

Approximations to the distributions of quadratic forms in normal variables

Sha Yu

January 10, 2016

1 Introduction

Many test statistics are asymptotically equivalent to quadratic forms of normal variables, which are further equivalent to the weighted sum of independent chi-square variables.

Consider the quadratic form $Q(X) = X'AX$, where X follows an n -dimensional multivariate normal distribution with mean vector μ_x and non-singular variance matrix Σ , and A represents a $p \times p$ symmetric and nonnegative definite matrix. A problem of interest is to evaluate the probability

$$Pr(Q(X) > q) \quad (1.1)$$

By spectral decomposition, a quadratic form can be converted to the weighted sum of independent chi-square variables. Let P be a $n \times n$ orthonormal matrix which converts $B = \Sigma^{1/2}A\Sigma^{1/2}$ to the diagonal form $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n) = PBP'$. Let $m = \text{rank}(A)$. We assume that $\lambda_1 \geq \dots \geq \lambda_n \geq 0$ and $\lambda_{m+1} = \dots = \lambda_n = 0$ if $m < n$. Since $Y = P\Sigma^{-1/2}X$ is normally distributed with mean $P\Sigma^{-1/2}\mu_x$ and variance I_n , thus

$$Q = X'AX = Y'\Lambda Y = \sum_{r=1}^m \lambda_r \chi_{hr}^2(\delta_r) \quad (1.2)$$

where $h_r = 1$, $\delta_r = \mu_{yr}^2$, and μ_{yr} is the i th component of μ_y for $r = 1, \dots, m$.

Estimating the tail probability of $Q(x)$ arises in many statistical applications. The power analysis of a test procedure if the (asymptotic) distribution test statistic (e.g., Pearson's chi-square statistic) takes the form $Q(X)$. A famous example is the Chernoff-Lehmann test statistic for goodness-of-fit to a fixed distribution, which converges to quadratic form under the null (alternative) hypothesis (see, e.g., Moore and Spruill (1975) and Spruill (1976)). In time series analysis, Box-Pierce-Ljung portmanteau test statistic for lack of t in ARMA models converges in distribution to a weighted sum of non-central chi-square random variables under local alternatives. The quadratic forms are also the building blocks for the commonly used F-statistics in ANOVA and regression.

It suggests the importance of determining (1.1) for quadratic forms in central and noncentral normal variables in level and power studies. Thus (1.1) should be computed accurately for quadratic forms in normal variables. Unfortunately, for the probability density function (pdf) of weighted sum of independent chi-square variables, there is no known closed analytic expression yet except in some special cases. For instance, if $A\Sigma$ is idempotent matrix of rank p , then $Q(X) \sim \chi_{p,\lambda}^2$, where $\lambda = \mu_x' A \mu_x$.

Because the exact distribution of weighted sum of independent chi-square variables is difficult to obtain in general, various approximations to its distribution have been proposed. In section 2, we will illustrate five most popular approximation approaches and their algorithms.

2 Five approximation approaches

As mentioned, many methods have been proposed for the problem of (1.1). In this section, five most influential approximation approaches will be illustrated.

2.1 Imhof's Approximation

Imhof derived this approximation using numerical integration to invert the characteristic function of $Q(X)$ in 1961. The characteristic function of (1.1) is

$$\phi(t) = \sum_{r=1}^m (1 - 2i\lambda_r t)^{-\frac{h_r}{2}} \exp(i \sum_{r=1}^m \frac{\delta_r^2 \lambda_r t}{1 - 2i\lambda_r t}) \quad (2.1)$$

The cumulative distribution function $F(x)$ of the variable Q defined by (1.1) can be obtained quite easily by straightforward numerical integration of an inversion formula. The standard inversion formula. Such a formula is implicit in work of Gurland (1948) and has been later derived explicitly by Gil-Pelaez (1951), namely

$$F(x) = \frac{1}{2} - \frac{1}{\pi} \int_0^\infty t^{-1} g(e^{-itx} \phi(t)) dt \quad (2.2)$$

where $g(z)$ denotes the imaginary part of z and where the characteristic function $\phi(t)$ of $Q(X)$ is given by (2.1). Using the relations

$$\begin{aligned} \arg[(1 - ibt^{-y})] &= g \tan^{-1}(bt), \quad |(1 - ibt)^{-g}| = (1 + b^2 t^2)^{-\frac{1}{2}g}, \\ \arg[\exp(iat/(1 - ibt))] &= at/(1 + b^2 t^2), \quad |\exp[iat/(1 - ibt)]| = \exp[-abt^2/(1 + b^2 t^2)], \end{aligned}$$

one finds that for quadratic form Q of (1.1), equation (2.2) can be written after the substitution $2t = u$ is made,

$$Pr(Q(X) > q) \approx \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \frac{\sin \theta(u)}{u \rho(u)} du \quad (2.3)$$

where

$$\theta(u) = \frac{1}{2} \sum_{r=1}^m [h_r \arctan(\lambda_r u) + \delta_r^2 \lambda_r u (1 + \lambda_r^2 u^2)^{-1}] - \frac{1}{2} q u,$$

$$\rho(u) = \prod_{r=1}^m (1 + \lambda_r^2 u^2)^{\frac{h_r}{4}} \exp\left(\frac{\sum_{r=1}^m (\delta_r \lambda_r u)^2}{2(1 + \lambda_r^2 u^2)}\right)$$

This approximation is not easy to calculate but it can be very accurate and bounds on accuracy can be found. Also, the eigenvalues need be specified explicitly in order to apply this method, which is very time-consuming in the case that A is complex and high deminsional.

2.2 Farebrother's Approximation

Farebrother (1984) and Sheil and O'Muircheartaigh (1977), based on the results of Ruben (1962), exploit the fact that (1.1) can be written as an infinite series of an infinite series of centra chi-square distriubuions.

$$Pr(Q(X) > q) = \sum_{k=0}^{\infty} c_k Pr(\chi_{2k+\tilde{h}}^2 > \frac{t}{\beta}) \quad (2.4)$$

For any $0 < \beta < \min(\lambda_1, \dots, \lambda_m)$, where $\tilde{h} = \sum_{r=1}^m h_r$, $\gamma_r = 1 - \beta/\lambda_r$
 $g_k = [\sum_{r=1}^m h_r \gamma_r^k + k \sum_{r=1}^m \delta_r \gamma_r^{k-1} (1 - \gamma_r)]/2$ $c_0 = \prod_{r=1}^m r = 1^m (\frac{\beta}{\lambda_r})^{\frac{h_r}{2}}; \exp(-\frac{\sum_{r=1}^m \delta_r}{2})$;
 $c_k = k^{-1} \sum_{k=0}^{k-r} g_{k-r} c_r$ for $k \geq 1$. If the series is truncated after N terms, the truncation error is bounded by

$$\sum_{k=N+1}^{\infty} c_k Pr(\chi_{2k+\tilde{h}}^2 > \frac{t}{\beta}) \leq \sum_{k=N+1}^{\infty} c_k = 1 - \sum_{k=1}^N c_k$$

This method also can be very accurate by bounding the error. However, it needs lots of computation and the eigenvalues need be specified explicitly. In later comparison, we use this method as our golden standard.

2.3 Pearson's three-moment central χ^2 Approximation

The Pearson (1959) approach was originally developed for approximating a non-central $\chi_l^2(\delta)$ distribution by a central $\chi_{l^*}^2$ distribution:

$$Pr\left(\frac{\chi_l^2(\delta) - \mu_X}{\sigma_X} > t^*\right) \approx Pr\left(\frac{\chi_{l^*}^2 - l^*}{\sqrt{2l^*}} > t^*\right) \quad (2.5)$$

where $\chi_l^2(\delta)$ and $\chi_{l^*}^2$ have equal skewness. Imhof (1961) extended Pearson's method to the non-negative quadratic form:

$$Pr(Q(X) > q) \approx Pr(\chi_{l^*}^2 > l^* + t^* \sqrt{2l^*}) \quad (2.6)$$

where l^* is determined so that $Q(X)$ and $\chi_{l^*}^2$ have equal skewness. The cumulant generating function of $Q(X)$ is given by Imhof (1961)

$$K(t) = -\frac{1}{2} \sum_{r=1}^m h_i \log(1 - 2t\lambda_r) + \sum_{r=1}^m \frac{\delta_r \lambda_r t}{1 - 2\lambda_r t} \quad (2.7)$$

The formula for the k th cumulant of $Q(X)$ is

$$\kappa_k = 2^{k-1}(k-1)! \left(\sum_{r=1}^m \lambda_r^k h_r + k \sum_{r=1}^m \lambda_r^k \delta_r \right) \quad (2.8)$$

It is necessary to calculate the λ_r 's and δ_r 's, since

$$\begin{aligned} \sum_{r=1}^m \lambda_r^k h_r &= \text{trace}(\Lambda^k) = \text{trace}((PBP')^k) = \text{trace}(B^k) = \text{trace}((A\Lambda)^k), \\ \sum_{r=1}^m \lambda_r^k \delta_r &= \mu'_y \Lambda^k \mu_y = \mu'_x (A\Sigma)^{k-1} A \mu_x. \end{aligned}$$

The mean, standard deviation and skewness of $Q(X)$ are given by

$$\mu_Q = \kappa_1 = c_1, \sigma_Q = \sqrt{\kappa_2} = \sqrt{2c_2}, \beta_1 = \frac{\kappa_3}{\kappa_2^{3/2}} = \sqrt{8}s_1,$$

where $c_k = \sum_{r=1}^m \lambda_r^k h_r + k \sum_{r=1}^m \lambda_r^k \delta_r$, $s_1 = c_3/c_2^{3/2}$.

In Pearson's approximation, $l^* = 1/s_1^2$. The accuracy of this method has been demonstrated in many empirical studies. Also, like other moment-based approaches, this method does not involve inverting a matrix or calculating the eigenvalues of a matrix.

2.4 Liu-Tang-Zhang's four-moment non-central χ^2 approximation

LTZ proposed this method using a non-central $\chi_l^2(\delta)$ distribution to approximate the distribution of Q , thus

$$Pr(Q(X) > q) \approx Pr(\chi_l^2(\delta) > t^* \sigma_X + \mu_X) \quad (2.9)$$

where $t^* = \frac{t - \mu_Q}{\sigma_Q}$, $\mu_x = l + \delta$, $\sigma_X = \sqrt{2(l + 2\delta)}$. The parameters δ and l are determined so that the skewness of $Q(X)$ and $\chi_l^2(\delta)$ are equal and the difference between the kurtosis of $Q(X)$ and $\chi_l^2(\delta)$ is minimized.

From (2.7), the kurtosis of $Q(X)$ is $\beta_2 = \frac{\kappa_4}{\kappa_2^2} = 12s_2$ and $s_2 = c_4/c_2^2$.

If $s_1^2 > s_2$, the minimum difference between the kurtosis $Q(X)$ and $\chi_l^2(\delta)$ is 0 when $s_1^2 - s_2 = (1/a - s_1)^2$. Thus

$$a = 1/(s_1 - \sqrt{s_1^2 - s_2}), \delta = s_1 a^3 - a^2 \text{ and } l = a^2 - 2\delta.$$

If $s_1^2 \geq s_2$, the difference between the kurtosis of $Q(X)$ and $\chi_l^2(\delta)$ is minimized when $a = 1/s_1$. Thus

$$a = 1/s_1, \delta = s_1^3 - a^2 = 0 \text{ and } l = 1/s_1^2 = c_2^3/c_3^2.$$

This method is closely related to Pearson's approximation and is equivalent to Pearson's approach when $s_1^2 \geq s_2$. The better high-order moment approximation often improves the tail probability approximation in practice. And also this method does not involve inverting a matrix or calculating the eigenvalues of a matrix. In the next section, we will use a simulation study to show that the actual approximation error of this method is usually much smaller than that of Pearson's approximation.

2.5 Kuonen's saddle point approximation

By contrast to the Pearson's method, this method uses the entire cumulant generating function (2.7).

$$Pr(Q(X) > q) \approx 1 - \Phi(w + \frac{1}{w} \log(\frac{v}{w})) \quad (2.10)$$

where

$$w = \text{sign}(\hat{\xi})[2(\hat{\xi}q - K(\hat{\xi}))]^{\frac{1}{2}}, v = \hat{\xi}(K''(\hat{\xi}))^{\frac{1}{2}}$$

and $\hat{\xi} = \hat{\xi}(q)$ is the saddle point which satisfies the equation $K'(\hat{\xi}) = q$; $K'(\xi)$ and $K''(\xi)$ are the first and second derivatives of $K(\xi)$ with respect to ξ .

Saddle point method gives high accuracy and is very easy to compute. However, the eigenvalues must be specified explicitly.

3 Numerical examples simulation

We consider five quadratic forms, which are defined in table 1.

$Q_1 = 0.5\chi_1^2(1) + 0.4\chi_2^2(0.6) + 0.1\chi_1^2(0.8)$
$Q_2 = 0.9\chi_1^2(0.2) + 0.1\chi_2^2(10)$
$Q_3 = 0.1\chi_1^2(0.2) + 0.9\chi_2^2(10)$
$Q_4 = 0.5\chi_1^2(1) + 0.4\chi_2^2(0.6) + \sum_{r=1}^{10} 0.01\chi_r^2(0.8)$
$Q_5 = \chi_1^2(1.0) + (0.6)^4\chi_1^2(7.0)$

Table 1

The quadratic form Q_1 is inspired from Liu et al.(2009). Q_2 is a quadratic form with small weight assigned to the chi-square term with large non-centrality parameter. Q_3 is a quadratic form with large weight assigned to the chi-square term with large non-centrality parameter. These two quadratic forms are designed to check whether the weight and non-centrality parameter will affect the accuracy of approximation. Q_4 is a modification of Q_1 . We decomposed the last term into 10 chi-square variables with same weight 0.01 and different degrees of freedoms. By comparing the results of Q_1 and Q_4 , we want to check whether the quantity of chi-square variables will have an effect on the performance of different approximation approaches. Q_5 is inspired from the asymptotic distribution of Box-Pierce-Ljung portmanteau test statistic, which is widely used in time series analysis. If a time series follows a ARMA model with parameter α , then the Box-Pierce-Ljung portmanteau test statistic asymptotically follows the quadratic form $\chi_{m-1}^2 + \alpha^{2m}\chi_1^2$. We set $m = 2$, $\alpha = 0.6$.

For each quadratic form, we simulated 101 different q values. For Q_1 , Q_2 , Q_4 and Q_5 , the sequence of q is from 1 to 11 with increment 0.1. To briefly analyze the results, we only include the results of $q = 2, 4, 6, 8, 10$ in the absolute error table and relative error table. For Q_3 , the sequence of q is from 1 to 20 with increment 0.2. The reason that the range of t is different for Q_3 is that

when $t = 10$, the value of (1.1) of quadratic form Q_3 is approximate 0.4, which is still quite large. So we extended the range of q values to make the analysis have more realistic meaning. We used all simulated points to construct qq plot. The results of these five quadratic forms in five approaches were presented in the next section.

<i>imhof</i> ($q, \lambda, h, \delta, epsabs = 10^{-6}, epsrel = 10^{-6}$)
<i>farebrother</i> ($q, \lambda, h, \delta, epsabs = 10^{-6}, eps = 10^{-6}$)
<i>liu</i> (q, λ, h)

Table 2

The computations of Farebrother's method, Imhof's method and LTZ's method were all performed using R package CompQuadForm developed by P.Lafaye in 2015 and the results were identical with a precision of 10^{-6} . The commands were presented in Table 2. Note that although LTZ's approximation does not involve calculating the eigenvalues of $Q(X)$, these values still need be specified explicitly in order to use this R package.

For Pearson's approach and Saddle point approach, the R demos were programmed according to their algorithms. R codes were presented in Figure 1 and Figure 2. We just used Q_1 as an example.

Figure 1: R demo of Pearson's approximation

Figure 2: R demo of Saddle point approximation

4 Comparisons and discussion

Table 3

The column **AE_I** gives the absolute errors of Imhof's approximation compared to Farebrother's approximation. The column **AE_P** gives the absolute errors of Pearson's three-moment center χ^2 approximation compared to Farebrother's approximation. The column **AE_L** gives the absolute errors of LTZ's four approximation compared to Farebrother's approximation. The column **AE_S** gives the absolute errors of Saddle point approximation compared to Farebrother's approximation.

Figure 3: qq plot of Q_2 (red points) and Q_3 (blue points)

Table 4

The column **RE_I** gives the relative error of Imhof's approximation compared to Farebrother's approximation. The column **RE_P** gives the relative errors of Pearson's three-moment center χ^2 approximation compared to Farebrother's approximation. The column **RE_L** gives the relative errors of LTZ's four approximation compared to Farebrother's approximation. The column **RE_S** gives the relative errors of Saddle point approximation compared to Farebrother's approximation.

Our numerical examples demonstrate that Farebrother's approximation and Imhof's approximation differ very little. The accuracy of Pearson's approximation is quite good except for Q_4 . LTZ's method yields better approximation than Pearson's method except Q_2 . Since for quadratic form Q_2 , it satisfies the condition that $s_1^2 \leq s_2$, under which case, the LTZ's approximation and Pearson's approximation are the same as we illustrated in section 2. Saddle point approximation also gives high accuracy but overallly it is not as good as LTZ's method. Pearson's approximation, LTZ's approximation and Saddle point approximation yield more errors when used to calculate tail probability of Q_2 compared to other quadratic forms.

We now discuss the results in more details. To check whether the weight and scale of non-centrality parameter have effect on the accuracy among different approaches, we constructed qqplots of Q_2 and Q_3 with y axes the results of Farebrother's approximation as in Figure 3. We can conclude that both Pearson's approximation and LTZ's approximation yield high accuracy in the case that the chi-square term with large non-centrality parameter has large weight.

To check whether the quantity of chi-square variables influences the accuracy of approximation, we constructed a qqplot comparing Q_1 and Q_4 as presented in Figure 4. The results indicate that Pearson's approximation performs poorly if the quantity of chi-square variable is large. Both LTZ's approximation and Saddle point approximation yield high accuracy for these two quadratic forms.

A qqplot of the results of Q_5 is also constructed as presented in Figure 5. We could conclude that all methods yield high accuracy.

Figure 4: qq plot of Q_1 (red points) and Q_4 (blue points)

Figure 5: qq plot of Q_5 : Pearson-red points, LTZ-blue dots, Saddle point-green dots

5 Conclusion

Our numerical results suggest that Farebrother's method and Imhof's method differ very little. Pearson's method is not applicable when the quantity of chi-square terms is large. Saddle point approximation performs well as a whole. LTZ's method performs much better compared to Pearson's method and it is the most accuracy except the Imhof's method. This method shows great advantage especially if the matrix A is complex and high-dimensional and thus has great realistic meaning.