AN APPROXIMATION FOR THE NONCENTRAL CHI-SQUARED DISTRIBUTION

D.A.S. Fraser
Department of Statistics
University of Toronto
Toronto, Canada. M5S 3G3

A.C.M. Wong
Department of Mathematics & Statistics
York University
North York, Canada. M3J 1P3

J. Wu
Department of Statistics
University of Toronto
Toronto, Canada. M5S 3G3

Key Words and Phrases: Asymptotic inference; Accurate approximation; Chi-squared distribution; Significance function

ABSTRACT

A simple and accurate method is discussed for approximating the noncentral chi-squared distribution; it is based on recent developments in third order asymptotic methods. The method is easy to apply and uses only a standard normal distribution function evaluation.

1. INTRODUCTION

The noncentral chi-squared random variable r^2 with p degrees of freedom and noncentrality parameter ρ^2 can be described by

$$r^2 = \sum_{i=1}^{p} z_i^2$$

where the z_i are independent normally distributed variables with means μ_i and variances 1, and

$$\rho^2 = \sum_{i=1}^{p} \mu_i^2.$$

The corresponding distribution function can be expressed, (for example, Johnson & Kotz 1970, p. 132) as

$$G_{\rho^2}(x) = P(r^2 \leq x) = \sum_{j=0}^{\infty} \frac{e^{-\rho^2/2}(\rho^2/2)^j}{j!} P(\chi^2_{p+2j} \leq x); \tag{1.1}$$

where $P(\chi^2_{p+2j} \leq x)$ is the distribution function of a chi-squared variable with $(p+2j)$ degrees of freedom.

The noncentral chi-squared distribution is an important distribution and is often used to calculate the power of tests on the mean of a multivariate normal distribution; see Anderson (1975, p. 75) and Patnaik (1949) for some discussion of various applications.

To avoid the infinite sum in (1.1), various approximations have been proposed. In particular, by inverting the cumulant generating function and using an Edgeworth expansion, Cox & Reid (1987) obtained the approximation

$$P(r^2 \leq x) \approx P\left(\chi^2_p \leq \frac{x}{1+\rho^2/p}\right). \tag{1.2}$$

An asymptotically equivalent approximation was also given in Cox & Reid (1987),

$$P(r^2 \leq x) \approx P\left(\chi_p^2 \leq x\left(1 - \frac{\rho^2}{p}\right)\right), \tag{1.3}$$

but this modification was not always satisfactory. An approximation closest in spirit to (1.2) and (1.3) was given by Bol'shev & Kuznetzov (1963):

$$P(r^2 \leq x) = P\left[\chi_p^2 \leq x\left(1 - \frac{\rho^2}{p} + \frac{1}{2}\rho^4 \frac{1 + x/(p+2)}{p^2}\right)\right] + 0(\rho^6). \tag{1.4}$$

Note that ignoring the third term in the braces gives (1.3) which is asymptotically equivalent to (1.2).

Wang & Gray (1993) suggested using the $G_p^{(m)}$-transform to approximate (1.1). The method requires the evaluation of an infinite sum and also the derivatives of a complicated function $f(x)$ defined in their equation (4). Even though this method is extremely accurate, it is not nearly as simple as the other approximations.

Cohen (1988) provided a procedure for evaluating a noncentral chi-squared distribution. The method requires the availability of a tabulation of the three lowest degrees of freedom of the noncentral chi-squared distribution function or equivalently an effective computer algorithm for their evaluation; also it requires recursive evaluations. Posten (1989) provided another recursive algorithm for evaluating the noncentral chi-squared distribution function.

In this paper, a new approximation is proposed. The method is simple, has third order accuracy, and is derived from a general inference procedure developed in Fraser & Reid (1995); it is particularly accurate in the extreme tail.

In Section 2, we discuss briefly the recent third order procedure. In Section 3, we show how this method can be applied to approximate the density and distribution functions for the noncentral chi-squared distribution and how it can approximate a percentile for that distribution. Examples are given in Section 4 and some concluding remarks are recorded in Section 5.

2. THIRD ORDER SIGNIFICANCE

The recent third order approximations have evolved from the saddlepoint method (Daniels, 1954; Barndorff-Nielsen & Cox, 1979) or from the direct analysis and Taylor series expansion of log density functions (Barndorff-Nielsen, 1986; DiCiccio, Field & Fraser, 1990; Fraser, 1990; Fraser & Reid, 1995). These third order approximations are typically based on likelihood and can give high accuracy even for very small samples.

Third order significance for testing whether a scalar parameter, say $\psi(\theta)$, has the value ψ can be obtained in many general contexts from a Lugannani & Rice (1980) type formula

$$p(\psi) = \Phi(R) + \phi(R)\left\{\frac{1}{R} - \frac{1}{Q}\right\} \tag{2.1}$$

or from a Barndorff-Nielsen (1986) r^* type formula

$$p(\psi) = \Phi(r^*) = \Phi\left(R - R^{-1}\log\frac{R}{Q}\right) \tag{2.2}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal density and distribution functions, R is of signed log likelihood root form, and Q is a standardized maximum likelihood departure. Detailed definitions for R and Q depend on the type and generality of the problem; see for example Lugannani & Rice (1980), Barndorff-Nielsen (1986), and Fraser & Reid (1995).

Note that $p(\psi)$ in (2.1) and (2.2) is generally known as the significance function and can be interpreted as giving the probability to the left of the observed data point for any chosen ψ, in the same way as the probability left of an observed t value in normal sampling provides a significance assessment of a value for the mean. Therefore, for testing the hypothesis $H_0 : \psi = \psi_0$, the one-sided observed level of significance is

$$p(\psi_0) \text{ or } 1 - p(\psi_0)$$

and the usual two-sided observed level of significance is

$$2min\{p(\psi_0), 1 - p(\psi_0)\}.$$

In addition, a $(1-\alpha) \times 100\%$ confidence interval for ψ is

$$(min\{p^{-1}(\alpha/2), p^{-1}(1-\alpha/2)\}, max\{p^{-1}(\alpha/2), p^{-1}(1-\alpha/2)\}).$$

A detail discussion of $p(\psi)$ is given in Fraser (1991).

It was shown in Lugannani & Rice (1980) and Barndorff-Nielsen (1986) that for specific problems, (2.1) and (2.2) have third order accuracy, $O(n^{-3/2})$, in providing a uniformly distributed significance function or p-value.

The definition of R is fairly standard to all problems and is obtained from the traditional likelihood ratio. It is called the signed likelihood root and has the form

$$R = \text{sgn}(\hat{\psi} - \psi)[2\{\ell(\hat{\theta}) - \ell(\hat{\theta}_\psi)\}]^{1/2} \tag{2.3}$$

where $\ell(\theta) = \ell(\theta; y^0)$ is the observed likelihood, $\hat{\theta}$ is the maximum likelihood value for the full parameter, and $\hat{\theta}_\psi$ is the same but subject to the constraint $\psi(\theta) = \psi$. Under the hypothesis $\psi(\theta) = \psi$, R is standard normal to the first order.

The definition of Q requires more than the observed likelihood function. The additional information is obtained from the gradient of the likelihood function at the data point

$$\varphi(\theta) = \frac{d}{dy}\ell(\theta; y)\big|_{y^0} \; ; \tag{2.4}$$

this give a new exponential-type parametrization. Formula (2.4) is for the special case where the dimension of the variable y is the same as the dimension say p of the parameter θ, a case that applies for the present problem. In more general contexts the gradient is taken in p specially chosen directions at the data point (Fraser & Reid, 1995).

A scalar combination $\chi(\theta)$ of the coordinates of $\varphi(\theta)$ is then chosen to behave like $\psi(\theta)$ near $\hat{\theta}_\psi$.

$$\chi(\theta) = \frac{\psi_{\varphi'}(\hat{\theta}_\psi)}{|\psi_{\varphi'}(\hat{\theta}_\psi)|}\varphi(\theta) \tag{2.5}$$

where $\psi_{\varphi'}(\theta) = (\partial/\partial\varphi)\psi(\theta) = \psi_{\theta'}(\theta)\{\varphi_{\theta'}^{-1}(\theta)\}$ evaluated at $\hat{\theta}_\psi$ gives the linear combination. The standardized maximum likelihood departure is then given as

$$Q = \mathrm{sgn}(\hat{\psi} - \psi)|\chi(\hat{\theta}) - \chi(\hat{\theta}_\psi)| \left\{ \frac{|\hat{\jmath}_{(\theta\theta)}|}{|\jmath_{(\lambda\lambda)}(\hat{\theta}_\psi)|} \right\}^{1/2} \tag{2.6}$$

where $\hat{\jmath}_{\theta\theta} = -\ell_{\theta\theta'}(\hat{\theta})$ and $j_{\lambda\lambda}(\hat{\theta}_\psi) = -\ell_{\lambda\lambda'}(\hat{\theta}_\psi)$ are full and nuisance information matrices, and

$$|\hat{\jmath}_{(\theta\theta)}| = |\hat{\jmath}_{\theta\theta}||\varphi'_\theta(\hat{\theta})|^{-2} \; , \quad |j_{(\lambda\lambda)}(\hat{\theta}_\psi)| = |j_{\lambda\lambda}(\hat{\theta}_\psi)||\varphi'_\lambda(\hat{\theta}_\psi)|^{-2}$$

are information determinants recalibrated to the φ parametrization; the matrix $\varphi_{\lambda'}(\hat{\theta}_\psi)$ is $p \times (p-1)$ and $|\varphi_{\lambda'}(\hat{\theta}_\psi)| = |\varphi'_{\lambda'}(\hat{\theta}_\psi)\varphi_{\lambda'}(\hat{\theta}_\psi)|^{1/2}$. Q is standard normal to the first order. The combination (2.1) or (2.2) of the present R and Q is however third order accurate as described earlier.

3. THIRD ORDER APPROXIMATION FOR THE NONCENTRAL CHI-SQUARED

Consider a generalization of Fisher's (1957) normal distribution on the circle. Let $y = \rho\alpha + e$ where ρ is positive, α is a p-dimensional unit vector, and $e_1, \ldots, e_p$ is a sample from the normal distribution with mean 0 and known variance σ_0^2. The parameter is $\theta = (\rho, \alpha)$ with ρ taken as the parameter of interest.

The distribution of $r^2/\sigma_0^2 = \sum y_i^2/\sigma_0^2$ is noncentral chi-squared with p degrees of freedom and noncentrality ρ^2/σ_0^2. As a consequence of scale properties, it suffices to carry out the calculations for the case $\sigma_0^2 = 1$.

The log likelihood function at the point y is

$$l(\theta) = l(\theta; y) = \rho \sum y_i \alpha_i - \rho^2 \sum \alpha_i^2/2 = \rho \sum y_i \alpha_i - \rho^2/2. \tag{3.1}$$

The full model maximum likelihood estimate is $\hat{\theta} = (\hat{\rho}, \hat{\alpha}) = (r, u)$ where u is the unit vector y/r. The maximum log likelihood is

$$l(\hat{\theta}) = r^2/2. \tag{3.2}$$

The constrained maximum likelihood estimate is $\hat{\theta}_\rho = (\rho, \hat{\alpha}_\rho) = (\rho, u)$ and the constrained maximum log likelihood is

$$l(\hat{\theta}_\rho) = \rho r - \rho^2/2. \tag{3.3}$$

We can now calculate the log likelihood ratio statistic

$$R^2 = 2[l(\hat{\theta}) - l(\hat{\theta}_\rho)] = 2(r^2/2 - \rho r + \rho^2/2) = (r - \rho)^2$$

and then the signed square root of the log likelihood statistic

$$R = sgn(r-\rho)[(r-\rho)^2]^{1/2} = r-\rho. \tag{3.4}$$

The observed information recalibrated on the φ scale is

$$|j_{\varphi\varphi}(\hat{\theta})| = 1. \tag{3.5}$$

For the nuisance parameter information, we rewrite the log likelihood function as

$$l(\theta) = \rho r \sum y_i \alpha_i / r - \rho^2/2 = \rho r \sum u_i \alpha_i - \rho^2/2$$

and initially examine the single component factor $\sum u_i \alpha_i = cos(u, \alpha)$ that depends on α. The second derivative determinant of the factor at $\alpha = \hat{\alpha}_\rho = u$ is -1, as calculated most easily using appropriately rotated α coordinates. It follows that the nuisance information determinant is $(\rho r)^{p-1}$ using α coordinates and is

$$j_{(\alpha\alpha)}(\hat{\theta}_\rho) = \left(\frac{r}{\rho}\right)^{p-1} \tag{3.6}$$

in the φ coordinates. The unit vector combination of the coordinates of φ that is used to measure departure from $\varphi(\hat{\theta}_\rho)$ is $\varphi(\theta) \cdot u$; the departure itself is $\varphi(\hat{\theta}) \cdot u - \varphi(\hat{\theta}_\rho) \cdot u = r - \rho$. The appropriate standardized version of this is

$$Q = (r-\rho)\left\{\frac{|j_{\theta\theta}(\hat{\theta})|}{|j_{(\alpha\alpha)}(\hat{\theta}_\rho)|}\right\}^{1/2} = (r-\rho)\left(\frac{\rho}{r}\right)^{(p-1)/2} . \tag{3.7}$$

The probability, $G_{\rho^2}(r^2)$, to the left of an observed r^2 for the noncentral chi-squared distribution with p degrees of freedom and noncentrality ρ^2 is given exactly by (1.1). This probability can approximated, with third order accuracy, by $1 - p(\rho)$ where $p(\rho)$ is defined in either (2.1) or (2.2) using R and Q from (3.4) and (3.7).

In many contexts, the Lugannani and Rice (2.1) formula seems to give excellent approximations although in extreme cases it can give values outside the $(0, 1)$ range for probabilities. The alternate way of combining R and Q using the Barndorff-Nielsen (2.2) formula is particularly attractive in its simplicity here: the distribution function for the noncentral chi-squared variable r^2 with noncentrality ρ^2 and p degrees of freedom can then be approximated by treating

$$z = (r - \rho) - \frac{p-1}{2} \frac{\log r - \log \rho}{r - \rho} \tag{3.8}$$

as a standard normal variable. Also, by taking normal percentiles for z in (3.8), we can solve for the corresponding percentiles of the noncentral chi-squared distribution; as the functions involved are simple, it is to be expected that the ordinary Newton method will work well for this. Also by differentiating the distribution function, say based on the standardized variable (3.8), we obtain an approximation of the noncentral chi-squared density function.

4. EXAMPLES

In all the examples, the exact probabilities are calculated from

$$\sum_{j=0}^{N} \frac{e^{-\rho^2/2}(\rho^2/2)^j}{j!} P(\chi^2_{p+2j} \leq r^2)$$

where N is an integer and

$$\frac{e^{-\rho^2/2}(\rho^2/2)^N}{N!} P(\chi^2_{p+2N} \leq r^2) < 10^{-10}.$$

Table 1 records selected values of the distribution function for the noncentral chi-squared distribution with p degrees of freedom and noncentrality ρ^2 using both the exact and the approximate formulas discussed above. Approximations (1.3) and (1.4) seem unsatisfactory; the Cox & Reid's approximation (1.2) is good when ρ is small. The proposed methods give good approximations, particularly (2.2) which involves evaluating the standard normal distribution function using (3.8).

Table 1a: Approximations to $G_{\rho^2}(r^2)$ with $p = 2$

ρ	1			2		
r	3	4	5	3	4	5
Cox & Reid (1.2)	0.9502	0.9952	0.9998	0.7768	0.9305	0.9845
Cox & Reid (1.3)	0.8946	0.9817	0.9981	0.0000	0.0000	0.0000
Bol'shev et al (1.4)	0.9831	0.9999	1.0000	1.0000	1.0000	1.0000
(2.1)	0.9575	0.9972	0.9999	0.7870	0.9661	0.9978
(2.2)	0.9578	0.9972	0.9999	0.7874	0.9661	0.9978
Exact (1.1)	0.9563	0.9971	0.9999	0.7856	0.9659	0.9978

ρ	5			10		
r	3	6	8	7	11	13
Cox & Reid (1.2)	0.2835	0.7364	0.9066	0.3815	0.6946	0.8093
Cox & Reid (1.3)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Bol'shev et al (1.4)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
(2.1)	0.0167	0.8182	0.9983	0.0011	0.8295	0.9984
(2.2)	0.0167	0.8183	0.9983	0.0011	0.8295	0.9984
Exact (1.1)	0.0166	0.8181	0.9983	0.0011	0.8295	0.9984

Table 1b: Approximations to $G_{\rho^2}(r^2)$ with $p = 5$

ρ	1			3		
r	2	4	5	2	5	6
Cox & Reid (1.2)	0.3513	0.9796	0.9992	0.0788	0.8881	0.9752
Cox & Reid (1.3)	0.3308	0.9747	0.9988	0.0000	0.0000	0.0000
Bol'shev et al (1.4)	0.3501	0.9834	0.9995	0.7781	1.0000	1.0000
(2.1)	0.1154	0.9765	0.9992	0.0242	0.9293	0.9942
(2.2)	0.3496	0.9810	0.9993	0.0351	0.9318	0.9944
Exact (1.1)	0.3472	0.9809	0.9993	0.0331	0.9313	0.9944

ρ	5			10		
r	3	6	8	8	11	13
Cox & Reid (1.2)	0.0869	0.6938	0.9416	0.3074	0.6699	0.8464
Cox & Reid (1.3)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Bol'shev et al (1.4)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
(2.1)	0.0055	0.7349	0.9963	0.0130	0.7905	0.9976
(2.2)	0.0060	0.7374	0.9964	0.0131	0.7909	0.9976
Exact (1.1)	0.0059	0.7368	0.9964	0.0131	0.7908	0.9976

Table 1c: Approximations to $G_{\rho^2}(r^2)$ with $p = 10$

ρ	3			5		
r	4	6	7	4	7	9
Cox & Reid (1.2)	0.4122	0.9591	0.9960	0.0821	0.8270	0.9899
Cox & Reid (1.3)	0.0014	0.0364	0.1022	0.0000	0.0000	0.0000
Bol'shev et al (1.4)	0.9192	1.0000	1.0000	1.0000	1.0000	1.0000
(2.1)	0.2003	0.9667	0.9985	0.0053	0.8815	0.9995
(2.2)	0.3842	0.9750	0.9988	0.0225	0.8931	0.9996
Exact (1.1)	0.3825	0.9750	0.9988	0.0221	0.8927	0.9996

ρ	10			20		
r	8	11	13	18	21	23
Cox & Reid (1.2)	0.1700	0.6425	0.8806	0.3616	0.6232	0.7708
Cox & Reid (1.3)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Bol'shev et al (1.4)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
(2.1)	0.0056	0.7118	0.9953	0.0126	0.7819	0.9974
(2.2)	0.0062	0.7160	0.9954	0.0126	0.7824	0.9974
Exact (1.1)	0.0062	0.7158	0.9954	0.0126	0.7824	0.9974

Figure 1 records the distribution function for the noncentral chi-squared distribution with 3 degrees of freedom and noncentrality parameter $\rho^2 = 2$ (i.e. it is a plot of $G_2(r^2)$ against r^2 with $p = 3$). Equation (2.2) gives approximations closest to the exact solution. Approximations obtained from equations (2.1) and (1.2) curves are also reasonable, but, the other two approximations are not satisfactory.

Figure 1: Noncentral chi-squared distribution function
(p=3, rho=sqrt(2))

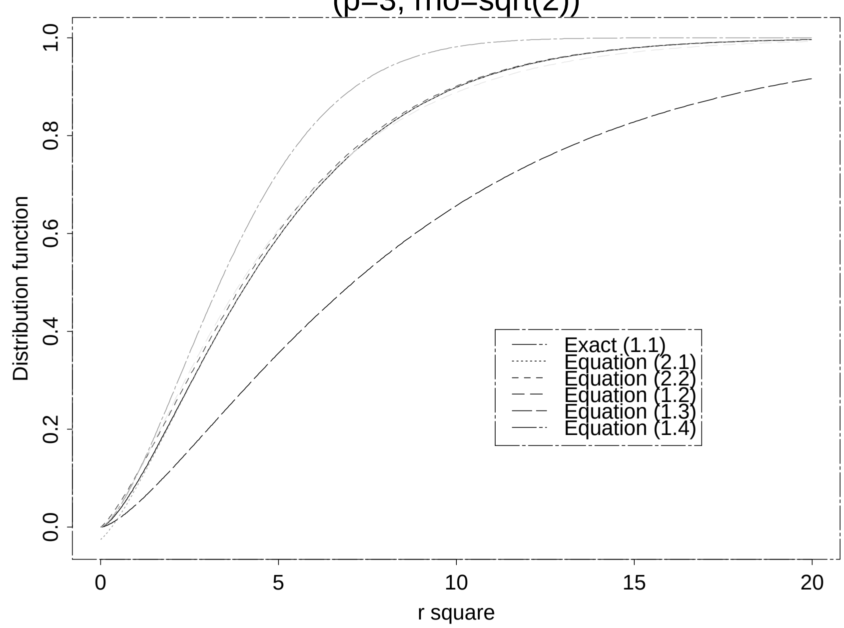
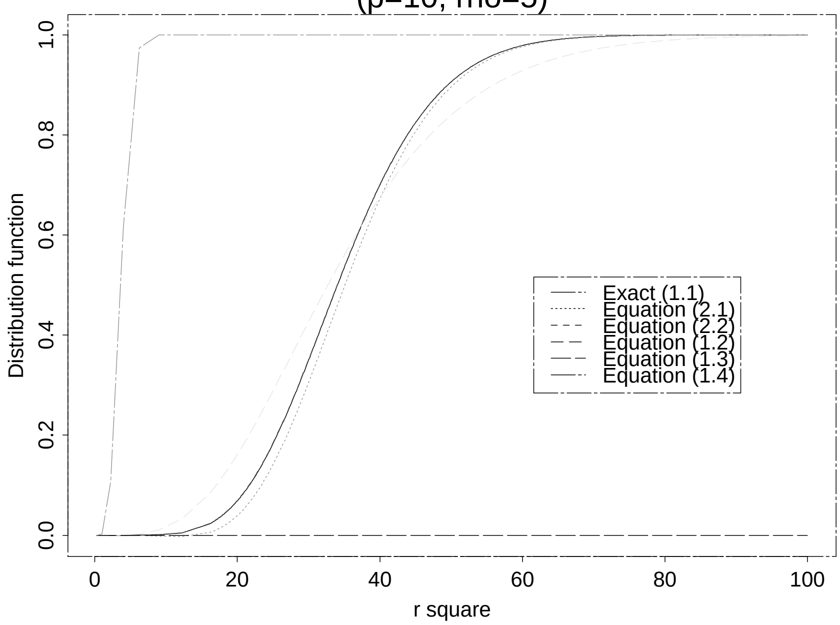


Figure 2 records the distribution function for the noncentral chi-squared distribution with 10 degrees of freedom and noncentrality parameter $\rho^2 = 25$ (i.e. it is a plot of $G_{25}(r^2)$ against r^2 with $p = 10$). Equation (2.2) and the exact solution are indistinguishable from each other. The other approximations are not satisfactory.

Figure 2: Noncentral chi-squared distribution function
(p=10, rho=5)



5. DISCUSSION

In this paper, we develop a simple third order approximation for the distribution of a noncentral chi-square variable with p degrees of freedom and noncentrality ρ^2. It is given by $\Phi\left\{r - \rho - (p-1)(\log r - \log \rho)/2(r-\rho)\right\}$ where $\Phi(\cdot)$ is the standard normal distribution function. It provides high accuracy. Furthermore, it is easily inverted to give the percentiles of the distribution.

REFERENCES

Anderson, T.W. (1975). *An Introduction to Multivariate Statistical Analysis.* New York: Wiley.

Barndorff-Nielsen O.E. (1986). Inference on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika* **73**, 307-322.

Barndorff-Nielsen O.E. (1991). Modified signed log likelihood ratio. *Biometrika* **78**, 557-564.

Barndorff-Nielsen O.E., and Cox, D.R. (1979). Edgeworth and saddlepoint approximations with statistical applications (with discussions). *Journal of the Royal Statistical Society B* **41**, 279-312.

Bol'shev, L.N., and Kuznetzov, P.I. (1963). On the calculation of the integral $p(x, y) = \cdots$ (English transl.). *U.S.S.R. Comput. Math. and Math. Phys.* **3**, 557-571.

Cohen, J.D. (1988). Noncentral chi-square: some observations on recurrence. *The American Statistician* **42**, 120-122.

Cox, D.R., and Reid, N. (1987). Approximations to noncentral distributions. *Canadian Journal of Statistics* **15**, 105-114.

Daniels, H.E. (1954). Saddlepoint approximations in statistics. *Annals of Mathematical Statistics* **25**, 631-650.

DiCiccio, T.J., Field, C.A., and Fraser, D.A.S. (1990). Approximation of marginal tail probabilities and inference for scalar parameters. *Biometrika* **77**, 77-95.

Fisher, R.A. (1957). *Statistical Methods and Scientific Inference.* Edinburgh: Oliver & Boyd.

Fraser, D.A.S. (1990). Tail probabilities from observed likelihood. *Biometrika* **77**, 65-76.

Fraser, D.A.S. (1991). Statistical inference: likelihood to significance. *Journal of the American Statistical Association* **86**, 258-265.

Fraser, D.A.S., and Reid, N. (1995). Ancillaries and third order significance. *Utilitas Mathematica* **47**, 33-53.

Johnson, N.L., and Kotz, S. (1970). *Distributions in Statistics: Continuous Univariate Distributions 2.* New York: Wiley.

Lugannani, R., and Rice, S. (1980). Saddlepoint approximation for the distribution of the sum of independent random variables. *Advanced Applied Probability* **12**, 475-490.

Pierce, D.A., and Peters, D. (1992). Practical use of higher order asymptotics for multiparameter exponential families (with discussion). *Journal of the Royal Statistical Society B* **54**, 701-738.

Patnaik, P.B. (1949). The noncentral χ^2 and F distributions and their applications. *Biometrika* **36**, 202-232.

Posten, H.O. (1989). An effective algorithm for the noncentral chi-squared distribution function. *The American Statistician* **43**, 261-263.

Wang, S., and Gray, H.L. (1993). Approximating tail probabilities of noncentral distribution. *Computational Statistics and Data Analysis* **15**, 343-352.