} \cdot \begin{vmatrix} \lambda_m & \mu_m \\ \lambda_l & \mu_l \end{vmatrix}^{-1}.$$

Substituting formula (A1) into (5.7), equating (5.3) with (5.7) and multiplying through by $y_1^m y_2^l$ we find that

$$\prod_{j\neq m,l}^k (y_3\alpha_j(m, l) + y_1\beta_j(m, l) + y_2\gamma_j(m, l))^{-\nu_j}$$
$$= \sum_{s=1}^{\nu_m} \sum_{t=1}^{\nu_l} A_{mlst} y_1^{\nu_m - s} y_2^{\nu_l - t} y_3^{\Sigma_{i=1}^k \nu_i + s + t}$$
$$+ \sum_{\substack{j\neq m \\ (j<i)}}^k \sum_{i\neq l}^k \sum_{u=1}^{\nu_j} \sum_{v=1}^{\nu_i} A_{jiuv}$$
$$\text{(A3)} \qquad \cdot \{[y_3\alpha_j(m, l) + y_1\beta_j(m, l) + y_2\gamma_j(m, l)]^{-u}$$
$$\cdot [y_3\alpha_i(m, l) + y_1\beta_i(m, l) + y_2\gamma_i(m, l)]^{-v} y_3^{-\Sigma_{j=1}^k \nu_j + u + v}\} y_1^{\nu_m} y_2^{\nu_l}.$$

Setting $y_1 = y_2 = 0$ in (A3) we have that

$$\text{(A4)} \qquad A_{ml\nu_m\nu_l} = \prod_{j\neq m,l}^k [\alpha_j(m, l)^{-\nu_j}].$$

To obtain the remaining coefficients differentiate both sides of (A3) h_1 times with respect to y_1 and h_2 times with respect to y_2 and then put $y_1 = y_2 = 0$. There is no contribution from the second member on the right hand side of (A3). The first term contributes

$$h_1 ! \, h_2 ! \, y_3^{-\nu+\nu_l+\nu_m-h_1-h_2} A_{m,l,\nu_m-h_1,\nu_l-h_2} .$$

Thus

$$y_3^{-\nu+\nu_l+\nu_m-h_1-h_2} A_{m,l,\nu_m-h_1,\nu_l-h_2} = f_{m,l}^{(h_1,h_2)}(0, 0)/h_1 ! \, h_2 !$$

where

$$f_{m,l}(y_1, y_2) = \prod_{j\neq m,l}^k [y_3\alpha_j(m, l) + y_1\beta_j(m, l) + y_2\gamma_j(m, l)]^{-\nu_j}.$$

Rewrite $f_{m,l}(y_1, y_2)$ as

$$f_{m,l}(y_1, y_2) = y_3^{-\nu+\nu_m+\nu_l} \prod_{j\neq m,l}^k [\alpha_j(m, l)]^{-\nu_j}$$
$$\cdot \exp\{-\sum_{j\neq m,l}^k \nu_j \log [1 + y_1\beta_j(m, l)/y_3\alpha_j(m, l) + y_2\gamma_j(m, l)/y_3\alpha_j(m, l)]\}.$$

The t_i's can always be chosen so that $|y_1/y_3| < 1$, $|y_2/y_3| < 1$. Hence we can expand the exponent into a series and equate coefficients. Thus

$$y_3^{-\nu+\nu_m+\nu_l} \sum_{h_1=0}^\infty \sum_{h_2=0}^\infty A_{m,l,\nu_m-h_1,\nu_l-h_2} (y_1/y_3)^{h_1} (y_2/y_3)^{h_2}$$
$$= \sum_{h_1=0}^\infty \sum_{h_2=0}^\infty f_{m,l}^{(h_1,h_2)}(0, 0) y_1^{h_1} y_2^{h_2} / h_1 ! \, h_2 !$$
$$= y_3^{-\nu+\nu_m+\nu_l} \prod_{j\neq m,l}^k [\alpha_j(m, l)]^{-\nu_j}$$
$$\cdot \exp\{\sum_{j\neq m,l}^k \nu_j \sum_{h=1}^\infty [[-\beta_j(m, l)/\alpha_j(m, l)] y_1/y_3$$
$$- [\gamma_j(m, l)/\alpha_j(m, l)] y_2/y_3]^h / h\}.$$

The exponent can be written as $\sum_{j\neq m,l}^{k} \nu_j \sum_{h=1}^{\infty} (h - 1)! \sum_{p=0}^{h} (y_1/y_3)^p/p! \cdot [(y_2/y_3)^{h-p}/(h - p)!][-\beta_j(m, l)/\alpha_j(m, l)]^p[-\gamma_j(m, l)/\alpha_j(m, l)]^{h-p}$. Interchanging the order of summation and setting $h_1 = p$, $h_2 = h - p$ we have that

$$
\begin{aligned}
&y_3^{-\nu+\nu_m+\nu_l} \sum_{h_1=0}^{\infty} \sum_{h_2=0}^{\infty} A_{m,l,\nu_m-h_1,\nu_l-h_2}(y_1/y_3)^{h_1}(y_2/y_3)^{h_2} \\
&\quad = y_3^{-\nu+\nu_m+\nu_l} \prod_{j\neq m,l}^{k} [\alpha_j(m, l)]^{-\nu_j} \\
&\qquad \cdot \exp \{\sum_{h_1=0}^{\infty} \sum_{h_2=0}^{\infty} (h_1 + h_2 - 1)! [(y_1/y_3)^{h_1}/h_1 !][(y_2/y_3)^{h_2}/h_2 !] \sum_{j\neq m,l}^{k} \nu_j \\
&\qquad \cdot [\{-\beta_j(m, l)/\alpha_j(m, l)\}^{h_1}\{-\gamma_j(m, l)/\alpha_j(m, l)\}]^{h_2}\}.
\end{aligned}
$$

Hence the equations expressing the relationship between the bivariate moments through the origin and the bivariate cumulants may be used, i.e. if

$$
\text{(A5)} \quad K_{m,l,h_1,h_2} = (h_1 + h_2 - 1)! \cdot \sum_{j\neq m,l}^{k} \nu_j\{[-\beta_j(m, l)/\alpha_j(m, l)]^{h_1}[-\gamma_j(m, l)/\alpha_j(m, l)]^{h_2}\}
$$

then

$$
\text{(A6)} \qquad A_{m,l,\nu_m-h_1,\nu_l-h_2} = (\prod_{j\neq m,l}^{k} [\alpha_j(m, l)]^{-\nu_j})\mu'_{m,l;h_1,h_2}/h_1 !h_2 ! .
$$

Using (A2) we have from (A4) and (A6) that

$$
\text{(A4)}' \qquad A_{ml\nu_m\nu_l} = \frac{\begin{vmatrix} \lambda_m & \mu_m \\ \lambda_l & \mu_l \end{vmatrix}^{\nu-\nu_j-\nu_m-\nu_l}}{\prod_{j\neq m,l}^{k} \begin{vmatrix} 1 & \lambda_j & \mu_j \\ 1 & \lambda_m & \mu_m \\ 1 & \lambda_l & \mu_l \end{vmatrix}^{\nu_j}}
$$

and

$$
\text{(A6)}' \qquad A_{m,l,\nu_m-h_1,\nu_l-h_2} = (A_{ml\nu_m\nu_l})\mu'_{m,l;h_1,h_2}/h_1 ! \, h_2 !
$$

where $\mu'_{m,l;h_1,h_2}$ is related to

$$
\text{(A5)}' \; K_{m,l;h_1,h_2} = (h_1 + h_2 - 1)! \sum_{j\neq m,l}^{k} \nu_j \left\{ \left[\frac{-\begin{vmatrix} \lambda_j & \mu_j \\ \lambda_l & \mu_l \end{vmatrix}}{\begin{vmatrix} 1 & \lambda_j & \mu_j \\ 1 & \lambda_m & \mu_m \\ 1 & \lambda_l & \mu_l \end{vmatrix}} \right]^{h_1} \left[\frac{-\begin{vmatrix} \mu_j & \lambda_j \\ \mu_m & \lambda_m \end{vmatrix}}{\begin{vmatrix} 1 & \lambda_j & \mu_j \\ 1 & \lambda_m & \mu_m \\ 1 & \lambda_l & \mu_k \end{vmatrix}} \right]^{h_2} \right\}
$$

as before.

REFERENCES

[1] Box, G. E. P. (1954). Some theorems on quadratic forms applied to the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification. *Ann. Math. Statist.* **25** 290–302.

[2] Cramér, H. (1946). *Mathematical Methods of Statistics*, Chapter 17. Princeton Univ. Press.

[3] Durbin, J. and Watson, G. S. (1951). Testing for serial correlation in least squares regression. II. *Biometrics* **38** 159–178.

[4] FISHER, R. A. (1928). *Statistical Methods for Research Workers*, Chapter IX. (second and later editions). Hafner, New York.
[5] KOOPMANS, T. (1942). Serial correlation and quadratic forms in normal variable. *Ann. Math. Statist.* **13** 14–33.
[6] LINDLEY, D. V. (1964). The Bayesian Analysis of Contingency Tables. *Ann. Math. Statist.* **35** 1622–1643.
[7] LINDLEY, D. V. (1965). *Probability and Statistics from a Bayesian point of View,* **2** Chapter 7. Cambridge Univ. Press, England.
[8] MCCARTHY, M. D. (1939). On the application of the z-test to randomized blocks. *Ann. Math. Statist.* **10** 337–359.
[9] PITMAN, E. J. G. (1937). The 'closest' estimates of statistical parameters. *Proc. Cambridge Phil. Soc.* **33** 212–222.
[10] VON NEUMANN, J. (1941). Distribution of the ratio of the mean square successive difference to the variance. *Ann. Math. Statist.* **12** 367–395.
[11] WATSON, G. S. (1956). On the joint distribution of the circular serial correlation coefficients. *Biometrics* **43** 161–168
[12] WATSON, G. S. (1965). Some Bayesian Methods related to χ^2. Technical Report No. 34, Johns Hopkins University.
[13] WILKS, S. S. (1962). *Mathematical Statistics*, Chapter 7. Wiley, New York.