QUADRATIC FORMS IN POISSON AND MULTINOMIAL VARIABLES

P. WHITTLE

(received 2 February 1959, revised 15 July 1959)

Summary: Let $\boldsymbol{\Delta}' = (\Delta_1 \Delta_2 \cdots \Delta_m)$ represent the deviations from expectation of a set of multinomial or independent Poisson variables, and $\boldsymbol{H}$ be a positive definite matrix. A lower bound is obtained for $Pr(\boldsymbol{\Delta}'\boldsymbol{H}\boldsymbol{\Delta} \leqq S)$ in terms of $Pr(\boldsymbol{\delta}'\boldsymbol{H}\boldsymbol{\delta} \leqq S)$, where $\boldsymbol{\delta}$ is a vector of normal variables with the same mean and covariance matrix as $\boldsymbol{\Delta}$.

1. One often has occasion to consider the distribution of a positive definite quadratic form in multinomial variables, the commonest instance being that of the χ^2 statistic. If the variables are denoted $\Delta_1, \Delta_2 \cdots \Delta_m$, then another instance is provided by the evaluation of probabilities of the type $Pr(|\sum_k a_{jk}\Delta_k| \leqq \alpha_j;\, j = 1, 2 \cdots p)$, because for suitably chosen $\boldsymbol{H}$ this probability is always greater than $Pr(\boldsymbol{\Delta}'\boldsymbol{H}\boldsymbol{\Delta} \leqq 1)$, which may be easier to calculate.

In many cases, that of the χ^2 statistic being the classic example, one obtains an approximation to the distribution function of $\boldsymbol{\Delta}'\boldsymbol{H}\boldsymbol{\Delta}$ by assuming the Δ_j to be normally distributed, with appropriate first and second moments. The order of magnitude of the error entailed by this assumption has been considered by several authors (see, for example, references 1, 2 and 3): in this paper I intend to calculate an explicit lower bound for $Pr(\boldsymbol{\Delta}'\boldsymbol{H}\boldsymbol{\Delta} \leqq S)$ which, while not as close as it might be (see the discussion in section 2) nevertheless gives a definite bound of the correct asymptotic form.

I consider first the case of a set of Poisson variables, since this is very similar to the multinomial case, but substantially simpler in one respect.

THEOREM: *Let $n_j (j = 1, 2 \cdots m)$ denote a set of independent Poisson variables with $E(n_j) = \lambda_j$, let $\boldsymbol{H} = (h_{jk})$ be a positive definite matrix, and let*

$$\Delta_j = n_j - \lambda_j \qquad (j = 1, 2 \cdots m) \tag{1}$$

Then

$$Pr(\boldsymbol{\Delta}'\boldsymbol{H}\boldsymbol{\Delta} \leqq S) > \prod_1^m \left[\frac{e^{-S/(24\lambda_j g)}}{(1 + (S/\lambda_j g)^{1/2})^{1/2}}\right] Pr(\boldsymbol{\delta}'\boldsymbol{H}\boldsymbol{\delta} \leqq \mu S) \tag{2}$$

provided

$$mh/4 \leqq S \leqq g \min_j \lambda_j \tag{3}$$

where h, g denote respectively the greatest eigenvalue of $\boldsymbol{H}$ and the least eigenvalue of $\boldsymbol{G} = (h_{jk}\sqrt{\lambda_j\lambda_k})$,

$$\mu = [1 - (mh/4S)^{1/2}]^2 \tag{4}$$

and $\boldsymbol{\delta}$ is a random normal vector with mean $E(\boldsymbol{\Delta})$ and covariance matrix $E(\boldsymbol{\Delta\Delta}')$.

THEOREM: *Let $n_j (j = 1, 2 \cdots m)$ denote a set of multinomial variables with $E(n_j) = Np_j$, $\Sigma p_j = 1$; let $\boldsymbol{H}$ be a positive definite matrix, and let*

$$\Delta_j = n_j - Np_j \tag{5}$$

Then

$$Pr(\boldsymbol{\Delta}'\boldsymbol{H}\boldsymbol{\Delta} \leqq S) > \frac{e^{-mS/(8Ng \min p_j)}}{\prod_j (1 + (S/Ngp_j)^{1/2})^{1/2}} Pr(\boldsymbol{\delta}'\boldsymbol{H}\boldsymbol{\delta} \leqq \mu S) \tag{6}$$

provided

$$mh/4 \leqq S \leqq Ng \min_j p_j \tag{7}$$

where h, g denote respectively the greatest eigenvalue of $\boldsymbol{H}$ and the least eigenvalue of $\boldsymbol{G} = (Nh_{jk}\sqrt{p_jp_k})$,

$$\mu = [1 - (mh/4S)^{1/2}]^2 \tag{8}$$

and $\boldsymbol{\delta}$ is a random normal vector with mean $E(\boldsymbol{\Delta})$ and covariance matrix $E(\boldsymbol{\Delta\Delta}')$.

Thus, for the case of the χ^2 statistic

$$\chi^2 = \sum \Delta_j^2/Np_j \tag{9}$$

we have

$$Pr(\chi^2 \leqq S) > \frac{e^{-mS/8N_p}}{\prod_j \left(1 + \left(\frac{S}{Np}\right)^{1/2}\right)^{1/2}} Pr(x_{m-1} \leqq (S^2 - (m/4Np)^2)^{1/2}) \tag{10}$$

provided

$$m/4Np \leqq S \leqq Np \tag{11}$$

where p is the least of the p_j, and x_{m-1} has a χ^2 distribution function with $(m - 1)$ degrees of freedom.

2. The two sources of error in the normal approximation are, first, that the non-zero ordinates of the joint $\boldsymbol{\Delta}$ distribution are not truly proportional to normal ordinates, and second, that the distribution is a discrete one. Bounds for the first type of deviation are relatively easily obtained; I have tried, however, to find bounds which are simple rather than close, since it is actually the second type of error which limits the accuracy of the estimate.

The most serious consequence of the discrete nature of the distribution is an edge-effect: the probability content of the region

$$\boldsymbol{\Delta}'\boldsymbol{H}\boldsymbol{\Delta} \leqq S \tag{12}$$

is not a continuous function of S. One would have to use advanced methods to obtain an adequate treatment of this effect, methods similar to those used to estimate the number of lattice points in an ellipsoid (see ref. 4), a problem not yet fully treated. The crude method used in the present paper (section 5) yields an estimate of order $N^{-\frac{1}{4}}$ for the error due to edge-effect; it seems likely that the actual order is $N^{-m/2}$, at least for the Poisson case.

It will be noted that in the above two theorems the quantity g, the least eigenvalue of $(h_{jk}\sqrt{\lambda_j\lambda_k})$ or its multinomial counterpart, plays a role. This would be a weakness if there were any considerable spread in the eigenvalues of this matrix, i.e., if the surfaces $\boldsymbol{\Delta}'\boldsymbol{H}\boldsymbol{\Delta} = \text{const.}$, $\sum \Delta_j^2/\lambda_j = \text{const.}$, were considerably different in form. The difficulty could be overcome by exploiting the fact that the part of the ellipsoid (12) which lies outside some particular contour $\sum \Delta_j^2/\lambda_j = \text{const.}$ will make little contribution to the total probability content of (12). However, I have not thought it worthwhile to introduce the extra modification here.

3. Consider a set of m independent Poisson variables n_j with means λ_j $(j = 1, 2, \cdots m)$. If

$$\Delta_j = n_j - \lambda_j \tag{13}$$

we wish to calculate a lower bound to the probability that relation (12) is fulfilled, i.e. we wish to calculate the sum of all terms

$$F(\boldsymbol{\Delta}) = \prod_j f(n_j) = \prod_j \frac{e^{-\lambda_j}\lambda_j^{n_j}}{n_j!} \tag{14}$$

for which the representative point $\boldsymbol{\Delta}$ lies inside the solid ellipsoid (12).

We shall suppose that

$$\Delta_j \leqq \theta\sqrt{\lambda_j} \leqq \lambda_j \qquad (j = 1, 2 \cdots m) \tag{15}$$

Inequality (12) will certainly be consistent with the first inequality of (15) if

$$S \leqq \theta^2 g \tag{16}$$

The simplest way to ensure the validity of (16) is to use the equality sign in this relation to define θ. The second inequality of (15) will then be fulfilled if

$$S \leqq g \min_j \lambda_j. \tag{17}$$

Now

$$\begin{aligned}\log f(n) &= -\lambda + n\log\lambda - \log n! \\ &> -\lambda + n\log\lambda - \tfrac{1}{2}\log 2\pi + n - (n+\tfrac{1}{2})\log n\end{aligned} \tag{18}$$

since the remainder in Stirling's approximation for $\log n!$ is positive. Expanding expression (18) in powers of $\Delta = n - \lambda$ we have then

$$\begin{aligned}\log f(n) &> -\tfrac{1}{2}\log 2\pi(\lambda+\Delta) + \sum_{j=1}^{\infty}\frac{\Delta^{j+1}}{j(j+1)(-\lambda)^j} \\ &> -\tfrac{1}{2}\log 2\pi(\lambda+\theta\lambda^{1/2}) - \frac{\Delta^2}{2\lambda}.\end{aligned} \tag{19}$$

It thus follows that

$$F(\boldsymbol{\Delta}) > \frac{e^{-\frac{1}{2}\sum \Delta_j^2/\lambda_j}}{(2\pi)^{m/2}\prod_j(\lambda_j+\theta\lambda_j^{1/2})^{1/2}}. \tag{20}$$

4. We wish now to relate the discrete probability distribution $F(\boldsymbol{\Delta})$ to a continuous normal density.

Let us associate with the point $\boldsymbol{\Delta}$ (corresponding to a particular set of integers n_j) the rectangular cell Ω_Δ whose vertices have coordinates $\Delta_1 \pm \frac{1}{2}$, $\Delta_2 \pm \frac{1}{2} \cdots \Delta_m \pm \frac{1}{2}$. These cells fill the whole $\boldsymbol{\Delta}$-space simply. We shall now prove that

$$F(\boldsymbol{\Delta}) > c\int_{\Omega_\Delta}\frac{e^{-\sum u_j^2/2\lambda_j}\,du}{\prod_1^m(2\pi\lambda_j)^{1/2}} \tag{21}$$

where

$$c = \prod_j \frac{e^{-\theta^2/24\lambda_j}}{(1+\theta/\lambda_j^{1/2})^{1/2}} \tag{22}$$

We have

$$\int_{\Omega_\Delta} e^{-\sum u_j^2/2\lambda_j}\,du = \prod_j \int_{-1/2}^{1/2} e^{-(\Delta_j+v)^2/2\lambda_j}\,dv \tag{23}$$

and

$$\begin{aligned}\int_{-1/2}^{1/2} e^{-(\Delta+v)^2/2\lambda}\,dv &\gtreqless e^{-\Delta^2/2\lambda}\int_{-1/2}^{1/2} e^{-\Delta v/\lambda}\,dv \\ &= e^{-\Delta^2/2\lambda}\sum_{j=0}^{\infty}\frac{(\Delta/2\lambda)^{2j}}{(2j+1)!} \gtreqless e^{-\Delta^2/2\lambda+\Delta^2/24\lambda^2} \gtreqless e^{-\Delta^2/2\lambda+\theta^2/24\lambda}.\end{aligned} \tag{24}$$

https://doi.org/10.1017/S1446788700025581 Published online by Cambridge University Press

Combining equations (20), (23) and (24) we obtain the desired result (21).

5. Let us denote the solid ellipsoid (12) by E. Relation (21) enables us to replace the summation of $F(\Delta)$ over all appropriate points Δ in E by the integration of a normal density over E. There will be an edge-effect, however, due to the fact that the surface of E will cut through a number of cells, some of whose centres Δ will lie in E, and some of whom will not.

We can obtain a conservative result if we take a smaller region E' which just excludes any cells which were cut by the surface of E, but whose centres did not lie in E. We are thus led to consider a region

$$\Delta' H \Delta \lesseqgtr \mu S \tag{25}$$

where μ is chosen so that the least radius of the ellipsoid has contracted by an amount equal to the greatest radius of a cell Ω_Δ (which will be half the principal diagonal, or $\sqrt{m}/2$). That is,

$$\sqrt{\frac{S}{h}} - \sqrt{\frac{\mu S}{h}} = \frac{\sqrt{m}}{2} \tag{26}$$

a relation equivalent to (4) if

$$S \gtreqless \frac{mh}{4}. \tag{27}$$

Extending the integration in (21) over the whole region (25), and setting $\theta = \sqrt{S/g}$, we obtain the result expressed in the first theorem.

6. The calculation for a multinomial distribution is largely analogous. In this case the inequality corresponding to (20) is

$$F(\Delta) = N! \prod_j \frac{p_j^{n_j}}{n_j!} > \frac{e^{-\Sigma \Delta_j^2/2Np_j}}{(2\pi N)^{\frac{m-1}{2}} \prod (p_j + \theta(p_j/N)^{1/2})^{1/2}}. \tag{28}$$

The conditions

$$\theta^2 = S/g \tag{29}$$

$$S \gtreqless Ng \min_j p_j \tag{30}$$

are certainly sufficient for the validity of (28).

A complication of the multinomial case is that the variables are constrained by the relation

$$\sum_1^m \Delta_j = 0. \tag{31}$$

It will be convenient to take coordinates in the hyperplane (31); we shall show in fact that (28) can be written

$$F(\Delta) > \left[\frac{m}{\prod_j (1 + \theta/(Np_j)^{1/2})}\right]^{1/2} \frac{e^{-\sum_1^{m-1} v_j^2/2N\rho_j}}{\prod_1^{m-1} (2\pi N \rho_j)^{1/2}} \tag{32}$$

https://doi.org/10.1017/S1446788700025581 Published online by Cambridge University Press

where $y_1, y_2 \cdots y_{m-1}$ constitute a set of orthogonal coordinates in the hyperplane (31) (and so together with $y_m = \sum \Delta_j / m^{1/2}$ are derived from the Δ_j by an orthogonal transformation).

We obtain the principal axes of the quadric formed by the section of $\sum \Delta_j^2 / N p_j =$ const. by the hyperplane (31) if we note that the stationary values ϕ of a quadratic form $\boldsymbol{\Delta}' \boldsymbol{A} \boldsymbol{\Delta}$, given $\boldsymbol{\Delta}' \boldsymbol{\Delta} = 1$, $\boldsymbol{\beta}' \boldsymbol{\Delta} = 0$, satisfy the relation

$$\boldsymbol{\beta}' \operatorname{adj} (\boldsymbol{A} - \phi \boldsymbol{I}) \boldsymbol{\beta} = 0 \tag{33}$$

as may be shown by a direct minimisation of the form, using Lagrange multipliers to allow for the two side-conditions.

For the case $\boldsymbol{\beta} = (1, 1 \cdots 1)$, $\boldsymbol{A} = \operatorname{diag.}(1/Np_1 \cdots 1/Np_m)$ we have then

$$\sum \frac{1}{\phi - \dfrac{1}{Np_j}} = 0 \tag{34}$$

or, for $\phi = 1/N\rho$,

$$\sum_{j=1}^{m} p_j \prod_{k \neq j} (\rho - p_k) = 0. \tag{35}$$

Now, the exponents of the two expressions (28), (32) are certainly equal, since $\sum y_j^2 / \rho_j$ is nothing but the standard form for the quadratic $\sum \Delta_j^2 / p_j$ on (31). To prove the equality of the remaining parts of the two expressions, we note from equation (35), an equation for ρ with roots $\rho_1, \rho_2 \cdots \rho_{m-1}$, that

$$\prod_{1}^{m-1} \rho_j = m \prod_{1}^{m} p_j \tag{36}$$

from which the equivalence of (28), (32) follows.

Another consequence of relation (31) is that it is not now such a direct matter to associate a cell with the representative points $\boldsymbol{\Delta}$ (or $\boldsymbol{y}$). The final integration must be restricted to the hyperplane (31), and so the cell associated with the point must also lie wholly in the hyperplane. However, the lattice formed by the representative points $\boldsymbol{\Delta}$ (or $\boldsymbol{y}$) in the $(m-1)$-space (31) is not a rectangular one, and the appropriate division of the space into cells not so obvious.

Probably the simplest way of dividing the space is to take the m-dimensional unit hypercube which has $\boldsymbol{\Delta}$ as centre, project it orthogonally on to the hyperplane (31), and take the outline of the figure thus formed as the cell to be associated with the corresponding $\boldsymbol{y}$. These cells cover the space (31) simply, and in fact form a regular honeycomb in the space. The cells may be regarded as being constructed of m $(m-1)$-dimensional unit hypercubes (the faces of the m-dimensional hypercube which can be seen from one vertex) foreshortened along a principal diagonal in the ratio $1/\sqrt{m}$ (the cosine of the angle between planes $\Delta_j = \text{const.}$, $\sum \Delta_j = \text{const.}$) and fitted together at the "blunt" vertices so as to fill the whole solid angle. Thus, for $m = 3$, the cell is a hexagon, composed of three foreshortened squares. For $m = 4$ it is a rhombic dodecahedron, composed of four foreshortened cubes.

The metric content of the cell is $m/\sqrt{m} = \sqrt{m}$, and the maximum radius is not greater than $\sqrt{m}/2$.

If the cell associated with y is denoted W_y, we shall now prove that

$$F(\Delta) > c' \int_{W_y} \frac{e^{-\sum_1^{m-1} u_j^2/2N\rho_j} \mathrm{d}u}{\prod_1^{m-1} (2\pi N\rho_j)^{\frac{1}{2}}} \tag{37}$$

where

$$c' = \frac{e^{-m\theta^2/(8N \min p_j)}}{\prod_j (1 + \theta/(Np_j)^{\frac{1}{2}})^{\frac{1}{2}}}. \tag{38}$$

We have, as in (24)

$$\int_{W_y} e^{-\sum u_j^2/2N\rho_j} \mathrm{d}u \gtreqless e^{-\sum y_j^2/2N\rho_j} \int_{W_0} e^{-\sum y_j v_j/N\rho_j} \mathrm{d}v. \tag{39}$$

Since the function of v, $\exp[-\sum y_j v_j/N\rho_j]$, is convex, its mean value over W_0 will be less than its mean value over the circumscribed hypersphere of radius $R = \sqrt{m}/2$. This observation leads to the inequality

$$\begin{aligned}\int_{W_0} e^{-\sum y_j v_j/N\rho_j} \mathrm{d}v &< \sqrt{m} \sum_{k=0}^{\infty} \frac{(DR)^{2k} \Gamma\left(\frac{m+1}{2}\right)}{2^{2k} k! \, \Gamma\left(\frac{m+1}{2} + k\right)} \\ &< \sqrt{m}\, e^{(DR)^2/2(m+1)} < \sqrt{m}\, e^{\sum y_j^2/8N^2\rho_j^2}\end{aligned} \tag{40}$$

where

$$D^2 = \sum y_j^2/N^2 \rho_j^2 . \tag{41}$$

Now

$$\begin{aligned}\sum y_j^2/N^2 \rho_j^2 &\gtreqless \max\left(\frac{1}{N\rho_j}\right) \sum y_j^2/N\rho_j \\ &\gtreqless \max\left(\frac{1}{Np_j}\right) \sum \Delta_j^2/Np_j \\ &\gtreqless \max\left(\frac{1}{Np_j}\right) m\theta^2.\end{aligned} \tag{42}$$

https://doi.org/10.1017/S1446788700025581 Published online by Cambridge University Press

Combining relations (32), (39), (40) and (42) we obtain the desired inequality (37). This inequality is the one analogous to (21), and all that remains now is to evaluate the edge-effect, which is done exactly as before.

By taking account of restriction (31) we could have replaced h and g by quantities respectively smaller and larger. However, this improvement will only occasionally be significant.

References

[1] Cochran, W. G. (1952). The χ^2 test of goodness of fit. Ann. Math. Stat., 23, 315—45.
[2] Darwin, J. H. (1958). On corrections to the χ^2 distribution. J. Roy. Stat. Soc., B. 20, 387—392.
[3] Hoel, P. G. (1938). On the Chi-square distribution for small samples. Ann. Math. Stat., 9, 158—65.
[4] Wilton, J. R. (1929). The lattice points of the circle: an historical account of the problem. Messenger of Math., 58, 67—80.

Applied Mathematics Laboratory, New Zealand D.S.I.R. (present address: Statistical Laboratory, Cambridge University).