APPROXIMATIONS TO THE DISTRIBUTION OF QUADRATIC FORMS

By M. M. Siddiqui

Colorado State University and Boulder Laboratories, National Bureau of Standards

1. Introduction and summary. Positive definite quadratic forms in normal variates, which do not necessarily reduce to a multiple of a χ^2 variate, arise quite naturally in estimation and hypothesis testing problems related to normal distributions and processes. A classical example is the problem of testing the difference between two sample means, $\bar{x} - \bar{y}$, where the x observations have variance other than of y observations (Welch [9]). More recent examples include: analysis of variance when errors are assumed to have unequal variance or are correlated (Box [1] [2]); regression analysis with stationary errors (Siddiqui [8]); and estimation of spectral density functions of stationary processes (Freiberger and Grenander [4]).

Let $Q = \frac{1}{2}Y'MY$, where $Y = [Y_1, \cdots, Y_n]$ is a $N(0; V)$ distributed column vector, Y' its transposed row vector, 0 zero vector, V a positive definite covariance matrix, and M a real symmetric matrix of rank $m \leqq n$. Let $a_1, \cdots, a_m$ be the non-zero characteristic roots of $A = MV$. It is well known (see, for example, Ruben [8]) that there exists a non-singular transformation from Y to X such that $X_1, \cdots, X_n$ are independent $N(0, 1)$ variates and $Q = \frac{1}{2}\sum_1^m a_j X_j^2$. Without loss of generality we therefore assume that Q has this canonical form.

Many papers have been written on the distribution of Q, especially when a_j are positive, and a more or less comprehensive list of these is included in the references of the two papers by Ruben [8] [9]. We shall therefore refer to only those which have direct bearing with the present paper.

In this paper, we will be mainly concerned with distribution of Q when m is an even number, say $2k$, and a_j positive. When m is an odd number a slight modification is necessary and this is mentioned in Remark (2) of Section 3. We will choose our subscripts so that $0 < a_1 \leqq a_2 \cdots \leqq a_{2k}$. After some notation and preliminaries in Section 2, a well known result will be stated as Theorem 1 under which $F(x) = \Pr(Q > x)$ can be evaluated as a finite linear combination of gamma df's. In other situations we require some methods of approximating to $F(x)$. In Sections 3 and 4 a simple approximation to $F(x)$ will be presented which reduces to the exact distribution when the condition of Theorem 1 is satisfied. The method is based on bounding Q by Q_1 and Q_2, where Q_1 and Q_2 are quadratic forms satisfying the condition of Theorem 1. The approximation is then obtained by minimizing $d(F, \hat{F})$ where $\hat{F}(x)$ is a linear combination of $F_i(x) = \Pr(Q_i > x)$, $i = 1, 2$, and $d(\cdot, \cdot)$ is the distance function of the metric space $L^2(0, \infty)$. In Section 5 a few numerical examples will be worked out for purposes of illustration.

Received 16 October 1963; revised 20 October 1964.

2. Notation and preliminaries. Throughout the paper, the letters, with or without subscripts, a, b, c, and p will denote positive numbers; k, m, and n positive integers; X, a $N(0, 1)$ variate and X_i and X_j, independent if $i \neq j$; and $\chi^2(p)$, a χ^2 variate with p degrees of freedom. If the Laplace transform of a function $h(x)$ exists, it will be denoted by $h^*(s)$.

Let $0 < a_1 \leqq a_2 \leqq \cdots \leqq a_{2k}$ and $Q = \frac{1}{2}\sum_{j=1}^{2k} a_j X_j^2$. Let $f(x)$ denote the pdf of Q and $F(x) = \Pr(Q > x) = \int_x^\infty f(y)\,dy$. We wish to evaluate $F(x)$ exactly if possible, approximately otherwise. Let $g_p(x; c) = c^{-p}[\Gamma(p)]^{-1}e^{-x/c}x^{p-1}$, if $x > 0$; 0, if $x \leqq 0$, and $G_p(x; c) = \int_x^\infty g_p(y; c)\,dy$. Note that $G_p(x; c) = G_p(x/c; 1)$. Also $G_m(x; c) = \Pr(\chi^2(2m) > 2x/c)$.

When $\operatorname{Re} s > 0$ all the following relations hold

$$(2.1)\quad g_p^*(s; c) = (1 + cs)^{-p}, \qquad G_p^*(s; c) = s^{-1}[1 - g_p^*(s; c)],$$
$$f^*(s) = \prod_1^{2k} (1 + a_j s)^{-\frac{1}{2}}, \qquad F^*(s) = s^{-1}[1 - f^*(s)].$$

The relations for g^* and f^* even hold for $\operatorname{Re} s > -c^{-1}$ and $\operatorname{Re} s > -a_{2k}^{-1}$, respectively.

When a_j's are equal within groups of even size $f(x)$ and $F(x)$ can be evaluated as a finite sum of g and G functions respectively. This result is well known (see, for example, Box [1] Theorem 2.4) and will be needed later. We state a slightly more general result as a theorem.

THEOREM 1. *If $f^*(s)$ can be developed as*

$$f^*(s) = \sum_{j=1}^r A_j(1 + c_j s)^{-p_j},$$

then

$$F(x) = \sum_{j=1}^r A_j G_{p_j}(x; c_j).$$

3. An approximation to $F(x)$. If Q is not of the type of Theorem 1, we obtain an approximation to $F(x)$. Let

$$2Q_1 = a_1(X_1^2 + X_2^2) + a_3(X_3^2 + X_4^2) + \cdots + a_{2k-1}(X_{2k-1}^2 + X_{2k}^2),$$
$$2Q_2 = a_2(X_1^2 + X_2^2) + a_4(X_3^2 + X_4^2) + \cdots + a_{2k}(X_{2k-1}^2 + X_{2k}^2).$$

Let $F_i(x) = \Pr(Q_i > x)$, $i = 1, 2$. Since Q_i, $i = 1, 2$, is of the form of Theorem 1, $F_i(x)$ can be evaluated exactly. If $a_{2j-1} = a_{2j}$, $j = 1, 2, \cdots, k$, then $Q = Q_1 = Q_2$, hence we assume that $a_{2j-1} \neq a_{2j}$, at least for one j. Then, almost surely, $Q_1 < Q < Q_2$, which implies, for all $x > 0$,

$$(3.1)\qquad F_1(x) < F(x) < F_2(x).$$

These inequalities motivate us to consider an approximation to $F(x)$ of the form

$$(3.2)\qquad \hat{F}(x) = F_1(x) + \theta[F_2(x) - F_1(x)],$$

where θ, $0 < \theta < 1$, is a constant.

To determine "optimum" θ many a criterion can be employed. Here, we employ the minimum "distance" method where the "distance" chosen is that of the (real Hilbert) space, $L^2(0, \infty)$, of functions which are square integrable over $(0, \infty)$. If p and q are in $L^2(0, \infty)$ we have the inner product (p, q), the norm $\|p\|$ and the distance $d(p, q)$ given by

$$(p, q) = \int_0^\infty p(x)q(x)\,dx, \qquad \|p\| = (p, p)^{\frac{1}{2}}, \qquad d(p, q) = \|p - q\|.$$

We determine θ in (3.2) by minimizing $d^2(F, \hat{F})$. We must first show, however, that F and $\hat{F}$ belong to $L^2(0, \infty)$. Now the function of x, $e^{-cx}x^{p-1}$, $c > 0$, $p \geqq 1$, belongs to $L^2(0, \infty)$. For $x > 0$, $G_m(x; c)$ is a linear combination of such functions; F_i, $i = 1, 2$, is a linear combination of such G's; finally, from (3.1), F is bounded by F_2. Hence, $d^2(F, \hat{F})$ is well defined, and minimizing it with respect to θ, we obtain

$$(3.3) \qquad \theta = (F - F_1, F_2 - F_1)\|F_2 - F_1\|^{-2}.$$

Remarks.

(1) If $x > 0$, $-\theta[F_2(x) - F_1(x)] < F(x) - \hat{F}(x) < (1 - \theta)[F_2(x) - F_1(x)]$, $\max|F(x) - \hat{F}(x)| \leqq \max(\theta, 1 - \theta)\max[F_2(x) - F_1(x)]$.

(2) Let $Q = \frac{1}{2}\sum_{j=1}^{n} a_j X_j^2$ where n is odd, say $n = 2k - 1$. In this case we construct $2Q_1 = a_1(X_1^2 + X_2^2) + a_3(X_3^2 + X_4^2) + \cdots + a_{2k-3}(X_{2k-3}^2 + X_{2k-2}^2)$, $2Q_2 = a_1(X_1^2 + X_{2k}^2) + a_3(X_2^2 + X_3^2) + \cdots + a_{2k-1}(X_{2k-2}^2 + X_{2k-1}^2)$, where X_{2k} is an additional independent $N(0, 1)$ variate. The approximation $\hat{F}(x)$ is then obtained in exactly the same way.

(3) Let Y_i, $i = 1, 2$, be independent variates with pdf's $g_{p_i}(x; c_i)$. The ratio $Z = Y_1/Y_2$ then has the beta pdf ([3], pp. 241–242).

$$(3.4) \quad h(x; p_1, p_2, c_1, c_2) = (c_2/c_1)^{p_1}[B(p_1, p_2)]^{-1}(1 + c_2x/c_1)^{-p_1-p_2}x^{p_1-1}, \quad x \geqq 0.$$

From this we conclude that if Y_i, $i = 1, 2$, are independent variates such that their pdf's can be represented each as a finite mixture of gamma pdf's then $Z = Y_1/Y_2$ can be represented as a finite mixture of the beta pdf's of type (3.4). Finally, let Y_i, $i = 1, 2$, be independent positive definite quadratic forms of type Q. We can approximate their pdf's each by a finite mixture of gamma pdf's as described in the earlier part of this Section and hence can obtain an approximation to pdf of the ratio $Z = Y_1/Y_2$ as a finite mixture of the beta pdf's of type (3.4).

4. Evaluation of θ. We have

$$\theta = [(F, F_2) - (F, F_1) - (F_1, F_2) + \|F_1\|^2]\|F_2 - F_1\|^{-2}.$$

Since F_1 and F_2 are linear combinations of functions of type $G_m(\cdot\,; c)$ we only require the evaluation of integrals of type $\int_0^\infty G_m(x; c)h(x)\,dx$, where h may be F, F_1 or F_2. For this purpose we note that

$$(4.1) \qquad G_m(x; c) = e^{-x/c}\sum_{j=0}^{m-1}(x/c)^j/j!.$$

If $h^*(s)$ exists and is differentiable j times,

$$(4.2) \qquad \int_0^\infty e^{-sx}x^j h(x)\,dx = (-1)^j h^{*(j)}(s),$$

where the superscript (j) denotes the jth differential coefficient. Recall from

(2.1) that $F^*(s) = s^{-1}[1 - \prod_1^{2k} (1 + a_j s)^{-\frac{1}{2}}]$; $F_1^*(s)$ is obtained from it by replacing a_{2j} by a_{2j-1}, $j = 1, \cdots, k$; and $F_2^*(s)$ by replacing a_{2j-1} by a_{2j}. Thus (4.2) is sufficient for the evaluation of θ.

To evaluate the differential coefficients in (4.2) a systematic procedure may be recommended. Let $p(x)$ be a function which is a product of several functions each differentiable n times at the point x. Let $u(x) = \log p(x)$. We then have

$$p^{(1)}(x) = p(x)u^{(1)}(x),$$
$$p^{(r+1)}(x) = \sum_{j=0}^{r} \binom{r}{j} p^{(r-j)}(x)u^{(j+1)}(x), \qquad r = 1, 2, \cdots, n-1.$$

We first calculate $p(x)$ and $u^{(r)}(x)$, $r = 1, 2, \cdots n$ at the given numerical value x and then $p^{(1)}(x), \cdots, p^{(n)}(x)$ recursively.

Illustration. Let $2Q = 3\sum_1^3 x_i^2 + 5\sum_4^6 x_i^2$. We have

$$2Q_1 = 3\sum_1^4 x_i^2 + 5\sum_5^6 x_i^2, \qquad 2Q_2 = 3\sum_1^2 x_i^2 + 5\sum_3^6 x_i^2,$$
$$F^*(s) = s^{-1} - s^{-1}(1 + 3s)^{-\frac{3}{2}}(1 + 5s)^{-\frac{3}{2}},$$
$$F_1^*(s) = s^{-1} - s^{-1}(1 + 3s)^{-2}(1 + 5s)^{-1},$$
$$F_2^*(s) = s^{-1} - s^{-1}(1 + 3s)^{-1}(1 + 5s)^{-2}.$$

By partial fraction expansion of $F_i^*(s)$ and applying Theorem 1, we obtain

$$F_1(x) = 6.25e^{-x/5} - 5.25e^{-x/3} - 0.5xe^{-x/3},$$
$$F_2(x) = 2.25e^{-x/3} - 1.25e^{-x/5} + 0.5xe^{-x/5}.$$

To evaluate (F, F_1), (F_1, F_2), etc. we make repeated use of (4.2). Thus, for example,

$$(F, F_1) = 6.25F^*(\tfrac{1}{5}) - 5.25F^*(\tfrac{1}{3}) + 0.5F^{*(1)}(\tfrac{1}{3}),$$
$$(F_2, F_1) = 6.25F_2^*(\tfrac{1}{5}) - 5.25F_2^*(\tfrac{1}{3}) + 0.5F_2^{*(1)}(\tfrac{1}{3}).$$

The terms $F^*(s)$, $F_2^*(s)$, etc. for $s = \frac{1}{5}, \frac{1}{3}$ can be directly evaluated. To evaluate the differential coefficients, say, $F^{*(1)}(\frac{1}{3})$, let $F^*(s) = s^{-1} - p(s)$, $u(s) = \log p(s)$, i.e., $u(s) = -\log s - \frac{3}{2}\log(1 + 3s) - \frac{3}{2}\log(1 + 5s)$. We then have

$$u^{(1)}(s) = p^{(1)}(s)/p(s) = -s^{-1} - \tfrac{9}{2}(1 + 3s)^{-1} - \tfrac{15}{2}(1 + 5s)^{-1},$$
$$F^{*(1)}(s) = -s^{-2} - p^{(1)}(s) = -s^{-2} - u^{(1)}(s)p(s).$$

We thus evaluate: $(F, F_1) = 7.8016$, $(F, F_2) = 8.4941$, $(F_1, F_1) = 7.4805$, $(F_1, F_2) = 8.0947$, and $(F_2, F_2) = 8.8635$. Finally, $\theta = 0.506$, and $\hat{F}(x) = 0.494F_1(x) + 0.506F_2(x)$.

Thus, the procedure for determining $\hat{F}(x)$ can be carried out in the following sequence.

(1) Write down $F^*(s)$ and $F_i^*(s)$, $i = 1, 2$.

(2) Use Box's [1] Theorem 2.4 to expand each $F_i^*(s)$ in partial fractions. Then, from Theorem 1, express each $F_i(x)$ as a finite linear combination of functions G.

(3) Use (4.1) to express each G as a finite linear combination of functions $x^r e^{-sx}$, $s > 0$, $r = 0, 1, \cdots$, thus representing each $F_i(x)$ as a finite linear combination of functions $x^r e^{-sx}$.

(4) Apply (4.2) to evaluate inner products (F, F_i) and (F_i, F_j), $i, j = 1, 2$. If differentiation of $F^*(s)$, $F_i^*(s)$ is needed let $F^*(s)$ or $F_i^*(s) = s^{-1} - p(s)$, $u(s) = \log p(s)$ and use the systematic procedure recommended in this Section before the illustration.

(5) Evaluate θ and obtain $\hat{F}(x)$.

A much simpler but rougher approximation to $F(x)$ can be easily obtained by bounding Q by $Q_1 = \frac{1}{2}a_1 \sum_{j=1}^{2k} X_j^2$ and $Q_2 = \frac{1}{2}a_{2k} \sum_{j=1}^{2k} X_j^2$. The procedure of Section 3 can be applied with $F_1(x) = G_k(x; a_1)$ and $F_2(x) = G_k(x; a_{2k})$. This approximation may suffice if a_{2k}/a_1 is near unity. If a_{2k}/a_1 is not near unity, consider

$$F_0(x) = \theta_1 G_k(x; a_1) + \theta_2 G_k(x; a_{2k}) + (1 - \theta_1 - \theta_2) G_k(x; a), \tag{4.3}$$

where θ_1 and θ_2 are constants to be determined by minimizing $d^2(F, F_0)$, and a is some average of $a_1, \cdots, a_{2k}$. For simplicity we will take a to be the arithmetic mean. Here there is no guarantee that θ_1 and θ_2 will be in the interval (0, 1). Let, for $i, j = 1, 2$,

$$A_{ij} = \int_0^\infty [G_k(x; a_i) - G_k(x; a)][G_k(x; a_j) - G_k(x; a)]\, dx,$$
$$\lambda_i = \int_0^\infty [F(x) - G_k(x; a)][G_k(x; a_i) - G_k(x; a)]\, dx.$$

These quantities can be evaluated by the use of relations (4.2). If $A = A_{11}A_{22} - A_{12}^2 \neq 0$, we have the solutions

$$\theta_1 = A^{-1}(A_{22}\lambda_1 - A_{12}\lambda_2), \tag{4.4}$$
$$\theta_2 = A^{-1}(A_{11}\lambda_2 - A_{12}\lambda_1).$$

5. Comparison with the exact distribution. The exact distribution of $Q = \frac{1}{2}\sum_{j=1}^{2k} a_j X_j^2$ is not available except when $k = 1$ (Grad and Solomon [5]) and when Theorem 1 is applicable. Since the approximation, $\hat{F}(x)$, proposed in Section 3 becomes exact when Theorem 1 applies it is not possible to produce examples with $k \geqq 2$ to judge the "goodness of fit" of $\hat{F}(x)$ to $F(x)$. We therefore confine ourselves to a few calculations with $k = 1$. To make our calculations comparable to those of Grad and Solomon [5, Table 1] we take $a_1 + a_2 = 2$.

TABLE 1

Pr $(Q \leqq x)$

a_1, a_2	$x =$ .1	.4	.7	1.0	2.0	3.0	4.0	5.0
.8, 1.2	09693	3340	5080	6358	8646	9487	9802	9923
	09825	3371	5108	6373	8632	9465	9784	9911
.6, 1.4	1029	3482	5221	6466	8638	9441	9761	9895
	1091	3614	5334	6530	8571	9351	9692	9851
.4, 1.6	1158	3755	5464	6630	8604	9365	9698	9853
	1344	4100	5712	6730	8421	9169	9556	9763

TABLE 2

Pr $(Q \leqq x)$

a_1, a_2	$x =$ .1	.4	.7	1.0	2.0	3.0	4.0	5.0
.4, 1.6	1158	3755	5464	6630	8604	9365	9698	9853
	1177	3776	5465	6618	8597	9373	9708	9859
.2, 1.8	1461	4226	5780	6785	8527	9269	9624	9803
	1570	4277	5734	6713	8526	9303	9654	9822
.1, 1.9	1813	4521	5904	6819	8478	9219	9585	9775
	2114	4520	5768	6694	8568	9275	9632	9805

There is no loss of generality in this constraint. In considering these examples as "for" or "against" $\hat{F}(x)$ one should bear in mind that $k = 1$ is an extreme situation and most unfavorable to the proposed method (the "degrees of freedom" in varying the a's is only 1).

Let $Q = \frac{1}{2}(a_1X_1^2 + a_2X_2^2)$, $0 < a_1 \leqq a_2$, $a_1 + a_2 = 2$, and $r = a_2/a_1$. We have

$$\hat{F}(x) = e^{-x/a_1} + \theta(e^{-x/a_2} - e^{-x/a_1}),$$

$$\theta = (r-1)^{-2}[(r-1)(2r+1) - (2+2r)^{\frac{1}{2}}(r^{\frac{3}{2}} - 1)].$$

Since $1 \leqq r < \infty$, $0.5 \leqq \theta < 2 - 2^{\frac{1}{2}} = 0.5858$. In Table 1 the first entry is $1 - F(x)$ and the second $1 - \hat{F}(x)$. It may be noted that the approximation becomes poorer as r increases.

When r is large we compute $F_0(x) = \theta_1 e^{-x/a_1} + \theta_2 e^{-x/a_2} + (1 - \theta_1 - \theta_2)e^{-x}$, where θ_1 and θ_2 are calculated from (4.4). In Table 2 the first entry is $1 - F(x)$ and the second $1 - F_0(x)$.

REFERENCES

[1] Box, G. E. P. (1954). Some theorems on quadratic forms applied in 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] Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems, II. Effects of inequality of variance and of correlation between errors in the two-way classification. *Ann. Math. Statist.* **25** 484–498.

[3] Cramér, Harald (1951). *Mathematical Methods of Statistics*. Princeton Univ. Press.

[4] Freiberger, Walter and Grenander, Ulf (1959). Approximate distributions of noise power measurement. *Quart. Appl. Math.* **17** 271–283.

[5] Grad, Arthur and Solomon, Herbert (1955). Distribution of quadratic forms and some applications. *Ann. Math Statist.* **26** 464–477.

[6] Ruben, Harold (1960). Probability content of regions under spherical normal distributions, I. *Ann. Math. Statist.* **31** 598–618.

[7] Ruben, Harold (1962). Probability content of regions under spherical normal distributions, IV: The distribution of homogeneous and non-homogeneous quadratic functions of normal variables. *Ann. Math. Statist.* **33** 542–570.

[8] Siddiqui, M. M. (1960). Tests for regression coefficients when errors are correlated. *Ann. Math. Statist.* **31** 929–938.

[9] Welch, B. L. (1937). The significance of the difference between two means when the population variances are unequal. *Biometrika* **29** 350–362.