Ann. Inst. Statist. Math.
33 (1981), Part A, 385-390

ON POSITIVE DEFINITE QUADRATIC FORMS IN CORRELATED t VARIABLES

ULRICH MENZEFRICKE

(Received Nov. 25, 1980)

Summary

In this paper we extend Ruben's [4] result for quadratic forms in normal variables. He represented the distribution function of the quadratic form in normal variables as an infinite mixture of chi-square distribution functions. In the central case, we show that the distribution function of a quadratic form in t-variables can be represented as a mixture of beta distribution functions. In the noncentral case, the distribution function presented is an infinite series in beta distribution functions. An application to quadratic discrimination is given.

1. Introduction

Let $\boldsymbol{y}$ denote a k-dimensional vector which has a k-variate t-distribution with ν degrees of freedom, k-dimensional mean vector $\boldsymbol{\mu}$ and $(k\times k)$ matrix parameter V, i.e., the density of $\boldsymbol{y}$ is

$$(1.1)\qquad f(\boldsymbol{y}|\nu, \boldsymbol{\mu}, V)=\frac{\Gamma[(\nu+k)/2]|V^{-1}|^{1/2}}{\Gamma(\nu/2)(\nu\pi)^{k/2}}[1+\nu^{-1}(\boldsymbol{y}-\boldsymbol{\mu})^t V^{-1}(\boldsymbol{y}-\boldsymbol{\mu})]^{-(\nu+k)/2}\,.$$

We shall consider the distribution of the quadratic form $(\boldsymbol{y}-\boldsymbol{a})^tC(\boldsymbol{y}-\boldsymbol{a})$, where $\boldsymbol{a}$ is a k-dimensional vector and C is a $(k\times k)$ positive definite symmetric matrix. We first make the linear transformation $\boldsymbol{y}=LK\boldsymbol{x}$ and $\boldsymbol{\mu}-\boldsymbol{a}=LK\boldsymbol{b}$, where $\boldsymbol{x}$ and $\boldsymbol{b}$ are k-dimensional vectors, L is a $(k\times k)$ lower triangular matrix defined by $V=LL^t$, and K is the $(k\times k)$ orthogonal matrix of the eigenvectors of L^tCL. Denoting the vector of eigenvalues of L^tCL by $\boldsymbol{a}^t=(a_1, \cdots, a_k)$, the desired probability

$$(1.2)\qquad \Pr\{(\boldsymbol{y}-\boldsymbol{a})^tC(\boldsymbol{y}-\boldsymbol{a})\leqq t\}=\Pr\left\{\sum_{i=1}^{k} a_i(x_i+b_i)^2\leqq t\right\}\,,$$

where $\boldsymbol{x}^t=(x_1, \cdots, x_k)$ has a k-variate t-distribution with ν degrees of

Key words: Quadratic form, correlated t-variables, beta distribution, quadratic discrimination.

freedom, mean vector $\mathbf{0}$, and matrix parameter I. $Q_{\boldsymbol{a},\boldsymbol{b}}(\boldsymbol{x})=\sum_{i=1}^{k} a_i(x_i+b_i)^2$ is a noncentral quadratic form in uncorrelated, but dependent t-variables whose distribution function will be expressed, using a result due to Ruben [4], as an infinite series of beta distribution functions. When $\boldsymbol{\mu}=\boldsymbol{\alpha}$ in (1.1), $\boldsymbol{b}=\mathbf{0}$ and the distribution function of $Q_{\boldsymbol{a},\mathbf{0}}(\boldsymbol{x})$ reduces to a mixture of beta distributions. The noncentral case will be studied in Section 2, and the central case in Section 3. In Section 4, we apply the result for the central case to the derivation of predictive probabilities of misclassification in the case of Bayesian quadratic discrimination when the common means are known.

2. Distribution function for the noncentral case

Letting $R_1=\left\{\boldsymbol{x}\,\middle|\,Q_{\boldsymbol{a},\boldsymbol{b}}(\boldsymbol{x})=\sum_{i=1}^{k} a_i(x_i+b_i)^2\leqq t\right\}$, we need

$$\Pr\{Q_{\boldsymbol{a},\boldsymbol{b}}(\boldsymbol{x})\leqq t\}=\frac{\Gamma[(\nu+k)/2]}{\Gamma(\nu/2)(\nu\pi)^{k/2}}\int_{R_1}(1+\nu^{-1}\boldsymbol{x}^t\boldsymbol{x})^{-(\nu+k)/2}d\boldsymbol{x}\,. \tag{2.1}$$

Since $\Gamma[(\nu+k)/2](1+\nu^{-1}\boldsymbol{x}^t\boldsymbol{x})^{-(\nu+k)/2}=\int_0^\infty z^{(\nu+k)/2-1}\exp[-(1+\nu^{-1}\boldsymbol{x}^t\boldsymbol{x})z]dz$, (2.1) can be rewritten as

$$\begin{aligned}&\Pr\{Q_{\boldsymbol{a},\boldsymbol{b}}(\boldsymbol{x})\leqq t\}\\ &\quad=\Gamma^{-1}(\nu/2)(\nu\pi)^{-k/2}\int_{R_1}\int_0^\infty z^{(\nu+k)/2-1}\exp[-(1+\nu^{-1}\boldsymbol{x}^t\boldsymbol{x})z]dz\,d\boldsymbol{x}\,.\end{aligned} \tag{2.2}$$

Changing the order of integration, making the transformation $\boldsymbol{w}=(2z/\nu)^{1/2}\boldsymbol{x}$, and collecting terms, we get

$$\begin{aligned}&\Pr\{Q_{\boldsymbol{a},\boldsymbol{b}}(\boldsymbol{x})\leqq t\}\\ &\quad=\Gamma^{-1}(\nu/2)\int_0^\infty e^{-z}(z^{\nu/2-1})\left\{(2\pi)^{-k/2}\int_{R_2}\exp(-\boldsymbol{w}^t\boldsymbol{w}/2)d\boldsymbol{w}\right\}dz\,,\end{aligned} \tag{2.3}$$

where $R_2=\{\boldsymbol{w}\,|\,\sum a_i[w_i+(2z/\nu)^{1/2}b_i]^2\leqq(2z/\nu)t\}$. The expression in braces on the right-hand side of (2.3) describes the distribution function $\Pr\{\sum a_i[w_i+(2z/\nu)^{1/2}b_i]^2\leqq(2z/\nu)t\}$, where the w_i's are uncorrelated and have standard normal densities. This distribution function was expressed by Ruben [4] as a series of chi-square distribution functions,

$$\Pr\{\sum a_i[w_i+(2z/\nu)^{1/2}b_i]^2\leqq(2z/\nu)t\}=\sum_{j=0}^{\infty}c_j\int_0^{2zt/(\nu\beta)}\frac{v^{k/2+j-1}e^{-v/2}}{2^{k/2+j}\Gamma(k/2+j)}dv\,, \tag{2.4}$$

where

$$c_j=\prod_{i=1}^{k}(\beta/a_i)^{1/2}\exp[-(z/\nu)\sum b_i^2]\beta^j\sum_{r=0}^{j}(2z/\nu)^{j-r}\frac{\mathrm{E}\,(D^rG^{2(j-r)})}{(2j-2r)!\,2^r r!}\,,$$

$G=\sum_i (b_i/a_i^{1/2})x_i$, $D=\sum_i (\beta^{-1}-a_i^{-1})x_i^2$, $\beta<\min a_i$ is an arbitrary positive constant, and the expectation is taken over the independent standard normal variables $x_1, \cdots, x_k$. $\mathrm{E}\,(D^r G^{2(j-r)})$ can be obtained from the cumulants κ_{rs} of D and G^2 which are given in Box and Tiao ([1], p. 516),

$$\kappa_{rs}=2^{r+s-1}(r+s-1)!\sum_{i=1}^{k}\left(\frac{a_i-\beta}{a_i-\beta+b_i^2\beta}\right)^r\left(\frac{b_i^2\beta}{a_i-\beta+b_i^2\beta}\right)^s \qquad (r+s\geqq 1)\,.$$

The relationship between bivariate cumulants and bivariate moments is given, e.g., by Kendall and Stuart ([3], § 3.28).

Using (2.4) in (2.3), first making the transformation $u=v\nu/(2z)$, and then writing v for u, and rearranging terms, we get

$$\begin{aligned}(2.5)\quad \Pr\{Q_{a,b}(x)\leqq t\} &=\Gamma^{-1}(\nu/2)\left[\prod_{i=1}^{k}(\beta/a_i)^{1/2}\right]\sum_{j=0}^{\infty}(\beta/2)^j\Gamma^{-1}(k/2+j)\\ &\quad\times\sum_{r=0}^{j}\frac{2^{2(j-r)}\,\mathrm{E}\,(D^rG^{2(j-r)})}{(2j-2r)!\,r!\,\nu^{\nu/2+2j-r}}\int_0^{t/\beta}v^{k/2+j-1}\\ &\quad\times\left\{\int_0^{\infty}z^{(k+\nu)/2+2j-r-1}\exp\left[-z(\nu+\sum b_i^2+v)/\nu\right]dz\right\}dv\,.\end{aligned}$$

The expression in braces on the right-hand side of (2.5) equals $\Gamma[(k+\nu)/2+2j-r][(\nu+\sum b_i^2+v)/\nu]^{-(k+\nu)/2+2j-r}$. Using this, making the transformation $u=v/[v+(\nu+\sum b_i^2)]$ so that $u\in(0,1)$, and then writing v for u, the right-hand side of (2.5) equals

$$\begin{aligned}(2.6)\quad &\sum_{j=0}^{\infty}\sum_{r=0}^{j}\Gamma^{-1}(\nu/2)\left[\prod_{i=1}^{k}(\beta/a_i)^{1/2}\right](1+\sum b_i^2/\nu)^{-\nu/2}\beta^j\\ &\times\frac{\mathrm{E}\,(D^rG^{2(j-r)})\Gamma(k/2+j-r)}{(2j-2r)!\,r!\,2^r[(\nu+\sum b_i^2)/2]^{j-r}}I(d\,|\,k/2+j,\,\nu/2+j-r)\,,\end{aligned}$$

where $d=t/[t+\beta(\nu+\sum b_i^2)]$ and $I(d\,|\,\alpha,\gamma)$ is the beta distribution function with α and γ degrees of freedom,

$$(2.7)\qquad I(d\,|\,\alpha,\gamma)=B^{-1}(\alpha,\gamma)\int_0^d u^{\alpha-1}(1-u)^{\gamma-1}du\,.$$

Now using the fact that

$$(2.8)\qquad I(d\,|\,\alpha,\gamma+j-r)=\sum_{i=0}^{j-r}(-1)^iC_i^{j-r}\frac{B(\alpha+i,\gamma)}{B(\alpha,\gamma+j-r)}I(d\,|\,\alpha+i,\gamma)\,,$$

we can rewrite (2.6) and get

$$(2.9)\qquad \Pr\{Q_{a,b}(x)\leqq t\}=\sum_{j=0}^{\infty}\left[\sum_{i=0}^{\langle j/2\rangle}\sum_{r=0}^{j-2i}c_{j-i,r,i}\right]I(d\,|\,k/2+j,\,\nu/2)\,,$$

where

$$c_{j-i,r,i}=\left[\prod_{i=1}^{k}(\beta/a_i)^{1/2}\right](1+\sum b_i^2/\nu)^{-\nu/2}\beta^{j-i}(-1)^i\frac{\mathrm{E}\,(D^rG^{2(j-i-r)})}{(2j-2i-2r)!\,2^r r!}$$
$$\times C_i^{j-i-r}\frac{\Gamma(k/2+j)}{\Gamma(k/2+j-i)}\,\frac{\Gamma[(k+\nu)/2+2j-2i-r]}{[(\nu+\sum b_i^2)/2]^{j-i-r}\Gamma[(k+\nu)/2+j]}\,,$$

and $\langle\alpha\rangle$ is the integer part of α.

Thus the desired distribution function for a noncentral quadratic form in t-variables is an infinite series in beta distribution functions.

3. Distribution function for the central case

In the central case $\boldsymbol{a}=\boldsymbol{\mu}$ and thus $\boldsymbol{b}=\mathbf{0}$ in (2.1). This leads to a considerable simplification of (2.9). When $\boldsymbol{b}=\mathbf{0}$, R_2 in (2.3) reduces to $R_3=\{\boldsymbol{w}\,|\sum a_i w_i^2\leqq(2z/\nu)t\}$. The same simplification is obtained in (2.4), in particular, the c_j's are not functions of z. Ruben [4] derived convenient recursion formulae for the c_j's.

$$(3.1)\qquad\begin{aligned}&c_0=\prod_{i=1}^{k}(\beta/a_i)^{1/2}\,,\\&c_j=(2j)^{-1}\sum_{i=0}^{j-1}g_{j-i}c_i\,,\qquad(j\geqq1)\\&g_r=\sum_{i=1}^{k}(1-\beta/a_i)^r\,,\qquad\text{and}\\&\beta<\min a_i\,.\end{aligned}$$

Using (3.1) in (2.4), making the transformation $u=v\nu/(2z)$, but writing v for u, (2.3) reduces to

$$(3.2)\qquad \Pr\{Q_{\boldsymbol{a},0}(\boldsymbol{x})\leqq t\}=\Gamma^{-1}(\nu/2)\sum_{j=0}^{\infty}c_j\int_0^{t/\beta}\frac{v^{k/2+j-1}}{\nu^{k/2+j}\Gamma(k/2+j)}\times\left\{\int_0^{\infty}e^{-z(\nu+v)/\nu}z^{(\nu+k)/2+j-1}dz\right\}dv\,.$$

Now the expression in braces on the right-hand side of (3.2) equals $\Gamma[(\nu+k)/2+j][(\nu+v)/\nu]^{-(\nu+k)/2+j}$. Using this and making the transformation $u=v/(\nu+v)$ so that $u\in(0,1)$, we get

$$(3.3)\qquad \Pr\{Q_{\boldsymbol{a},0}(\boldsymbol{x})\leqq t\}=\sum_{j=0}^{\infty}c_jI\left(\frac{t}{t+\beta\nu}\,\middle|\,k/2+j,\,\nu/2\right),$$

where $I(d|\alpha,\gamma)$ is given in (2.7). Thus the desired distribution function for a central quadratic form can be expressed as a mixture of beta distributions.

If the series in (3.3) is truncated after n terms, the accuracy of this approximation can be assessed by noting that $\sum_{j=n+1}^{\infty}c_jI(t(t+\beta\nu)^{-1}|k/2$

$+j, \nu/2) \leqq \left(1-\sum_{j=0}^{n} c_j\right) I(t(t+\beta\nu)^{-1}|k/2+n, \nu/2)$, since $I(d|\alpha, \gamma) \geqq I(d|\alpha+1, \gamma)$ and $\sum_{j=0}^{\infty} c_j = 1$.

4. An application

Desu and Geisser [2] examine a discrimination problem with two populations where the units in each population come from k-variate normal distributions with known common mean vector, taken to be $\mathbf{0}$, and unknown different covariance matrices. They give a Bayesian treatment of the problem which involves predictive distributions as a basis for discrimination. In one case examined, the two predictive distributions are k-variate t-distributions, denoted by $f(\boldsymbol{y}|\nu, \mathbf{0}, V_r)$, $r=1, 2$, with $f(\boldsymbol{y}|\cdot, \cdot, \cdot)$ as in (1.1). The discrimination rule is then to assign a unit $\boldsymbol{y}$ of unknown origin to population 1 if

$$D(\boldsymbol{y}) = \boldsymbol{y}^t C \boldsymbol{y} > t\,, \tag{4.1}$$

where $C = V_2^{-1} - wV_1^{-1}$, and w and t are constants depending on the costs of misclassification, prior probabilities for the two populations, $|V_1|$, $|V_2|$, ν and k. Predictive probabilities of misclassification can now be found. For example, the predictive probability of misclassifying an observation from population 1 into population 2 is $\Pr\{D(\boldsymbol{y})<t\}$, where $\boldsymbol{y}$ has density $f(\boldsymbol{y}|\nu, \mathbf{0}, V_1)$. Note that C in (4.1) is not necessarily positive definite (or negative definite in which case it can be made positive definite by interchanging the labels of the two populations), but in many cases it will be so. If C is positive definite, $\Pr\{D(\boldsymbol{y})<t\}$ is just the probability in (1.2) with $\boldsymbol{\mu}=\boldsymbol{\alpha}=\mathbf{0}$, and $V=V_1$, i.e., a probability for a central quadratic form. The other probability of misclassification can be found in a similar way.

In the more usual discrimination problem when the means of the two populations are unknown and not equal, predictive probabilities of misclassification involve noncentral quadratic forms in t variables.

5. Remarks

The distribution function for the central case can be evaluated conveniently using a modification of the algorithm described by Sheil and O'Muircheartaigh [5]. Some numerical work carried out suggests that if $\beta \in [(.90) \min a_i, (.96) \min a_i]$ then the number of terms needed in (3.3) to achieve a required level of accuracy will be smaller than for other values of β.

The series representation presented for the noncentral case, however, requires a heavier computational effort.

UNIVERSITY OF TORONTO

REFERENCES

[1] Box, G. E. P. and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*, Addison-Wesley Publishing Co., Reading.
[2] Desu, M. M. and Geisser, S. (1973). Methods and applications of equal-mean discrimination, *Discriminant Analysis and Applications* (ed. T. Cacoullos), Academic Press, Inc., New York, 139-159.
[3] Kendall, M. G. and Stuart, A. (1969). *The Advanced Theory of Statistics I*, 3rd edition, Hafner Publishing Co., New York.
[4] Ruben, H. (1962). Probability content of regions under spherical normal distribution IV, *Ann. Math. Statist.*, **33**, 542-570.
[5] Sheil, J. and O'Muircheartaigh, I. (1977). The distribution of non-negative quadratic forms in normal variables, *Appl. Statist.*, **26**, 92-98.