SOME MODIFIED RECURSION FORMULAE FOR COMPUTING THE DISTRIBUTION OF A LINEAR COMBINATION OF CHI-SQUARES

BU-556-M by May, 1975

R. H. Brown

Abstract

This note presents some modifications to an algorithm given by Johnson and Kotz for computing the distribution of a linear combination of central chi-squares. The modifications result in less storage requirements and improved computational efficiency. The extension of the algorithm to the non-central case is also presented.

Paper No. BU-556-M in the Biometrics Unit Mimeo Series, Department of Plant Breeding and Biometry, Cornell University, Ithaca, New York 14853.

SOME MODIFIED RECURSION FORMULAE FOR COMPUTING THE DISTRIBUTION OF A LINEAR COMBINATION OF CHI-SQUARES

BU-556-M by May, 1975

R. H. Brown

1. Introduction

In many practical situations the statistics used in testing or estimation procedures can be expressed as quadratic forms in jointly normally distributed random variables where the matrix of the quadratic form is not idempotent. Examples of such situations include inferences on variance components, certain variable selection problems, and synthesis of error terms in unbalanced data.

The problem of computing the cumulative distribution function of these statistics is complicated. A convenient approximation to the c.d.f. for positive definite quadratic forms is given by Jensen and Solomon [2] who also review various other approximations. Algorithms for the computation of the exact c.d.f. are given by Imhof [1] for general non-definite forms, and Johnson and Kotz [3] for positive definite central forms.

This note presents some modifications to the recursion formulae given by Johnson and Kotz which alleviate some of the computational disadvantages of their algorithm. The extension of the algorithm to the non-central case is also provided and some results on the stability of the modified formulae are presented.

2. The Central Case

Let $Q = \underset{\sim}{Y}'A\underset{\sim}{Y}$ where $\underset{\sim}{Y}$ is an $n \times 1$ vector of random variables with a multivariate normal distribution with mean vector $\underset{\sim}{0}$ and variance matrix $\Sigma = T'T$ which

---

Paper No. BU-556-M in the Biometrics Unit Mimeo Series, Department of Plant Breeding and Biometry, Cornell University, Ithaca, New York 14853.

is assumed positive definite. Then

$$Q = \sum_{i=1}^{n} \alpha_i z_i^2 \tag{2.1}$$

where $\alpha_1 \geq \cdots \geq \alpha_n$ are the characteristic roots of TAT' and $z_i = \underline{p}_i' T'^{-1} \underline{Y}$ where $\underline{p}_i$ is the characteristic vector of TAT' associated with α_i. The z_i are i.i.d. N(0,1). Based on earlier analytic work (Kotz, Johnson, and Boyd [4]) Johnson and Kotz give the c.d.f. of Q as

$$F(y) = \Pr(Q \leq y) \tag{2.2}$$

$$= \Pr(\chi_n^2 \leq y) + \frac{(y/2\beta)^{n/2} e^{-y/2\beta}}{\Gamma(n/2 + 1)} \cdot \sum_{j=1}^{\infty} c_j \ell_j^{(n+2)}(y/2\beta)$$

where χ_n^2 denotes a central chi-squared random variable with n degrees of freedom and

$$\ell_j^{(n)}(x) = {}_1F_1(-j; \tfrac{1}{2}n; x) = \frac{\Gamma(n/2) \cdot j!}{\Gamma(n/2 + j)} \cdot L_j^{(n/2-1)}(x) \tag{2.3}$$

with $L_j^{(a)}(x)$ being the generalized Laguerre polynomial of degree j,

$$\beta = (\alpha_1 + \alpha_n)/2 \tag{2.4}$$

$$c_0 = 1, \; c_j = j^{-1} \cdot \sum_{i=0}^{j-1} d_{j-i} \cdot c_i, \; j = 1, 2, \cdots \tag{2.5}$$

with

$$d_k = \tfrac{1}{2} \cdot \sum_{i=1}^{n} (1 - \alpha_i/\beta)^k = \tfrac{1}{2} \cdot \sum_{i=1}^{n} \gamma_i^k \, .$$

Let $t = y/\beta$. Then the Johnson and Kotz algorithm computes the $\ell_j^{(n+2)}(t/2)$ from the recursion

$$p_0 = 1,\ p_1 = 1 - t/n,$$

$$p_j = \ell_j^{(n)}(t/2) = \left(2 - \frac{n + t}{n + 2 \cdot j - 2}\right) \cdot p_{j-1} - \left(1 - \frac{n}{n + 2 \cdot j - 2}\right) \cdot p_{j-2}$$

and computes the c_j directly from (2.5). The summation is carried out until two successive terms are less than some prescribed tolerance.

The use of (2.5) to compute the c_j implies two related disadvantages:

(a) All of the previously computed c_j and d_j must be held in storage. This means that two work vectors are required, the dimension of which depends on the number of partial sums required for convergence. This is generally much greater than n, the dimension of $\underline{Y}$.

(b) The number of operations required to compute c_j is a function of j. As more coefficients are computed, more and more operations are needed to compute the next coefficient.

The following modifications of (2.5) overcome both of these disadvantages. We have from (2.5):

$$j \cdot c_j = \sum_{i=0}^{j-1} d_{j-i} \cdot c_i$$

$$= \sum_{s=1}^{j} d_s c_{j-s}$$

$$= \tfrac{1}{2} \cdot \sum_{s=1}^{j} \left(\sum_{r=1}^{n} \gamma_r^s \right) \cdot c_{j-s}$$

$$= \tfrac{1}{2} \cdot \sum_{r=1}^{n} \left(\sum_{s=1}^{j} \gamma_r^s \cdot c_{j-s} \right)$$

$$= \tfrac{1}{2} \cdot \sum_{r=1}^{n} b_{r,j} \quad \text{with} \quad b_{r,j} = \sum_{s=1}^{j} \gamma_r^s \cdot c_{j-s} \, .$$

Similarly,

$$(2.6) \qquad (j+1) \cdot c_{j+1} = \tfrac{1}{2} \cdot \sum_{r=1}^{n} b_{r,j+1}$$

$$\text{with} \quad b_{r,j+1} = \sum_{s=1}^{j+1} \gamma_r^s \cdot c_{j+1-s}$$

$$= \sum_{i=0}^{j} \gamma_r^{i+1} \cdot c_{j-i}$$

$$= \gamma_r \cdot c_j + \gamma_r \cdot \sum_{i=1}^{j} \gamma_r^i \cdot c_{j-i}$$

$$(2.7) \qquad = \gamma_r \cdot (c_j + b_{r,j}).$$

There are only n of the $b_{r,j}$'s which are updated (i.e., overwritten) recursively using (2.7), and then c_{j+1} computed using (2.6). Only the last coefficient needs to be saved rather than all of the previous coefficients.

A comparison of the number of multiplications required by the two methods illustrates the advantages of equations (2.6) and (2.7). Using (2.5) directly, n multiplications are required in order to compute d_{j+1} with j additional multiplications needed in order to form the inner product of the d's and c's. With

equation (2.7), n multiplications are needed to update the $b_{r,j}$'s. No additional multiplications are needed in equation (2.6). The d_j's are never computed explicitly and hence do not require any storage. Clearly, as j increases the accumulated savings become quite significant.

3. The Non-Central Case

As before, let $Q = \underline{Y}'A\underline{Y}$ where $\underline{Y}$ is an n × 1 vector of random variables with a multivariate normal distribution. However, now the mean vector of $\underline{Y}$ is $\underline{\mu} \neq \underline{0}$ and the variance matrix of $\underline{Y}$ is $\Sigma = T'T$ which is positive definite. Then

$$(3.1) \qquad Q = \sum_{i=1}^{n} \alpha_i (z_i + \delta_i)^2$$

where $\alpha_1 \geq \cdots \geq \alpha_n$ are the characteristic roots of TAT', $z_i = \underline{p}_i'(T'^{-1}\underline{Y} - T'^{-1}\underline{\mu})$, $\delta_i = \underline{p}_i'T'^{-1}\underline{\mu}$, where $\underline{p}_i$ is the characteristic vector of TAT' associated with α_i. The z_i are i.i.d. N(0,1).

Kotz, Johnson, and Boyd [5] give the c.d.f. of Q by the expansion (2.2) where the coefficient c_j are now determined by

$$(3.2) \qquad c_0 = 1, \; c_j = j^{-1} \cdot \sum_{i=0}^{j-1} d_{j-i} \cdot c_i$$

$$\text{with} \quad d_k = -\tfrac{1}{2}k \cdot \sum_{i=1}^{n} \delta_i^2 \cdot \gamma_i^{k-1} + \tfrac{1}{2} \cdot \sum_{i=1}^{n} (1 + k \cdot \delta_i^2)\gamma_i^k$$

$$\text{where} \quad \gamma_i = (1 - \alpha_i/\beta).$$

Algebra similar to that preceding (2.6) gives:

$$(3.3) \qquad j \cdot c_j = \tfrac{1}{2} \cdot \sum_{r=1}^{n} b_{r,j} + \tfrac{1}{2} \cdot \sum_{r=1}^{n} \gamma_r \cdot f_{r,j} - \tfrac{1}{2} \cdot \sum_{r=1}^{n} f_{r,j}$$

where $b_{r,j}$ is given by (2.6) and

$$f_{r,j} = \delta_r^2 \cdot \sum_{i=1}^{j} i \cdot \gamma_r^{i-1} \cdot c_{j-i} \cdot$$

We have also that:

$$(j + 1) \cdot c_{j+1} = \tfrac{1}{2} \cdot \sum_{r=1}^{n} b_{r,j+1} + \tfrac{1}{2} \cdot \sum_{r=1}^{n} \gamma_r \cdot f_{r,j+1} - \tfrac{1}{2} \cdot \sum_{r=1}^{n} f_{r,j+1} \cdot$$

Equation (2.7) can be used to update the $b_{r,j}$'s. For the $f_{r,j}$'s, we have:

$$f_{r,j+1} = \delta_r^2 \cdot \sum_{i=1}^{j+1} i \cdot \gamma_r^{i-1} \cdot c_{j+1-i}$$

$$= \delta_r^2 \cdot \sum_{s=0}^{j} (s + 1) \cdot \gamma_r^s \cdot c_{j-s}$$

$$= \delta_r^2 \cdot \sum_{s=0}^{j} \gamma_r^s \cdot c_{j-s} + \gamma_r \cdot \delta_r^2 \cdot \sum_{s=1}^{j} s \cdot \gamma_r^{s-1} \cdot c_{j-s}$$

$$= \delta_r^2 \cdot (c_j + b_{r,j}) + \gamma_r \cdot f_{r,j}$$

$$= \delta_r^2 \cdot (c_j + b_{r,j}) + f_{r,j+\frac{1}{2}} \cdot \tag{3.4}$$

Then the algorithm to compute c_j from the $b_{r,j-1}$'s and the $f_{r,j-\frac{1}{2}}$'s would be as follows:

(i) Compute $b_{r,j}$ using (2.7).

(ii) Compute $f_{r,j}$ using (3.4).

(iii) One $= \left(\sum_{r=1}^{n} b_{r,j}\right) - \left(\sum_{r=1}^{n} f_{r,j}\right)$.

(iv) Compute $f_{r,j+\frac{1}{2}} = \gamma_r \cdot f_{r,j}$.

(v) $\text{Two} = \sum_{r=1}^{n} f_{r,j+\frac{1}{2}}$.

(vi) $c_j = (\text{one} + \text{two})/(2 \cdot j)$.

4. A Simple Error Analysis

Assume that at the k^{th} stage the coefficients c_i have been computed such that $c_i = c_i' + \epsilon_i$, $i = 0, 1, \cdots, k-1$, where c_i' is the true coefficient. Using (2.5), c_k would be evaluated as:

$$c_k = \sum_{i=0}^{k-1} d_{k-i'} \cdot c_i$$

$$= \sum_{i=0}^{k-1} d_{k-i} \cdot (c_i' + \epsilon_i)$$

$$= \sum_{i=0}^{k-1} d_{k-i} \cdot c_i' + \sum_{i=0}^{k-1} d_{k-i} \cdot \epsilon_i$$

$$= c_k' + \sum_{i=0}^{k-1} d_{k-i} \cdot \epsilon_i .$$

Using (2.6) and (2.7), we have:

$$c_k = \sum_{r=1}^{n} b_{r,k}$$

$$= \sum_{r=1}^{n} \gamma_r \cdot (c_{k-1} + b_{r,k-1})$$

$$= \sum_{r=1}^{n} \gamma_r \cdot (c'_{k-1} + \epsilon_{k-1} + b_{r,k-1})$$

$$= \sum_{r=1}^{n} \gamma_r \cdot (c'_{k-1} + b_{r,k-1}) + \epsilon_{k-1} \sum_{r=1}^{n} \gamma_r$$

$$= c'_k + \epsilon_{k-1} \cdot d_1 \cdot$$

In other words, using (2.6) and (2.7) the error in the computed value of c_k is a function of the error in only the last coefficient, whereas using (2.5) involves an error which is a linear combination of the errors in all the previous coefficients. The same would be true in the non-central case.

The above comments neglect errors in the d_r's and the $b_{r,k}$'s. However, it is clear from (2.5) that the d_r's would be computed as

$$d_r = \sum_{r=1}^{n} \gamma_r \cdot d_{r,k-1} \quad \text{where} \quad d_{r,k-1} = \gamma_r^{k-1}.$$

This is essentially the same recurrence as used in (2.7) and it then seems reasonable to assume that the $d_{r,k}$'s and the $b_{r,k}$'s are computed with approximately the same relative precision.

5. Some Refinements

A considerable increase in efficiency in terms of both storage requirements and number of operations can be achieved if the number of distinct roots of TAT' is much less than n. If there are k distinct roots occurring with multiplicities m_i, i = 1, 2, ···, k, then there are only k distinct γ_r's and equation (2.6) becomes:

$$(5.1) \qquad (j+1) \cdot c_{j+1} = \tfrac{1}{2} \cdot \sum_{r=1}^{k} m_r \cdot b_{r,j+1} \cdot$$

Only the distinct $b_{r,j}$'s are required in storage (along with the corresponding m_r) and are updated in the same way as before. A similar modification to (3.3) gives:

$$(5.2) \qquad j \cdot c_j = \tfrac{1}{2} \cdot \sum_{r=1}^{k} m_r \cdot b_{r,j} + \tfrac{1}{2} \cdot \sum_{r=1}^{k} \gamma_r \cdot f_{r,j} - \tfrac{1}{2} \cdot \sum_{r=1}^{k} f_{r,j} \cdot$$

Suppose that it is desired to evaluate

$$(5.3) \qquad \Pr\Big[\sum_{i=1}^{n} \alpha_i \cdot (z_i + \delta_i)^2 \le \sum_{j=1}^{m} \beta_j \cdot (w_j + \lambda_j)^2\Big]$$

where the z_i and w_i are i.i.d. $N(0,1)$. Writing this as

$$E_{\underline{W}}\Big\{\Pr\Big[\sum_{i=1}^{n} \alpha_i \cdot (z_i + \delta_i)^2 \le \sum_{j=1}^{m} \beta_j \cdot (x_j + \lambda_j)^2\Big] \Big| w_j = x_j,\ j = 1, \cdots, m\Big\},$$

the preceding algorithm can be used at each point of a quadrature formula to numerically evaluate (5.3). Noting that the coefficients c_j depend only on the α_i and the δ_i and not on the point at which the c.d.f. is evaluated, it is not necessary to recompute them at every point of the quadrature if they can be held in storage. If m is small, this can be expected to be reasonably efficient. In particular if there is only one distinct value of β_j occurring with multiplicity m, the modifications presented above lead to an efficient algorithm for computing the c.d.f. of the doubly non-central F statistic.

References

[1] Imhof, J. P. (1961). Computing the distribution of quadratic forms in normal variables. Biometrika 48:419-426.

[2] Jensen, D. R. and Solomon, H. (1972). A Gaussian approximation to the distribution of a definite quadratic form. JASA 67:898-902.

[3] Johnson, N. L. and Kotz, S. (1968). Tables of distributions of positive definite quadratic forms in central normal variables. Sankhyā Series B, 30:303-314.

[4] Kotz, Samuel, Johnson, N. L., and Boyd, D. W. (1967a). Series representations of distributions of quadratic forms in normal variables. I. Central Case. Ann. Math. Statist. 38:823-837.

[5] Kotz, Samuel, Johnson, N. L., and Boyd, D. W. (1967b). Series representations of distributions of quadratic forms in normal variables. II. Non-Central Case. Ann. Math. Statist. 38:838-848.