

# Saddlepoint approximations for distributions of quadratic forms in normal variables

BY D. KUONEN

*Department of Mathematics, Swiss Federal Institute of Technology, 1015 Lausanne,  
Switzerland*

diego.kuonen@epfl.ch

## SUMMARY

An extensive study and several applications show that a saddlepoint approximation for the distribution of quadratic forms in normal variates works very well and outperforms existing approximate methods in accuracy.

*Some key words:* Chi-squared variable; Nonparametric regression; Normal variable; Pseudolikelihood ratio test; Quadratic form; Saddlepoint approximation.

## 1. INTRODUCTION

Let  $X = (X_1, \dots, X_n)^T$  be a multivariate normal random vector with mean vector  $\mu = (\mu_1, \dots, \mu_n)^T$  and covariance matrix  $\Omega$ . The quadratic form associated with the  $n \times n$  matrix  $A$  is defined as

$$Q(X) = X^T A X = \sum_{i=1}^n \sum_{j=1}^n a_{ij} X_i X_j; \quad (1)$$

without loss of generality we assume that  $A$  is symmetric. Quadratic forms enter into many statistics associated with normally distributed random variables, so we may want to calculate

$$\text{pr} \{Q(X) > q\}, \quad (2)$$

where  $q$  is a given scalar. In the simplest case,  $A = \Omega = I_n$  and  $Q(X)$  is a noncentral chi-squared variable with  $n$  degrees of freedom and noncentrality parameter  $\xi^2 = \sum \mu_i^2$  (Scheffé, 1959, Appendix IV). If the matrix  $A$  is neither idempotent nor positive definite, classical results such as Cochran's theorem (Scheffé, 1959, Appendix VI) implying a chi-squared distribution for the quadratic form do not apply, and another approach to the calculation of (2) is needed. Johnson & Kotz (1970, Ch. 29) discuss representations of the distributions of quadratic forms. Imhof (1961) gives exact methods for computing (2) using real arithmetic. This method has been programmed in Fortran by Koerts & Abrahamse (1969) and in Pascal by Farebrother (1990). Moreover, as any quadratic form in independent normal variables can be reduced to a linear combination of chi-squared variables, the Algol algorithm of Davies (1980) can also be used; a C version is available.

This note proposes a saddlepoint approximation to (2). It is comparable in speed with exact methods, almost as accurate and is much easier to program. In § 2, theory for quadratic forms is recalled. Section 3 presents the saddlepoint approximation to (2). The methods are compared in § 4, and § 5 illustrates their use in nonparametric regression.

## 2. THE GENERAL QUADRATIC FORM

We first consider a central quadratic form in which  $\mu_1 = \dots = \mu_n = 0$ . Since  $\Omega$  is positive definite and symmetric, it can be factored by Choleski decomposition as  $\Omega = \Gamma \Gamma^T$ , where  $\Gamma$  is a nonsingular

lower triangular matrix. It follows that  $\Gamma^T A \Gamma$  is symmetric and therefore its eigenvalues are all real. Thus there is a nonsingular orthogonal linear transformation of  $X$  such that  $Q(X)$  has the same distribution as  $Q(Y) = \sum \lambda_i Y_i^2$ , where the  $Y_i$  are independent standard normal variables and  $\lambda_1 \geq \dots \geq \lambda_n$  denote the eigenvalues of  $\Gamma^T A \Gamma$  (Scheffé, 1959, Appendix II).

For the general noncentral case a similar reduction is (Imhof, 1961)

$$Q(Y) = \sum_{i=1}^n \lambda_i \chi_{h_i; \sigma_i^2}^2,$$

where the  $h_i$  are the orders of multiplicity of the  $\lambda_i$ , the  $\chi_{h_i; \sigma_i^2}^2$  are independent chi-squared variables with  $h_i$  degrees of freedom and noncentrality parameters  $\sigma_i^2$ , and they are defined by the relation

$$\chi_{h_i; \sigma_i^2}^2 = (Y_1 + \sigma_i)^2 + \sum_{r=2}^{h_i} Y_r^2,$$

where the  $Y_i$  are independent standard normal deviates. The cumulant generating function of  $Q(Y)$  is (Imhof, 1961, eqn (2.3); Johnson & Kotz, 1970, p. 152)

$$K(\zeta) = -\frac{1}{2} \sum_{i=1}^n h_i \log(1 - 2\zeta \lambda_i) + \sum_{i=1}^n \frac{\sigma_i^2 \lambda_i}{1 - 2\zeta \lambda_i}, \quad \zeta < \frac{1}{2} \min_i \lambda_i^{-1}.$$

Note that  $\lambda_1 \geq \dots \geq \lambda_n$  must be determined explicitly. Imhof (1961) extended Pearson's (1959) three-moment central chi-squared approximation to the distribution of noncentral chi-squared variables. A special case of this is  $\text{pr}\{Q(X) > q\} \approx \text{pr}\{\chi_b^2 > r\}$ , where  $\chi_b^2$  denotes a chi-squared variable with  $b = c_2^3/c_3^2$  degrees of freedom,  $r = (q - c_1)(b/c_2)^{\frac{1}{2}} + b$  and  $c_s = \sum \lambda_i^s (h_i + s\sigma_i^2)$ , for  $s = 1, 2, 3$ . If the quadratic form is nonpositive one has to assume that  $c_3 > 0$ . Otherwise, one must approximate the distribution of  $-Q(X)$ .

In the central case, where  $\sigma_1^2 = \dots = \sigma_n^2 = 0$ ,  $h_1 = \dots = h_n = 1$ ,

$$c_s = \sum_{i=1}^n \lambda_i^s = \text{tr}\{(\Gamma^T A \Gamma)^s\} = \text{tr}\{(\Omega A)^s\},$$

because  $\Gamma^T A \Gamma$  and  $\Omega A$  have the same eigenvalues. In the central case, Bowman & Azzalini (1997) rewrite Pearson's approximation in terms of the first three cumulants of  $Q(X)$ , which are

$$\kappa_s = 2^{s-1}(s-1)! \text{tr}\{(\Omega A)^s\} = 2^{s-1}(s-1)!c_s$$

(Johnson & Kotz, 1970, p. 153). They fit a distribution of type  $a\chi_b^2 + c$ , where  $a = |\kappa_3|/(4\kappa_2)$ ,  $b = (8\kappa_2^3)/\kappa_3^2$  and  $c = \kappa_1 - ab$ . The advantage is that the eigenvalues only enter through  $\text{tr}\{(\Omega A)^s\}$  and therefore need not be calculated explicitly.

### 3. SADDLEPOINT APPROXIMATION

Saddlepoint methods (Daniels, 1954; Reid, 1988; Jensen, 1995) give highly accurate approximations to density and distribution functions. By contrast with Pearson's approximation they use the entire cumulant generating function. One form of saddlepoint approximation to the distribution of  $Q(X)$  or  $Q(Y)$  at  $q$  is (Barndorff-Nielsen, 1990)

$$\text{pr}\{Q(X) > q\} \approx 1 - F_s(q) = 1 - \Phi\left\{w + \frac{1}{w} \log\left(\frac{v}{w}\right)\right\}, \quad (3)$$

where

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

and  $\hat{\zeta} = \hat{\zeta}(q)$ , known as the saddlepoint, is the value of  $\zeta$  satisfying the equation  $K'(\hat{\zeta}) = q$ ;  $K'(\zeta)$  and  $K''(\zeta)$  are the first and second derivatives of  $K(\zeta)$  with respect to  $\zeta$ . An alternative to (3) is the Lugannani–Rice approximation (Lugannani & Rice, 1980), but they are usually indistinguishable in practice. The existence and uniqueness of  $\hat{\zeta}$  follow from Daniels (1954). Calculation of (3) requires computation of  $\hat{\zeta}$  for each  $q$  of interest. Under additional conditions (3) gives a relative error of  $O(n^{-3/2})$  for the distribution (Jensen, 1995, § 6.5). When  $q$  is close to  $E\{Q(X)\}$ , approximation (3) can become numerically unstable, and is best avoided. Fortunately such values of  $q$  are not of interest in practice.

#### 4. COMPARISON

Simulation shows that Imhof's (1961) and Davies's (1980) exact methods differ little so only Davies's method will be considered here. In an extensive numerical study, the author considered 39 different quadratic forms. For the 12 most interesting of them, Table 1 records the values of (2) found by Davies's method. For the approximations the relative error compared to this method is given. The quadratic forms  $Q_1$  to  $Q_7$  are represented in the first column of Table 1 by  $\lambda_i$ ,  $h_i$  and  $\sigma_i^2$ , while  $Q_8$  to  $Q_{12}$  are represented by  $A$ ,  $\mu$  and  $\Omega$ . The forms are as follows:  $F_n$  is the symmetric banded  $n \times n$  matrix with bandwidth one, diagonal  $(1, 2, \dots, 2, 1)$  and subdiagonal consisting of  $-1$ 's (Farebrother, 1990);  $S_n$  is a positive definite symmetric banded  $n \times n$  matrix with bandwidth two,  $7$ 's on the diagonal,  $2$ 's in the first band and  $1$ 's in the second band; and  $D_n$  is a positive definite symmetric banded  $n \times n$  matrix with bandwidth one,  $10$ 's on the diagonal, and  $-1$ 's in the subdiagonal, in the  $n$ th line and in the  $n$ th column. Quadratic forms  $Q_1$  to  $Q_3$  were used by Imhof (1961, Table 1) and  $Q_4$  and  $Q_5$  were used by Farebrother (1990, Table 1).

Pearson's three-moment central chi-squared approximation (ii) encounters problems for negative definite or indefinite quadratic forms, such as  $Q_6$ ,  $Q_9$ ,  $Q_{10}$ ,  $Q_{11}$  and  $Q_{12}$ , and its relative error can be large. The saddlepoint approximation (iii) yields very accurate approximations, even for nonpositive quadratic forms. Even with  $n$  large, as in  $Q_{11}$  and  $Q_{12}$ , the saddlepoint approximation (iii) is more precise than (ii).

The saddlepoint approximation (3) is very easy to compute: in the statistical package S-Plus it takes about 20 lines of commands, whereas Davies's method requires more than 200 lines of C code, and to have an estimate of desired accuracy one has to adjust all the input variables. Thus the saddlepoint approximation can be applied more easily, though there is some additional cost in computation time because of the use of an interpreted language like S-Plus, and to evaluation of all the eigenvalues; this is of the order of a few seconds when  $n = 300$ .

#### 5. APPLICATION TO NONPARAMETRIC REGRESSION

Suppose that data  $(x_1, y_1), \dots, (x_n, y_n)$  of dimension  $(d + 1)$  have been collected. Their regression relationship can be modelled as

$$y_i = m(x_i) + \varepsilon_i \quad (i = 1, \dots, n),$$

where  $m(\cdot)$  is an unknown function and the  $\varepsilon_i$  are independent errors with mean zero and variance  $\sigma^2$ . A nonparametric estimator of  $m(x)$ , such as the local least squares estimator, can be written as  $\hat{m}_H(x) = S_H y$ , where  $S_H$  is the  $n \times n$  smoothing matrix and  $H$  is a  $d \times d$  symmetric positive matrix, the bandwidth matrix (Ruppert & Wand, 1994). Suppose that one wants to compare the hypotheses  $H_0: E(y_i) = \mu$  and  $H_1: E(y_i) = m(x_i)$ ;  $H_0$  posits no effect. The standard approach from classical linear models was extended by Azzalini, Bowman & Härdle (1989) to the nonparametric setting. They were led to the so-called pseudolikelihood ratio test statistic,  $T = (\text{RSS}_0 - \text{RSS}_1) / \text{RSS}_1$ , where  $\text{RSS}_i$  denotes the residual sum of squares under the hypothesis  $H_i$  ( $i = 0, 1$ ). An explicit expression is  $T = y^T B y / y^T C y$ , where  $C = (I_n - S)^T (I_n - S)$ ,  $B$  is the matrix  $I_n - n^{-1}L - C$  and  $L$  is the  $n \times n$  matrix with all its entries equal to one (Bowman & Azzalini, 1997, § 5.2). Hence, the corresponding

Table 1. Probability that the quadratic form exceeds  $q$ . (i) Davies's (1980) method with accuracy 0.0001, (ii) relative error (%) for Pearson's (1959) three-moment central chi-squared approximation, and (iii) relative error (%) for the saddlepoint approximation

Quadratic form	$q$	(i)	(ii)	(iii)
$Q_1 = 0.6, 2, 0; 0.3, 2, 0; 0.1, 2, 0$	0.2	0.9936	0.65	0.007
	3	0.1869	0.89	0.727
	6	0.0161	0.76	0.709
$Q_2 = 0.6, 6, 0; 0.3, 4, 0; 0.1, 2, 0$	1	0.9973	0.11	0.001
	3	0.8156	0.19	0.031
	10	0.0311	0.39	0.132
$Q_3 = 0.6, 2, 0; 0.3, 4, 0; 0.1, 6, 0$	1	0.9666	1.04	0.026
	4	0.2115	0.99	0.260
	8	0.0087	1.85	0.042
$Q_4 = 30, 1, 0; 1, 10, 0$	5	0.9846	1.56	0.058
	25	0.4892	1.76	1.021
	100	0.0837	0.69	0.266
$Q_5 = 30, 1, 0; 1, 20, 0$	10	0.9951	0.49	0.010
	40	0.4267	1.98	0.942
	100	0.1035	1.35	1.152
$Q_6 = -0.6, 2, 0; 0.3, 4, 0; 0.1, 6, 0$	-4	0.9900	1.01	0.026
	1	0.4010	12.38	3.117
	4	0.0098	168.48	1.813
$Q_7 = 0.6, 1, 0.1; 0.3, 2, 0.2; 0.1, 1, 0.2$	0.1	0.9883	1.18	0.001
	0.4	0.8694	1.14	0.337
	3.2	0.0775	3.35	1.363
$Q_8 = F_5; 0_5; I_5$	0.5	0.9881	1.20	0.017
	5	0.5949	1.77	0.519
	50	0.0006	11.70	1.578
$Q_9 = -F_5; 0_5; I_5$	-25	0.9731	2.77	0.050
	-5	0.4051	40.78	0.761
	-0.5	0.0119	913.46	1.423
$Q_{10} = -S_7; 0_7; I_7$	-200	0.9993	0.07	0.002
	-50	0.5976	28.68	0.523
	-5	0.0023	3341.06	0.714
$Q_{11} = -S_{15}; 0_{15}; I_{15}$	-200	0.9722	2.86	0.026
	-100	0.5106	22.94	0.231
	-50	0.0580	77.49	0.282
$Q_{12} = -D_{10}; 0_{10}; I_{10}$	-200	0.9687	3.24	0.004
	-50	0.1124	19.97	0.002
	-10	0.0002	24457.37	0.008

$p$ -value can be written as

$$\text{pr}(T > t | H_0) = \text{pr}(y^T U y > 0) = \text{pr}(\varepsilon^T U \varepsilon > 0), \quad (4)$$

where  $U = B - tC$ ,  $t$  is the observed value of  $T$  and  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T$  is the error vector. If the  $\varepsilon_i$  are normally distributed one can set  $\sigma^2 = 1$  without loss of generality as  $T$  is scale-invariant. Hence, if we set  $A = U$  and  $\Omega = I_n$ , the covariance matrix of  $\varepsilon$ , the  $p$ -value (4) can be calculated easily.

For illustration, we consider the Great Barrier Reef data for the closed zone in 1993 (Bowman & Azzalini, 1997, p. 52). Two of its variables are the bottom depth,  $x$ , and the catch score,  $y$ . A

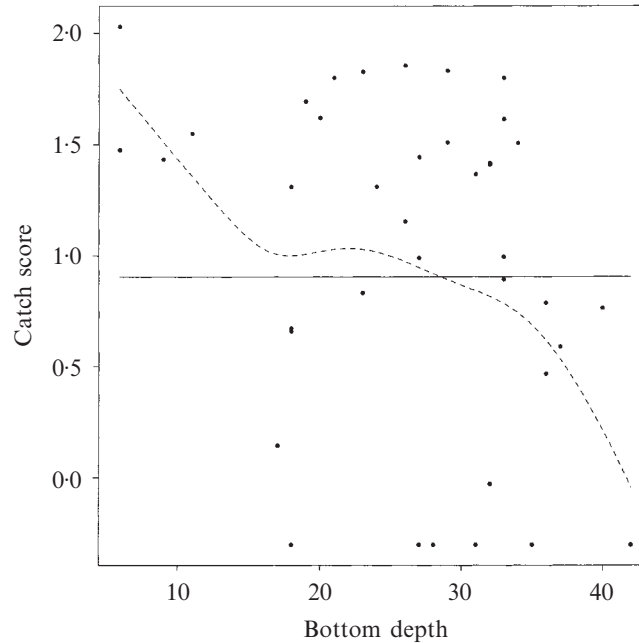


Fig. 1. Plot of the relationship between bottom depth (metres) and catch score in a subsample of the Great Barrier Reef data. The nonparametric regression curve with  $h = 5$ , dashed line, and the curve of no effect, solid line, have been superimposed.

Table 2. The  $p$ -value as a function of the smoothing parameter  $h$  for testing a relationship between catch score and bottom depth in the Great Barrier Reef data. (i) Davies's (1980) method with accuracy 0.0001, (ii) relative error (%) for Pearson's (1959) three-moment central chi-squared approximation, and (iii) relative error (%) for the saddle-point approximation

$h$	(i)	(ii)	(iii)	$h$	(i)	(ii)	(iii)
3	0.0604	5.433	0.028	11	0.0334	5.890	1.192
5	0.0633	7.737	0.135	13	0.0289	4.867	1.288
7	0.0516	7.545	0.019	15	0.0258	4.164	0.240
9	0.0407	6.765	0.738	17	0.0234	3.460	1.431

plot of their relationship is given in Fig. 1. Here  $d = 1$ , so  $H$  is a scalar smoothing parameter  $h$ . Figure 1 displays the nonparametric regression curve with  $h = 5$  and the curve of no effect. Table 2 gives the  $p$ -value (4) as a function of the smoothing parameter  $h$ . Bowman & Azzalini (1997) stated that their approach (ii) is sufficiently accurate to approximate the  $p$ -value, but Table 2 shows that Pearson's approximation underestimates the exact  $p$ -value. The saddlepoint approximation is extremely accurate. For a significance level of 5%, the  $p$ -values recorded in Table 2 suggest that there exists a relationship between bottom depth and catch score for  $h > 7$ , but not for  $h \leq 7$ . This may result from the fact that when  $h$  increases the nonparametric local regression curve becomes smoother and differs from the fitted constant curve significantly.

In practice it is useful to have some idea of the power of a test. Suppose that  $y_1, \dots, y_n$  are generated from  $y_i = \hat{m}_h(x_i) + e_i$ , where  $\hat{m}_h(x_i)$  is the local least squares estimate evaluated at  $x_i$ , and the  $e_i$  are drawn independently from the normal distribution  $N(0, \eta^2)$ . Various combinations of  $h$

Table 3. The simulated power (%) as a function of  $h$  and  $\eta$  for testing a relationship between catch score and bottom depth in the Great Barrier Reef data. (i) Davies's (1980) method with accuracy 0.0001, (ii) Pearson's (1959) three-moment central chi-squared approximation, and (iii) saddlepoint approximation

$h$	$\eta$	(i)	(ii)	(iii)	$h$	$\eta$	(i)	(ii)	(iii)
1	1	93.4	93.4	93.4	9	1	94.6	94.6	94.6
	0.5	94.6	94.4	94.6		0.5	95.0	94.8	95.0
	0.25	95.4	95.0	95.4		0.25	95.6	95.2	95.6
5	2	93.4	93.0	93.4	11	2	94.4	93.8	94.4
	1	94.0	93.8	94.0		1	95.0	94.2	95.0
	0.25	94.8	94.4	94.8		0.5	96.0	95.8	96.0
7	2	94.0	93.8	94.0					
	1	94.6	94.6	94.6					
	0.1	95.4	95.4	95.4					

and  $\eta$  were considered and in each case 500 samples were generated. Table 3 presents the percentages of times  $H_0$  was rejected for significance level 5% and some selected values of  $h$  and  $\eta$ . The pseudolikelihood ratio test seems to be very powerful in the present context, though the power decreases when  $\eta$  gets large, because large  $\eta$ 's increase the noise. The saddlepoint approximation gives the same power as does Davies's method, whereas Pearson's approximation delivers slightly smaller estimated power.

#### ACKNOWLEDGEMENT

The author thanks A. C. Davison for his helpful comments and suggestions. Many thanks go as well to R. B. Davies for the C version of his algorithm, and to the referees and the editor for their valuable suggestions. The author was supported by a grant from the Swiss National Science Foundation.

#### REFERENCES

- AZZALINI, A., BOWMAN, A. W. & HÄRDLE, W. (1989). On the use of nonparametric regression for model checking. *Biometrika* **76**, 1–11.
- BARNDORFF-NIELSEN, O. E. (1990). Approximate interval probabilities. *J. R. Statist. Soc. B* **52**, 485–96.
- BOWMAN, A. W. & AZZALINI, A. (1997). *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations*. Oxford: Oxford University Press.
- DANIELS, H. E. (1954). Saddlepoint approximations in statistics. *Ann. Math. Statist.* **25**, 631–50.
- DAVIES, R. B. (1980). Algorithm AS 155: the distribution of a linear combination of  $\chi^2$  random variables. *Appl. Statist.* **29**, 323–33.
- FAREBROTHER, R. W. (1990). Algorithm AS 256: the distribution of a quadratic form in normal variables. *Appl. Statist.* **39**, 294–309.
- IMHOF, J. P. (1961). Computing the distribution of quadratic forms in normal variables. *Biometrika* **48**, 419–26.
- JENSEN, J. L. (1995). *Saddlepoint Approximations*. Oxford: Oxford University Press.
- JOHNSON, N. L. & KOTZ, S. (1970). *Distributions in Statistics: Continuous Univariate Distributions*, **2**. New York: Wiley.
- KOERTS, J. & ABRAHAMSE, A. P. J. (1969). *On the Theory and Application of the General Linear Model*. Rotterdam: Rotterdam University Press.
- LUGANNANI, R. & RICE, S. O. (1980). Saddlepoint approximation for the distribution of the sum of independent random variables. *Adv. Appl. Prob.* **12**, 475–90.

- PEARSON, E. S. (1959). Note on an approximation to the distribution of non-central  $\chi^2$ . *Biometrika*, **46**, 364.
- REID, N. (1988). Saddlepoint methods and statistical inference (with Discussion). *Statist. Sci.* **3**, 213–38.
- RUPPERT, D. & WAND, M. P. (1994). Multivariate locally weighted least squares regression. *Ann. Statist.* **22**, 1346–70.
- SCHEFFÉ, H. (1959). *The Analysis of Variance*. New York: Wiley.

[Received June 1998. Revised April 1999]