Calculating the Density and Distribution Function for the Singly and Doubly Noncentral F

Ronald W. Butler [a] * Marc S. Paolella [b] †

[a] *Department of Statistics, Colorado State University, Fort Collins, CO 80525, USA*

[b] *Institute of Statistics and Econometrics, Kiel University, D–24098, Kiel, Germany*

April 29, 1999

Preliminary Version

Abstract

We derive a saddlepoint approximation to the pdf and cdf of both a singly and doubly noncentral F random variable and demonstrate its high accuracy over the entire parameter space. In contrast to usual saddlepoint applications, the method admits a *closed form* solution implying that, particularly for the doubly noncentral case, the proposed method is several orders of magnitude faster than existing computational methods. We also draw attention to existing errors in popular software packages, prove uniformity of error of the approximation, and provide an example demonstrating use of the new method for sample size calculation in experimental design.

Keywords—Design of Experiments; Ratios of Quadratic Forms in Normal Variables; Saddlepoint Approximation; Statistical Computing.

*Research supported in part by NSF grant DMS-9625396.

†Corresponding author. e-mail: paolella@stat-econ.uni-kiel.de

1 Introduction

A large number of test statistics can be expressed as ratios of quadratic forms in normal variables, say

$$G = \frac{\mathbf{Z}'\mathbf{A}\mathbf{Z}}{\mathbf{Z}'\mathbf{B}\mathbf{Z}}, \quad \mathbf{Z} \sim \mathrm{N}_n(\boldsymbol{\mu}, \Sigma). \tag{1}$$

Despite such prevalence, tractable expressions that are useful for efficient computation of their probability densities (pdfs) and cumulative distributions (cdfs) functions are not available (see, for example, Provost and Rudiuk, 1996).

Arguably the most important example of (1) is the noncentral F distribution. It plays an ubiquitous role in the analysis of linear models where it is used in the calculation of power for tests of linear hypotheses. In the usual case, it is the singly noncentral F, say $F^{(1)}$, which is relevant. However, numerous examples exist where the doubly noncentral case, $F^{(2)}$, is required, for example, in power calculations for ANOVA designs with interaction or bias effects (Tang, 1938; Scheffé, 1959, pp. 134-5, 415; Bulgren, 1971; and the references therein) and two-way cross classification ANOVA (Tiko, 1972). In econometrics, it arises naturally in testing linear models with proxy variables (Kurumai and Ohtani, 1998) and with Ramsey's (1969) popular regression specification error (RESET) test (DeBenedictis and Giles, 1998). Distribution $F^{(2)}$ also occurs in signal processing and pattern recognition applications (Price, 1964; Helstrom and Ritcey, 1985).

A method for evaluating the cdf of (1) that involves numerical integration was provided by Imhof (1961). This method, however, has some drawbacks: it can be quite slow when n, the dimension of $\mathbf{Z}$, is large, and can occasionally fail either when n is small, and/or far in the tails, as discussed below.

In this paper, we develop saddlepoint approximations to the pdf and cdf of both noncentral F distributions $F^{(1)}$ and $F^{(2)}$. We show that the pdf and cdf approximations preserve relative error over the entire support of the distribution and are extremely accurate throughout their range of support. Furthermore, these methods yield *closed–form* expressions for saddlepoints which result in computations that are trivial to implement in practice. Such explicitness also results in computational speeds that are several thousand times faster than those of the Imhof method.

The paper is organized as follows. Section 2 develops saddlepoint expressions for the cdf and pdf of (1) and specializes these results for the singly and doubly noncentral F distributions. Section 3 briefly reviews computationally exact evaluation methods and provides some explanations as to why the numeric implementations in three popular software packages occasionally fail. Correct exact methods are used in Section 4 to demonstrate the accuracy of the SPA. Section 5 illustrates a practical application of the proposed methods in sample size determination and 6 provides concluding remarks.

2 Saddlepoint Approximation

We present general results on saddlepoint approximations followed by their application to ratio G and then more specifically to the noncentral F distribution. For clarity, technical details and general discussions of the relative orders of accuracy are avoided.

2.1 Saddlepoint Distribution Approximations

A saddlepoint approximation (SPA) to the cdf of random variable X with cumulant generating function $K(s)$ that converges on $s \in (a,b)$ has been derived in Lugannani and Rice (1980), as

$$\tilde{F}(x) = \begin{cases} \Phi(\hat{w}) + \phi(\hat{w})\left\{\hat{w}^{-1} - \hat{u}^{-1}\right\} & \text{if} \quad x \neq \mathcal{E}[X] \\ \frac{1}{2} + \frac{K'''(0)}{6\sqrt{2\pi}K''(0)^{3/2}} & \text{if} \quad x = \mathcal{E}[X] \end{cases} \tag{2}$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ denote the distribution and density function of a standard normal random variable, respectively and

$$\hat{w} = \operatorname{sgn}(\hat{s})\sqrt{2\{\hat{s}x - K(\hat{s})\}}, \quad \text{and} \quad \hat{u} = \hat{s}\sqrt{K''(\hat{s})}. \tag{3}$$

Value $\hat{s}$ is the saddlepoint that is defined as the unique solution to the saddlepoint equation

$$K'(\hat{s}) = x \tag{4}$$

for $\hat{s}$ in (a,b). (see also Barndorff-Nielsen and Cox, 1990; Field and Ronchetti, 1990; Jensen, 1994; and Reid, 1988, 1996). Expression (2) is the leading term in an asymptotic expansion and the next most important term has been provided by Daniels (1987) to give a second order approximation as

$$\hat{F}(x) = \tilde{F}(x) - \phi(\hat{w})\left\{\hat{u}^{-1}\left(\frac{\hat{\kappa}_4}{8} - \frac{5}{24}\hat{\kappa}_3^2\right) - \hat{u}^{-3} - \frac{\hat{\kappa}_3}{2\hat{u}^2} + \hat{w}^{-3}\right\} \tag{5}$$

for $x \neq \mathcal{E}[X]$ where $\hat{\kappa}_i = K^{(i)}(\hat{s})/K''(\hat{s})^{i/2}$.

2.1.1 Application to G

To evaluate the cdf of G, rewrite it as

$$\Pr(G < g) = \Pr\left(\mathbf{Z}'(\mathbf{A} - g\mathbf{B})\mathbf{Z} < 0\right) = \Pr\left(\mathbf{X}'\Lambda\mathbf{X} < 0\right) = \Pr(X < 0)$$

where $\mathbf{X} = \mathbf{P}'\Sigma^{-1/2}\mathbf{Z} \sim \mathrm{N}_n(\boldsymbol{\nu}, \mathbf{I}_n)$, $\Sigma^{1/2}$ is the unique positive definite square root of Σ, and $\mathbf{P}\Lambda\mathbf{P}'$ is the spectral decomposition of $\Sigma^{1/2}(\mathbf{A} - g\mathbf{B})\Sigma^{1/2}$ such that $\mathbf{P}$ is orthogonal and $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. Random variable $X = \mathbf{X}'\Lambda\mathbf{X}$ has the distribution of a linear combination of independent noncentral

χ^2_1 random variables of the form $\sum_{i=1}^n \lambda_i \chi^2_i\left(1,\nu_i^2\right)$. Here, $\chi^2_i\left(1,\nu_i^2\right)$ denotes a noncentral χ^2_1 variable with one degree of freedom and noncentrality parameter ν_i^2, and $\{\lambda_i : i = 1,\ldots,n\}$ are the eigenvalues of $\Sigma\left(\mathbf{A}-g\mathbf{B}\right)$.

The moment generating function (mgf) of X is given by

$$M_X(s) = \left\{\prod_{i=1}^n \left(1-2s\lambda_i\right)\right\}^{-1/2} \exp\left(\sum_{i=1}^n \frac{s\,\lambda_i\nu_i^2}{1-2s\lambda_i}\right) \tag{6}$$

when $1-2s\lambda_i > 0$ for all i. The mgf converges over the open neighborhood of zero

$$\frac{1}{2\lambda_{(1)}} < s < \frac{1}{2\lambda_{(n)}} \tag{7}$$

when $\lambda_{(1)} := \min_i \lambda_i < 0 < \lambda_{(n)} := \max_i \lambda_i$, which is the setting for the noncentral F. In the other cases involving $\lambda_{(1)}$ and $\lambda_{(n)}$, the convergence strip is unbounded on one side. The cgf and its derivatives are easily obtained from (6). The bounded solution to the saddlepoint equation $K'\left(\hat{s}\right) = 0$ within (7) is found using standard numerical univariate root search algorithms.

2.1.2 Application to the Noncentral F Distribution

There is considerable simplification to the above in the context of the noncentral F distribution. Let $U_i = \mathbf{Y}_i'\mathbf{Y}_i$, where $\mathbf{Y}_i \sim \mathrm{N}_{n_i}\left(\boldsymbol{\mu}_i, \mathbf{I}_{n_i}\right)$ for $i = 1,2$ so that $U_i \sim \chi^2\left(n_i,\theta_i\right)$, where n_i and $\theta_i = \boldsymbol{\mu}_i'\boldsymbol{\mu}_i$ denote the degrees of freedom and noncentrality parameter, respectively. If U_1 and U_2 are independent, then F follows the doubly noncentral $F_{n_1,n_2}\left(\theta_1,\theta_2\right)$ distribution, where

$$F = \frac{U_1/n_1}{U_2/n_2} = \frac{n_2}{n_1}\frac{\mathbf{Z}'\begin{pmatrix}\mathbf{I}_{n_1} & 0\\ 0 & 0\end{pmatrix}\mathbf{Z}}{\mathbf{Z}'\begin{pmatrix}0 & 0\\ 0 & \mathbf{I}_{n_2}\end{pmatrix}\mathbf{Z}} := \frac{n_2}{n_1}\frac{\mathbf{Z}'\mathbf{A}\mathbf{Z}}{\mathbf{Z}'\mathbf{B}\mathbf{Z}} \tag{8}$$

and $\mathbf{Z} = \begin{bmatrix}\mathbf{Y}_1\\ \mathbf{Y}_2\end{bmatrix} \sim \mathrm{N}_{n_1+n_2}\left(\begin{bmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{bmatrix}, \mathbf{I}_{n_1+n_2}\right) := \mathrm{N}_n\left(\boldsymbol{\nu}, \mathbf{I}_n\right)$, i.e., $\mathbf{A}$ and $\mathbf{B}$ are defined in (8), $n = n_1+n_2$ and $\boldsymbol{\nu} = \begin{bmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{bmatrix}$. Taking $\theta_2 = 0$ yields the singly noncentral F, while $\theta_1 = \theta_2 = 0$ is the central F distribution.

The cdf of F at f can be expressed as

$$\begin{aligned}\Pr\left(F < f\right) &= \Pr\left(\frac{n_2}{n_1}\mathbf{Z}'\mathbf{A}\mathbf{Z} < f\mathbf{Z}'\mathbf{B}\mathbf{Z}\right) = \Pr\left\{\mathbf{Z}'\left(\frac{n_2}{n_1}\mathbf{A} - f\mathbf{B}\right)\mathbf{Z} < 0\right\}\\ &= \Pr\left\{\sum_{i=1}^n \lambda_i\chi_i^2\left(1,\nu_i^2\right) < 0\right\} = \Pr\left(X_f < 0\right)\end{aligned} \tag{9}$$

where

$$\lambda_i = \begin{cases} \ell_1 = n_2/n_1, & \text{if } 1 \le i \le n_1,\\ \ell_2 = -f, & \text{if } n_1 < i \le n.\end{cases}$$

The mgf of random variable X_f in (9) simplifies to

$$M_{X_f}(s) = \left(1 - 2s\frac{n_2}{n_1}\right)^{-\frac{n_1}{2}} (1+2sf)^{-\frac{n_2}{2}} \exp\left\{\frac{s\theta_1\frac{n_2}{n_1}}{1-2s\frac{n_2}{n_1}} - \frac{s\theta_2 f}{1+2sf}\right\} \tag{10}$$

and is convergent on the neighborhood of zero given by

$$-\frac{1}{2f} < s < \frac{n_1}{2n_2}. \tag{11}$$

This convergence region does not depend upon the noncentrality parameters. By defining $\vartheta_i = 1/(1-2s\ell_i)$ for $i = 1, 2$, the d^{th} order derivative of $K_{X_f}(s)$ is given by

$$K^{(d)}_{X_f}(s) = k_d \sum_{i=1}^{2} \ell_i^d \vartheta_i^d \left(n_i + d\theta_i \vartheta_i\right), \quad d \geq 1, \tag{12}$$

where $k_1 = 1$ and $k_d = 2(d-1)k_{d-1}$ for $d > 1$.

In most saddlepoint applications, $\hat{s}$ in (4) needs to be found numerically. However, in this case, an explicit root can be found for $K'_{X_f}(\hat{s}) = 0$ where

$$K'_{X_f}(s) = \frac{n_2}{n_1}\frac{1}{1-2s\frac{n_2}{n_1}}\left(n_1 + \frac{\theta_1}{1-2s\frac{n_2}{n_1}}\right) - \frac{f}{1+2sf}\left(n_2 + \frac{\theta_2}{1+2sf}\right). \tag{13}$$

Using the notation of Abramowitz and Stegun (1972, p. 17) equation (13) can be expressed as the root of the cubic $s^3 + a_2 s^2 + a_1 s + a_0$ where

$$\begin{aligned}
a \cdot a_2 &= 8f(1-f)n_1 n_2^2 + 4f\left(n_2^3 + \theta_2 n_2^2 - n_1^2 n_2 f - n_1 n_2 \theta_1 f\right), \\
a \cdot a_1 &= 2\left(n_2^2 n_1 + n_1^2 n_2 f^2\right) - 4f n_1 n_2 \left(n_1 + n_2 + \theta_1 + \theta_2\right), \\
a \cdot a_0 &= f\theta_2 n_1^2 - (1-f) n_1^2 n_2 - n_1 n_2 \theta_1 \\
a &= 8f^2 n_2^2 (n_1 + n_2).
\end{aligned}$$

Upon further defining

$$q = \frac{1}{3}a_1 - \frac{1}{9}a_2^2, \quad r = \frac{1}{6}(a_1 a_2 - 3a_0) - \frac{1}{27}a_2^3, \quad m = q^3 + r^2 \tag{14}$$

and the two values $s_{1,2} = \sqrt[3]{r \pm m^{1/2}}$, the three roots to the cubic equation are given as in (40) of the Appendix. The following result is also shown in the Appendix:

Theorem 1 *For any $n_1, n_2 > 0$ and any values of $, \theta_1, \theta_2 \geq 0$, the three roots to the saddlepoint equation given in (40) are real and ordered according to $z_2 < z_3 \leq z_1$. Among them, the saddlepoint solution is always*

$$\hat{s} = z_3 = -\frac{1}{2}(s_1 + s_2) - \frac{a_2}{3} - \frac{i\sqrt{3}}{2}(s_1 - s_2).$$

Root z_3 is the only root in the range $-1/(2f) < s < n_1/(2n_2)$ and the only root for which $K''(z_i) > 0$.

An alternative expression for the saddlepoint that is useful in an environment that does not support complex arithmetic is

$$\hat{s} = z_3 = \sqrt{-q}\left\{\sqrt{3}\sin(\phi) - \cos(\phi)\right\} - \frac{a_2}{3} \tag{15}$$

where

$$\phi = \frac{1}{3}\arg\left(r + i\sqrt{-m}\right) = \begin{cases} \tan^{-1}(-m/r) & \text{if} \quad r \geq 0 \\ \pi + \tan^{-1}(-m/r) & \text{if} \quad r < 0 \end{cases},$$

and m and q are always < 0. These results are derived in the Appendix.

When $\theta_2 = 0$, we obtain the single root

$$\hat{s} = \frac{f n_1 (n_1 + 2n_2 + \theta_1) - n_1 n_2 - \sqrt{n_1 y}}{4 n_2 f (n_1 + n_2)}, \tag{16}$$

where

$$y = f^2 n_1^3 + 2f^2 n_1^2 \theta_1 + 2n_1^2 f n_2 + 4f^2 n_1 n_2 \theta_1 + n_1 \theta_1^2 f^2 + 2n_1 \theta_1 f n_2 + n_2^2 n_1 + 4f n_2^2 \theta_1.$$

For the central F case with $\theta_1 = 0 = \theta_2$, this simplifies further to

$$\hat{s} = \frac{n_1 (f - 1)}{2f (n_1 + n_2)}. \tag{17}$$

The benefit of explicit saddlepoints is that the saddlepoint cdf and pdf have closed form expressions (except for evaluation of the normal cdf Φ) and are trivial to implement in all cases.

With fractional degrees of freedom the mgf development is still correct and the saddlepoint approximations remain valid in approximating these generalized distributions. By comparson, the Imhof algorithm, that has often been used in computing the noncentral F, only supports integer degrees of freedom in its computations. Apart from simulation, there appears to be no obvious alternative means of computation in this case.

The saddlepoint $\hat{s} = 0$ when $f = (1 + \theta_1/n_1) / (1 + \theta_2/n_2)$ for all parameter values. This occurs at the mean of X_f but not the mean of F. For this value of f, the limiting approximation in (2) should be used. In the singly noncentral case, $\hat{s} = 0$ for $f = 1 + \theta_1/n_1$, which can be compared to the mean $\mathcal{E}\left(F^{(1)}\right) = (1 + \theta_1/n_1)\left(n_2/(n_2 - 2)\right)$.

2.2 Saddlepoint Density Approximations

A density saddlepoint approximation (SPA) of random variable X at x was given in Daniels (1954, Sec. 2) as

$$\tilde{g}(x) = \frac{1}{\sqrt{2\pi K''(\hat{s})}} \exp\left\{K(\hat{s}) - \hat{s}x\right\}. \tag{18}$$

where $\hat{s}$ is the same saddlepoint that solves (4) for the cdf approximation. A second-order approximation has been provided in Daniels (1954, 1987) as

$$\hat{g}(x) = \tilde{g}(x)\left(1 + \frac{\hat{\kappa}_4}{8} - \frac{5}{24}\hat{\kappa}_3^2\right). \tag{19}$$

The density of ratio $F = U/V$ can be expressed as the density of a "constructed" random variable to which the approximation in (19) may be applied. This is the approach used in Daniels (1954, Sec. 9) and later in Lieberman (1994). The connection makes use of the Geary (1944) representation for the density of a ratio of random variables. Geary showed that the density of F at f, or $g_F(f)$ for fixed value f, can be expressed in terms of the density of "constructed" random variable W_f at 0, or $g_{W_f}(0)$, where W_f is the random variable associated with mgf

$$M_{W_f}(s) = \frac{1}{\mathcal{E}(V)}\frac{\partial}{\partial t} M_{U,V}(s,t)|_{t=-sf}. \tag{20}$$

The relationship is

$$g_F(f) = \mathcal{E}(V)\, g_{W_f}(0) \tag{21}$$

and is developed in Stuart and Ord (1994, Sec. 11.10).

Using the SPA (19) via (20) for $g_{W_f}(0)$, relationship (21) yields the density SPA to $g_F(f)$ as

$$\hat{g}_F(f) = \mathcal{E}(V)\, \hat{g}_{W_f}(0) \tag{22}$$

which is Daniels' (1954) approximation.

2.2.1 Application to the Noncentral F Distribution

Take $U = \frac{n_2}{n_1}\mathbf{Z}'\mathbf{A}\mathbf{Z}$ and $V = \mathbf{Z}'\mathbf{B}\mathbf{Z}$ so that $F = U/V$. Then $\mathcal{E}(V) = n_2 + \theta_2$ and the joint mgf of (U,V) is

$$M_{U,V}(s,t) = \left(1 - 2s\frac{n_2}{n_1}\right)^{-\frac{n_1}{2}} (1-2t)^{-\frac{n_2}{2}} \exp\left(\frac{s\theta_1\frac{n_2}{n_1}}{1-2s\frac{n_2}{n_1}} + \frac{t\theta_2}{1-2t}\right).$$

The computation in (20) leads to

$$\begin{aligned} M_{W_f}(s) &= \left(1 - 2s\frac{n_2}{n_1}\right)^{-\frac{n_1}{2}} (1+2sf)^{-\frac{n_2}{2}} \exp\left(\frac{s\theta_1\frac{n_2}{n_1}}{1-2s\frac{n_2}{n_1}} - \frac{sf\theta_2}{1+2sf}\right) && (23) \\ &\quad \times (n_2+\theta_2)^{-1}\left\{\frac{\theta_2}{(1+2sf)^2} + \frac{n_2}{1+2sf}\right\}. && (24) \end{aligned}$$

If the portion of M_{W_f} in (24) is discarded in saddlepoint determination, so only (23) is used, the saddlepoint equation becomes (13) which makes the density saddlepoint the same as the cdf saddlepoint

in (15). In this case the saddlepoint density is

$$\tilde{g}_F(f) = \frac{(n_2+\theta_2)\, M_{W_f}(\hat{s})}{\sqrt{2\pi\left(\ln M^{\circ}_{W_f}\right)''(\hat{s})}} \tag{25}$$

where $M^{\circ}_{W_f}$ is $M_{W_f}(s)$ using only the terms in the first line (23).

Further justification for excluding the terms in row (24) when determining $\hat{s}$ is obtained by examining the exactness of the approximation in the central setting. Using the saddlepoint in (17) with $\theta_1 = 0 = \theta_2$ then (25) reduces to

$$\tilde{g}_F(f) = \left(\frac{n_2}{n_1}\right)^{\frac{n_2}{2}} \frac{1}{\hat{B}\left(\frac{n_1}{2}, \frac{n_2}{2}\right)} \frac{f^{\frac{n_1}{2}-1}}{\left(1+\frac{n_1}{n_2}f\right)^{(n_1+n_2)/2}}, \tag{26}$$

where $\hat{B}\left(\frac{n_1}{2}, \frac{n_2}{2}\right)$ is Stirling's approximation to the beta function $B\left(\frac{n_1}{2}, \frac{n_2}{2}\right)$. This is the exact F_{n_1,n_2} density apart from the use of Stirling's approximation; therefore $\tilde{g}_F(f)$ becomes exact upon normalization.

2.2.2 Moments

Given the inherent speed of the saddlepoint approximation, the calculation of moments of G becomes feasible. In particular, one may compute $\mathcal{E}\left[G^k\right] \approx \int_{-\infty}^{\infty} g^k \hat{f}(g)\, dg$, as pointed out by Lieberman (1994, p. 925). An alternative approach is to base the computation on the cdf approximation and use the fact that $\mathcal{E}\left[(G-\mu)^k\right]$ may be expressed as

$$k\int_{\mu}^{\infty} (g-\mu)^{k-1}\left(1-\hat{F}(g)\right) dg - k\int_{-\infty}^{\mu} (g-\mu)^{k-1}\hat{F}(g)\, dg,$$

for $\mu = E(G)$. The latter expression has two advantages. First, it can be expected to provide greater accuracy than $\tilde{f}(g)$ and perhaps $\hat{f}(g)$ when these saddlepoint densities are used without normalization. Upon numerical normalization, the density approximations then require twice the numerical effort and so it is more efficient. Secondly, $\hat{F}$ can be replaced with the Imhof approximation so the accuracy can be checked without resorting to simulation.

As an example, Morin-Wahhab (1985) gave expressions for the positive integer moments of

$$\frac{\sum_{i=1}^{p_1} a_i X_i + \sum_{j=1}^{p_3} c_j Z_j}{\sum_{i=1}^{p_2} b_i Y_i + \sum_{j=1}^{p_3} d_j Z_j}$$

where X_i, $i = 1, \ldots, p_1$, Y_j, $j = 1, \ldots, p_2$ and Z_k, $k = 1, \ldots, p_3$ are independent χ^2 random variables with ℓ_i, m_j and n_k degrees of freedom, respectively. This is clearly a special case of G in (1) with

$$\begin{aligned} \mathbf{A} &= \text{blockdiag}\left(a_1\mathbf{I}_{\ell_1}, \ldots, a_{p_1}\mathbf{I}_{\ell_{p_1}}, \mathbf{0}_{m_\bullet}, c_1\mathbf{I}_{n_1}, \ldots, c_{p_3}\mathbf{I}_{n_{p_3}}\right) \\ \mathbf{B} &= \text{blockdiag}\left(\mathbf{0}_{\ell_\bullet}, b_1\mathbf{I}_{m_1}, \ldots, b_{p_2}\mathbf{I}_{m_{p_2}}, d_1\mathbf{I}_{n_1}, \ldots, d_{p_3}\mathbf{I}_{n_{p_3}}\right) \end{aligned}$$

where $\mathbf{I}_h$ denotes the $h \times h$ identity matrix, $\mathbf{0}_h$ is an $h \times h$ matrix of zeros, $\ell_\bullet = \sum_{i=1}^{p_1} \ell_i$, $m_\bullet = \sum_{j=1}^{p_2} m_j$, $n_\bullet = \sum_{k=1}^{p_3} n_k$, $\mathbf{Z} \sim \mathrm{N}_n(\mathbf{0}, \mathbf{I})$ and $n = \ell_\bullet + m_\bullet + n_\bullet$. Because her expressions also involve numeric approximation (of hypergeometric functions of many variables), the saddlepoint method not only serves as a highly accurate (and most likely much faster) approximation from which "exact" calculations might be verified, but could turn out to be more accurate in certain cases as well. The SPA is also valid for any real moment (provided it exists), and not just the positive integers.

2.3 Relative Errors in Approximation

Relative errors for first and second order SPAs of the cdf and density are uniform for f in the right tail. This is a consequence of determining the asymptotic relative error of the SPAs as $f \to \infty$ as derived in the Appendix. The method of proof is somewhat nonstandard but broadly follows the general approach used in Jensen (1988).

We have not addressed the left tail of F as $f \to 0$. The left tail of F, however, is the right tail of $1/F$ that has a different noncentral F distribution for which the relative error is uniform.

Expansion of the saddlepoint equation $K'_{X_f}(\hat{s}) = 0$ determined from (13), shows that

$$\hat{s} = \omega_0 + O\left(f^{-1}\right)$$

as $f \to \infty$ where

$$0 < \omega_0 = \frac{n_1}{n_1 + \theta_1 + 2n_2 + \sqrt{(n_1 + \theta_1)^2 + 4\theta_1 n_2}} < \frac{n_1}{2n_1}.$$

The saddlepoint $\hat{s}$ is monotonic increasing in f and has upper bound ω_0 so the saddlepoint root never reaches $n_1/(2n_1)$, the endpoint of the convergence strip for X_f. The term

$$\hat{w} = \sqrt{-2K_{X_f}(\hat{s})} \to \infty$$

as $f \to \infty$, not because $\hat{s}$ reaches the endpoint of convergence, but because the mgf term $(1 + 2\hat{s}f)^{-n_2/2} \to 0$. The following result is derived in the Appendix.

Theorem 2 *The limiting relative error of the first order SPA in (2) is*

$$\lim_{f \to \infty} \frac{\Pr(X_f > 0)}{\widehat{\Pr}(X_f > 0)} = \sqrt{2\pi} g_{Y_1 - Y_2}(0), \tag{27}$$

where $g_{Y_1 - Y_2}(0)$ denotes the density of $Y_1 - Y_2$ at 0 where $Y_1 \sim \nu_0 \chi^2(n_1, 2\nu_1)$ independently of $Y_2 \sim (2\omega_0\sigma_0)^{-1} \chi^2_{n_2+2}$. Parameters ν_0, ν_1, and σ_0 are given in (46), (48), and (47) of the Appendix and $g_{Y_1 - Y_2}(0)$ is given in (49). The first order SPA density has the same limiting relative error

$$\lim_{f \to \infty} \frac{g_F(f)}{\tilde{g}_F(f)} = \sqrt{2\pi} g_{Y_1 - Y_2}(0). \tag{28}$$

In the central case with $\theta_1 = 0 = \theta_2$, this works out to be $\hat{B}\left(n_1/2, n_2/2\right)/B\left(n_1/2, n_2/2\right)$ which agrees with the error indicated in (26).

Relative errors for the second order approximations, given in (5) for the cdf and (19) for the density, are also uniform in the right tail. The additional correction terms for second order remain bounded as $\hat{s} \to \omega_0$. This occurs because they are either standard cumulants at $\hat{s}_0$, $\hat{u}$ or $\hat{w}^{-3}$ and all these converge as $\hat{s} \to \omega_0$.

3 Exact Evaluation

We briefly discuss numerical methods of evaluation whose accuracy level can be explicitly controlled. When calculated to machine precision, we deem such methods "exact".

3.1 Singly Noncentral F

The density and cdf of $F^{(1)} \sim F_{n_1,n_2}\left(\theta_1, 0\right)$ at f can be expressed as (Johnson *et al.*, 1995, eq. 30.7 and 30.10)

$$p_{F^{(1)}}\left(f\right) = \frac{n_1}{n_2} \sum_{j=0}^{\infty} \frac{\omega_{j,\theta_1}}{B\left(\frac{n_1}{2}+j, \frac{n_2}{2}\right)} \frac{\left(\frac{n_1}{n_2} f\right)^{n_1/2+j-1}}{\left(1+\frac{n_1}{n_2} f\right)^{(n_1+n_2)/2+j}} := \frac{n_1}{n_2} \sum_{j=0}^{\infty} d_j \tag{29}$$

and

$$\Pr\left(F^{(1)} < f\right) = \sum_{j=0}^{\infty} \omega_{j,\theta_1} I_x\left(\frac{n_1}{2}+j, \frac{n_2}{2}\right) := \sum_{j=0}^{\infty} c_j, \tag{30}$$

respectively, where

$$\omega_{i,\theta} = e^{-(\theta/2)} \frac{(\theta/2)^i}{i!}, \tag{31}$$

$x = n_1 f / \left(n_2 + n_1 f\right)$ and $I_x\left(a, b\right)$ is the incomplete beta function ratio $\int_0^x t^{a-1}\left(1-t\right)^{b-1} dt \,/\, B\left(a, b\right)$.

Consider first the evaluation of (30). Schader and Schmid (1986), Lenth (1987) and Randall (1994) made use of the recursions

$$\begin{aligned} I_x\left(a+1, b\right) =& I_x\left(a, b\right) + \frac{, \ (a+b)}{, \ (a+1) \, , \ (b)} x^a \left(1-x\right)^b \\ :=& I_x\left(a, b\right) + T_x\left(a, b\right) \end{aligned}$$

and

$$T_x\left(a+1, b\right) = T_x\left(a, b\right) \frac{a+b}{a+1} x$$

to realize considerable speed improvements. Also, the summands d_j and c_j are not always monotonically decreasing, so that the finite upper limit on j, say $j_{\max}$, cannot simply be determined as when the summand is small relative to the cumulative sum.[1] This is illustrated in Figure 1, which plots c_j with both $\log_{10}$ and regular scales, for values

$$n_1 = 1, \quad n_2 = 12, \quad \theta_1 = 2316 \text{ and } f = 990, \tag{32}$$

which arose in an application by Chow and Shao (1990). (The exact cdf value was computed as 0.0057818, the SPA returns 0.0057812.) Note that c_j reaches its maximum at $1149 \approx \theta_1/2$, which suggests starting the sum in (30) at $j = \theta_1/2$ and sum in both directions (increasing and decreasing j) until convergence. This was originally proposed by Posten (1993), who showed that the error incurred from truncating (30) is less than ϵ when

$$\sum_{j=\max(0,[\theta_1/2]-k)}^{[\theta_1/2]+k} \omega_j \geq 1 - \epsilon.$$

Empirical evidence suggests that

$$\arg\max_j (d_j) \approx \arg\max_j (c_j) \approx \theta_1/2 \tag{33}$$

when $f \geq \mathcal{E}\left[F^{(1)}\right]$, with both $\arg\max_j (d_j)$ and $\arg\max_j (c_j)$ decreasing as f moves to the left of $\mathcal{E}\left[F^{(1)}\right]$. For large θ_1, this "summing outwards" is both time saving and numerically more stable. This holds for the density (29) as well, but the d_i need to be evaluated in terms of their logs for numeric precision.

Figure 1 somewhere here.

As an example, for the $F^{(1)}$ density with parameters (32), the renormalized density SPA approximation (with constant of integration 1.01389) is graphically indistinguishable from the exact density, and yielded an almost constant relative percentage error, i.e., $100 \times (\text{exact} - \text{SPA}) / \text{exact}$, of -0.00884 over the support for which $p_{F^{(1)}}(f) > 10^{-10}$. Besides its accuracy, it is also worth emphasizing that the density SPA is completely closed form and far faster than even the SPA cdf evaluation, with its biggest computational bottleneck being the accurate evaluation of the normal cdf. Thus, except for the numerical integration of the SPA density (which is also extremely fast and can often be ignored given its closeness to one), a grid of several hundred density values are virtually instantaneously calculated on a typical desktop computer—for both the singly and doubly noncentral F density.

[1] Such a method is implemented in function `ncfcdf` in Matlab (version 5.1), whereby they stop summing when $t_j / \left(T_j + \epsilon^{1/4}\right) < \epsilon^{1/2}$, where $\epsilon = 2.2 \times 10^{-16}$ represents machine tolerance and $T_j = \sum_{i=0}^{j} t_i$. For the example in (32), $j_{\max}$ would be set to zero and the algorithm fails.

Note finally that evaluation of $e^{-\theta_1/2}$ will result in underflow[2] for large θ_1 ($\approx$ 1,416 on double precision machines), so that instead of factoring the term out of the ω_j, it is numerically safer (albeit slower) to evaluate the Poisson weights as (with $\lambda = \theta_1/2$)

$$\frac{e^{-\lambda}\lambda^j}{j!} = \exp\left\{-\lambda + j\ln\lambda - \ln\,,\ (j+1)\right\}.$$

Randall (1994) provides an alternative method which entails splitting up the exponent $e^{-\lambda}$ when necessary.

We take as exact values those delivered by the program available from Randall. An alternative method is to use the more general algorithm of Imhof (1961), as given in (36) with $\theta_2 = 0$.[3] Besides requiring far more computing time, the method breaks down for small n_1 and n_2. For example, with $n_1 = n_2 = 1$ and $f = 10$, Imhof fails (returns cdf values outside of $[0,1]$) for $0 < \theta_1 < 31$, while for $31 \le \theta_1 < 40$, its computed cdf differs from the exact value, worsening as θ_1 decreases. As f decreases (increases), matters get worse (better). As n_1 and n_2 increase, accuracy improves quickly, but it is not clear what values are completely safe (for $f = 0.01$, $n_1 = 3$, $n_2 = 6$ and $\theta = 0.1$, Imhof still differs in the third significant digit).

3.2 Doubly Noncentral F

The density and cdf of $F^{(2)} \sim F_{n_1,n_2}(\theta_1,\theta_2)$ at f can be expressed as (Johnson *et al.*, 1995, eq. 30.49 and 30.51)

$$p_{F^{(2)}}(f) = \frac{n_1}{n_2}\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}\frac{\omega_{j,\theta_1}\omega_{k,\theta_2}}{B\left(\frac{n_1}{2}+j,\frac{n_2}{2}+k\right)}\frac{\left(\frac{n_1}{n_2}f\right)^{n_1/2+j-1}}{\left(1+\frac{n_1}{n_2}f\right)^{(n_1+n_2)/2+j+k}} \tag{34}$$

and

$$\Pr\left(F^{(2)} < f\right) = \sum_{k=0}^{\infty}\omega_{k,\theta_2}\sum_{j=0}^{\infty}\omega_{j,\theta_1}I_x\left(\frac{n_1}{2}+j,\frac{n_2}{2}+k\right) \tag{35}$$

respectively, with $\omega_{i,\theta}$ given in (31) and $x = n_1 f\,/\,(n_2 + n_1 f)$.

First consider the pdf (34). Clearly, it will be computationally burdensome to evaluate the double sum until convergence and, as in the $F^{(1)}$ case, possibly numerically unstable as well. For the cdf (35), it appears that a similar result to (33) applies, so that each sum in (34) and (35) may be evaluated

[2] As noted by Randall, this might explain the underflow in the `probf` function in SAS (version 6.07) for the parameters given in (32). In addition, Knüsel (1995) documented several computing errors in Gauss-386 (version 3.2.6), in particular for the singly noncentral cdf function `cdffnc` with parameters $f = 100$, $n_1 = 10$, $n_2 = 1$ and $\theta_1 = 38$ and 39.

[3] We use the Pascal implementation from Farebrother (1990) in extended (19 significant digit) accuracy, with an error tolerance of 10^{-14}.

by summing outwards, i.e., k would start at $\theta_2/2$ and work outwards and, for each k, j would work outwards from $\theta_1/2$. This works well in practice, and is the method we deem exact when referring to the evaluation of (35).

A practical alternative to computing (35) might be the Imhof algorithm, which can evaluate the cdf of a general weighted sum of noncentral χ^2 random variables. For the cdf of $F^{(2)} \sim F_{n_1,n_2}(\theta_1,\theta_2)$ at f, Imhof computes

$$\Pr\left(F^{(2)} < f\right) = \Pr\left(X \leq 0\right) = \frac{1}{2} - \frac{1}{\pi}\int_0^\infty \frac{\sin\theta\left(x\right)}{x\,\rho\left(x\right)}dx \tag{36}$$

where

$$\theta\left(x\right) = \frac{1}{2}\left(n_1 \tan^{-1}\kappa_2 + \frac{\theta_1\kappa_2}{1+\kappa_2^2} - n_2\tan^{-1}\kappa_1 - \frac{\theta_2\kappa_1}{1+\kappa_1^2}\right),$$

$$\rho\left(x\right) = \left(1+\kappa_2^2\right)^{n_1/4}\left(1+\kappa_1^2\right)^{n_2/4}\exp\left\{\frac{1}{2}\left(\frac{\theta_1^2\kappa_2^2}{1+\kappa_2^2} + \frac{\theta_2^2\kappa_1^2}{1+\kappa_1^2}\right)\right\}$$

and $\kappa_1 = n_1 x f$, $\kappa_2 = n_2 x$ (Ennis and Johnson, 1993). Imhof derived expressions to explicitly control the error arising in (36) so that, at least in principle, a specified degree of accuracy can be attained. Improvements on and minor corrections to the implementation by Koerts and Abrahamse (1969) were provided by Farebrother (1980a, 1990). Some performance aspects of the Imhof procedure were investigated by Farebrother (1984), while Helstrom (1996) drew attention to the potential inaccuracy of the Imhof method in the tails of the distribution.

As in the $F^{(1)}$ case, Imhof fails or exhibits lack of accuracy for the cdf of $F^{(2)}$ when θ_1 and/or θ_2 is small, particularly so when f is not in the far right tail of the distribution.

4 Saddlepoint Accuracy

4.1 Singly Noncentral F CDF

We evaluate the accuracy delivered from the second order Lugannani-Rice saddlepoint method (5) via a design over the four parameters n_1, n_2, θ_1 and f. In particular, for $n_1 = 1$, we take $f = 1.1, 11.1, \ldots, 91.1$; and for $n_2 = 10$, $f = 1.1, 3.1, \ldots, 19.1$, so as to capture a reasonable portion of the support for both cases. For both values of n_1, we take $n_2 = 1, 11, \ldots, 101$ and $\theta_1 = 0, 10, \ldots, 50$, yielding a total of 1,320 "sample points". From the discussion in Section 2, the saddlepoint approximation should exhibit more accuracy both as n_1 and n_2 increase (more summing) and as f increases (farther into the tail). One might postulate that, all things being equal, the smaller θ_1, the higher the accuracy, given the elongated right tail for large θ_1.

This is precisely what we observe. Figures 2 and 3 correspond to $n_1 = 1$ and $n_1 = 10$, respectively, and plot the absolute relative percentage error (APE, $100\,|t-a|\,/t$ where t is the true and a is the approximate value) on a $\log_{10}$ scale for fixed θ_1 and f, as a function of n_2. The largest APE was 2.9 (not surprisingly corresponding to the most extreme case, for which $n_1 = n_2 = 1$, $\theta_1 = 50$ and $f = 1.1$, with exact $\Pr(F^{(1)} < f) = 1.06 \times 10^{-6}$, saddlepoint cdf 1.03×10^{-6}). Particularly for smaller θ_1, it is clear that accuracy increases smoothly as both n_2 and f increase and decreases for larger θ_1. Although it would appear from a juxtaposition of Figures 2 and 3 that much higher accuracy is obtained for $n_1 = 10$ as opposed to $n_1 = 1$, one should keep in mind that these graphs are not directly comparable, given the nonlinear nature of the cdf as a function of the parameters. Nevertheless, it is clear that a severalfold increase in accuracy occurs as n_1 goes from one to ten.

Figure 2 somewhere here.

Figure 3 somewhere here.

In assessing the increase in accuracy due to the second order term in (5), we found that, when measured in terms of APE, (5) was, on average, over 32 times as accurate as (2) but still, in 67 of the 1,320 cases, (2) was preferable to (5). Of these, 29 had $n_1 = n_2 = 1$, and for the remaining, at least one n_i was 1. In the majority of the 67 cases, f was such that $\Pr\left(F^{(1)} < f\right) < 0.6$. In the worst case discussed above, however, (2) yielded an APE of 7.4 (compared to 2.9 for with the second order term), showing that clear-cut rules when one method is to be preferred are not available. In general, the second order correction made the highest contributions for n_1 and/or n_2 large, and f farther into the right tail, irrespective of θ_1.

It is worth mentioning that the cdf evaluation for all 1,320 points took 0.13 seconds using the Randall program, as opposed to just over 10 minutes using the Imhof procedure,[4] implying about a 4,600 factor speed increase. The SPA required 0.040 seconds, or a 14,000 factor speed increase over Imhof.

4.2 Doubly Noncentral F CDF

With the additional parameter θ_2, a study as in the previous section is not feasible. Instead, we use degrees of freedom values $n_1 = n_2 = 1$, for which we expect the saddlepoint performance to be the worst. For each pair of $\theta_1 = 1, 21, \ldots, 501$ and $\theta_2 = 1, 21, \ldots, 501$, we compute the accuracy of the cdf at that value f such that $\Pr\left(F^{(2)} < f\right) = \alpha$, $\alpha = 0.80$ and 0.99. The value of f is iteratively found using the saddlepoint approximation (due to its large speed advantage). Figure 4 shows the

[4]Both were computed using Borland Pascal programs using a 300Mz Pentium-II PC.

APE values for all θ_1 and several θ_2, including the worst case, $\theta_2 = 1$. Roughly speaking, performance increases as one or both θ_i increase. It appears that, for constant θ_2, error first decreases as θ_1 increases, reaching an "optimal point", and then increases slightly and levels off to a constant value. Accuracy is overall higher for $\alpha = 0.99$ compared to $\alpha = 0.80$, most noticeably for $\theta_2 = 1$.

Figure 4 somewhere here.

¿From these plots, it might appear that accuracy starts to breaks down as both θ_i approach zero. To investigate this in more detail, Figure 5 shows the percentage error for $n_1 = n_2 = 1$, and $\theta_1 = \theta_2 = 0, 0.3, 0.6, 0.9, 1.2$.[5] The top panel uses f values 2, 12, 22, ..., 992, while the bottom panel uses $f = 0.02, 0.04, \ldots, 2.0$. From the right (i.e., large f), the percent error worsens as f approaches a certain point, then improves, reaching zero at the median $f = 1.0$, then steadily increases again, reaching its largest values (about 7.5% for $\theta_1 = \theta_2 = 1.2$) and finally moving towards zero as f approaches zero in the left tail.

Figure 5 somewhere here.

As noted, we chose the values $n_1 = n_2 = 1$ hoping to capture the SPA at its worst, but observed reasonable performance, particularly as the θ_i increase. In comparison, Figure 6 shows the same plot as in Figure 4, but using $n_1 = 10$ and $n_2 = 20$. Qualitatively, the graphs are very similar, but, particularly for the smaller values of θ_i, accuracy increases approximately 100 fold. Clearly, for values of n_i which typically occur in practice, the SPA exhibits remarkable accuracy.

Figure 6 somewhere here.

To appreciate the speed gains from using the SPA, Table 1 indicates by how much longer the exact methods took, i.e., the entries are normalized such that the saddlepoint time is unity. One sees immediately that both exact routines require several orders of magnitude longer to compute. Whereas the computing time for the closed-form SPA is unaffected by the parameters, we see that the time required for the exact methods, most notably Imhof, vary greatly. As θ_2 (and θ_1) increase, more summing is required in (35), so that, for large values of θ_1, θ_2, the Imhof algorithm becomes more attractive, provided of course, that n_1 and n_2 are large enough so that the Imhof method does not break down.

Table 1 somewhere here.

[5] For $\theta_1 = \theta_2 = 0$, the distribution simplifies to the square of a Cauchy random variable, say $G = Z_1^2/Z_2^2$ with cdf $\Pr(G \leq g) = \pi^{-1}\left(\tan^{-1}\sqrt{g} - \tan^{-1}\left(-\sqrt{g}\right)\right)$.

4.3 Comparison with the Tiku Approximation

Tiku (1972) proposed approximating the distribution of $F^{(2)}$ with that of a location-scale shifted central F, i.e.,

$$\Pr\left(F_{n_1,n_2}(\theta_1,\theta_2) < x\right) \approx \Pr\left(F_{\hat{v}_1,n_2}(0,0) < \frac{x+\hat{a}}{\hat{h}}\right) \tag{37}$$

by matching the first three moments and using the fact that

$${}_1F_1\left(r, \frac{n_2}{2}, -\frac{\theta_2}{2}\right) \approx (1+\theta_2/n_2)^{-r} \tag{38}$$

for θ_2/n_2 near zero. In particular, denoting by ${}_i\mu'_r$ and ${}_i\mu_r$ the r^{th} raw and central moments of $F^{(i)}$, respectively, $i = 1, 2$, Tiku (1964) has shown that

$${}_2\mu'_r = {}_1\mu'_r \cdot {}_1F_1\left(r, \frac{n_2}{2}, -\frac{\theta_2}{2}\right), \tag{39}$$

and expressions for the first four moments of $F^{(1)}$ are well-known (Johnson *et al.*, 1995). With (38), moments ${}_2\mu'_r$ are easily computed.

With $\beta_1 = {}_2\mu_3^2 \,/\, {}_2\mu_2^3$, this results in

$$\hat{a} = \frac{hn_2}{n_2-2} - {}_2\mu'_1, \quad \hat{h} = \frac{n_1(n_2-2)(n_2-6)\ {}_2\mu_3}{4n_2(2v+n_2-2)\ {}_2\mu_2},$$

and

$$\hat{v}_1 = \frac{n_2-2}{2}\left(-1+\sqrt{\frac{(n_2-6)^2\beta_1}{\beta_1(n_2-6)^2 - 32(n_2-4)}}\right).$$

Although his calculations are correct, the formula for $\hat{v}_1$ in Tiku (1972, eq. 8) contains a misprint. We denote this method $T3$.

Tiku and Yip (1978) gave the extension to the first four moments by providing values $\hat{a}$, $\hat{h}$, $\hat{v}_1$ and $\hat{v}_2$ for (37) but with $F_{\hat{v}_1,\hat{v}_2}$, which can be used with any target distribution, provided that its first four moments can be readily computed. We denote as $T4$ this method applied to $F^{(2)}$ in conjunction with (38). Note that $T3$ is only valid for $n_2 > 6$; likewise, $T4$ is only valid for $n_2 > 8$.

For illustration, we reproduce part[6] of Table 1 of Tiku (1972) in Table 2, except we show the relative percentage error (instead of absolute differences) of both $T3$ and the SPA. In the central F case for which $\theta_1 = \theta_2 = 0$, $T3$ is exact. As ϕ_2 increases, approximation (38) loses its accuracy and $T3$ begins to break down. The SPA is not only far more accurate in virtually all the cases (except, of

[6]The results for the entire table in Tiku (1972) were computed and are available upon request. The results were qualitatively identical to those shown here, except that $T3$ performs even worse relative to the SPA as n_1 increases.

course, for the central F, but also for the two cases $n_1 = 4$, $n_2 = 60$, $\phi_1 = 0$, $\phi_2 = 0.5$ and $n_1 = 4$, $n_2 = 8$, $\phi_1 = 1$, $\phi_2 = 0$ for which they are close), but its error is also far more consistent.

Table 2 somewhere here.

What is not clear in the $T3$ approximation is the contribution to error from (38). Naturally, the moments-based approximations will lose their simplicity once ${}_1F_1$ needs to be computed numerically, but it is worth knowing if they lead to better approximations to the cdf of $F^{(2)}$. For comparison, Table 3 is similar to Table 2, but shows the accuracy of both the $T3$ and $T4$ approximations using the exact value of ${}_1F_1$ (computed via numeric integration) instead of approximation (38). Because $T4$ does not exist for $n_2 = 8$, we replace those cases by $n_2 = 10$.

Table 3 somewhere here.

For $T3$, as would be expected, there is a significant increase in accuracy for larger values of ϕ_2 from using the exact ${}_1F_1$ instead of (38), although not all cases improve. Worse yet, the method fails occasionally by returning a negative scale parameter; and this occurs more frequently with increasing ϕ_2, precisely when approximation (38) breaks down, rendering $T3$ using either (38) or the exact value of ${}_1F_1$ useless for values of θ_2/n_2 much larger than zero.

The $T4$ approximation turned out to be somewhat of a disappointment. It is much more reliable than $T3$, but still breaks down occasionally, and provides an increase in accuracy in only about half the cases in Table 3. It also also exhibits non-uniformity in error, which increase with ϕ_2.

Summarizing, with the exception of the central F case, the SPA is almost always several orders of magnitude more accurate than both Tiku approximations, never fails, and is accurate over the entire parameter space. Also, the two Tiku approximations require that n_2 be greater than 6 and 8, respectively. Finally, the SPA only requires evaluation of the normal cdf and not the incomplete beta function, and is otherwise closed–form.

5 Application

To be finished.

6 Conclusions

We present a saddlepoint approximation (SPA) to the density and cdf of the singly and doubly noncentral F distribution and, in both cases, give a closed form solution to the saddlepoint equation, thus obviating the need for numeric root searching.

Given the availability of cheap computing power and the stability and accuracy of the Randall algorithm, we recommend its use for the singly noncentral F cdf, particularly when extreme accuracy is needed. Plots of the singly noncentral F pdf, however, benefit tremendously from the SPA in terms of speed, as well as simplicity. Furthermore, at least three statistical software packages return incorrect values for the $F^{(1)}$ cdf, while very few even offer calculation of the cdf (or pdf) of the doubly noncentral F.

The time savings associated with the SPA in the doubly noncentral case compared with either the Imhof algorithm or "outward summing" of (35) is so dramatic that, not only for possible simulation purposes, but also for routine calculation and inference, the saddlepoint method is recommended. Moreover, the saddlepoint density offers (possibly the only) computational method which is not only numerically reliable and virtually exact, but also evaluated practically instantaneously.

As Johnson *et al.* (1995, p. 491) note regarding research contributions for evaluation of the noncentral F, "...clear examples of multiplication and overlap of results are prevalent in the statistical literature, especially in the area of statistical algorithms. Much of it is due to lack of coordination, almost identical publication in different journals, and unjustified publication of results providing 'epsilon' improvement". We hope that this contribution does not fall victim to that criticism, but instead provides a highly workable solution to a complex issue.

Appendix

Saddlepoint Roots

The three roots to the cubic equation are

$$z_1 = (s_1 + s_2) - \frac{a_2}{3}, \quad z_{2,3} = -\frac{1}{2}(s_1 + s_2) - \frac{a_2}{3} \pm \frac{i\sqrt{3}}{2}(s_1 - s_2) \tag{40}$$

and they are all real if $m < 0$. It is known in general that the a unique root of the saddlepoint equation exists, so that one may take any real z_i that is in the range (11). We now show that all three roots are real and that z_3 is always the unique saddlepoint solution Showing that all three roots in (40) are real is equivalent to showing that m in (14) is negative (Abramowitz and Stegun, 1972, p. 17). Using Maple, m can be factored as

$$m = -\frac{1}{6912}(n_1 f + n_2)^2 \times \frac{n_1}{f^6 n_2^5 (n_1 + n_2)^4} \times Q,$$

where Q consists of the sum of 47 terms in n_1, n_2, θ_1, θ_2 and f, all of which are positive, except for one, given by $-8\theta_2^2 n_2 f^2 \theta_1^2 n_1$. By combining this term with two other (positive) terms $4n_2^2\theta_2^3\theta_1 f$ and

$4\theta_2 n_1^2\theta_1^3 f^3$ in Q (found after some trial and error), we have

$$-8\theta_2^2 n_2 f^2\theta_1^2 n_1 + 4n_2^2\theta_2^3\theta_1 f + 4\theta_2 n_1^2\theta_1^3 f^3 = 4\theta_2 f\theta_1 \left(n_2\theta_2 - n_1\theta_1 f\right)^2 > 0,$$

showing that $m < 0$ and all three roots are real.

To show that z_3 is always the correct root, we first demonstrate that, if the roots are ordered along the real line, then $K_X'' > 0$ for the middle root and is negative at the other two. Then we show that $z_2 < z_3 \leq z_1$ demonstrating that z_1 and z_2 cannot be the saddlepoint. To see the former, define

$$L\left(s\right) := \left(1 - 2s\frac{n_2}{n_1}\right)^2 \left(1 + 2sf\right)^2 K_X'\left(s\right)$$

and note that a root of K_X' is also a root for L. Differentiating gives

$$L'\left(s\right) = \frac{\partial}{\partial s}\left[\left(1 - 2s\frac{n_2}{n_1}\right)^2 \left(1 + 2sf\right)^2\right] \times K_X'\left(s\right) + \left(1 - 2s\frac{n_2}{n_1}\right)^2 \left(1 + 2sf\right)^2 K_X''\left(s\right)$$

and, when evaluated at roots z_i, $i = 1, 2, 3$,

$$L'\left(z_i\right) = \left(1 - 2z_i\frac{n_2}{n_1}\right)^2 \left(1 + 2z_i x\right)^2 K_X''\left(z_i\right)$$

so that

$$\operatorname{sgn}\{L'\left(z_i\right)\} = \operatorname{sgn}\{K_X''\left(z_i\right)\}. \tag{41}$$

Now $L\left(s\right)$ is a cubic polynomial whose leading term is $-8n_2^2 f^2\left(n_1 + n_2\right) n_1^{-2} s^3$ with a negative coefficient; thus $L'\left(s\right) > 0$ when s is the middle root and negative at the first and third ordered roots. From (41), the same holds for K_X'' so the middle root must be the saddlepoint.

To prove that $z_2 < z_3 \leq z_1$, first note that r in (14) is real and recall that $m < 0$ implying that also $q < 0$. Then

$$r + \sqrt{m} = r + i\sqrt{-\left(q^3 + r^2\right)} = \sqrt{-q^3}e^{i\lambda}$$

where $\lambda = \arg\left(r + \sqrt{m}\right) \in \left(0, \pi\right)$ and, more specifically, if $r < 0$, then $\lambda \in \left(\pi/2, \pi\right)$. Then

$$s_{1,2} = \left(r \pm \sqrt{m}\right)^{1/3} = \left(\sqrt{-q^3}e^{\pm i\lambda}\right)^{1/3} = \sqrt{-q}e^{\pm i\lambda/3}$$

and thus,

$$s_1 + s_2 = \sqrt{-q}\left(e^{i\lambda/3} + e^{-i\lambda/3}\right) = 2\sqrt{-q}\cos\left(\lambda/3\right) > 0 \tag{42}$$

is real and positive, since $0 \leq \lambda/3 < \pi/3 = 60°$. Similarly,

$$s_1 - s_2 = \sqrt{-q}\left(e^{i\lambda/3} - e^{-i\lambda/3}\right) = 2i\sqrt{-q}\sin\left(\lambda/3\right)$$

so that

$$i\frac{\sqrt{3}}{2}(s_1 - s_2) = -\sqrt{-3q}\sin(\lambda/3) < 0. \tag{43}$$

From (43) it follows directly that $z_2 < z_3$. Now, comparing z_1 and z_3,

$$z_3 = -\sqrt{-q}\cos(\lambda/3) - \frac{a_2}{3} + \sqrt{3}\sqrt{-q}\sin(\lambda/3) \overset{?}{\leq} 2\sqrt{-q}\cos(\lambda/3) - \frac{a_2}{3} = z_1$$

or

$$\sqrt{3}\sqrt{-q}\sin(\lambda/3) \overset{?}{\leq} 3\sqrt{-q}\cos(\lambda/3)$$

or

$$\tan(\lambda/3) \overset{?}{\leq} \sqrt{3}.$$

But

$$\tan(\lambda/3) \leq \tan(\pi/3) = \sqrt{3}$$

so that $z_3 \leq z_1$ and $z_2 < z_3 \leq z_1$. ■

Simplifying the above expressions for z_3 shows that it can be expressed as

$$z_3 = \sqrt{-q}\left\{\sqrt{3}\sin(\phi) - \cos(\phi)\right\} - \frac{a_2}{3}$$

where

$$\phi = \frac{1}{3}\arg\left(r + \sqrt{m}\right).$$

Relative error of the cdf approximation

The tail probability of F from the inversion formula is

$$\begin{aligned}
\Pr(X_f > 0) &= \frac{1}{2\pi i}\int_{\hat{s}-i\infty}^{\hat{s}+i\infty} z^{-1} M_{X_f}(z)\, dz \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} (\hat{s}+it)^{-1} M_{X_f}(\hat{s}+it)\, dt \\
&= \frac{M_{X_f}(\hat{s})}{\hat{s}\hat{\sigma}}\frac{1}{2\pi}\int_{-\infty}^{\infty}\left(1+\frac{it}{\hat{s}\hat{\sigma}}\right)^{-1}\frac{M_{X_f}\left(\hat{s}+\frac{it}{\hat{s}\hat{\sigma}}\right)}{M_{X_f}(\hat{s})}dt
\end{aligned}$$

where the last integral results from a scale change to the variable of integration. Taking $\hat{\sigma} = \sqrt{K''_{X_f}(\hat{s})}$, then the first order SPA

$$\widehat{\Pr}(X_f > 0) \sim \frac{M_{X_f}(\hat{s})}{\hat{s}\hat{\sigma}\sqrt{2\pi}}$$

as $f \to \infty$ so that

$$\lim_{f\to\infty}\frac{\Pr(X_f > 0)}{\widehat{\Pr}(X_f > 0)} = \sqrt{2\pi}\lim_{f\to\infty}\frac{1}{2\pi}\int_{-\infty}^{\infty}\left(1+\frac{it}{\hat{s}\hat{\sigma}}\right)^{-1}\frac{M_{X_f}\left(\hat{s}+\frac{it}{\hat{s}\hat{\sigma}}\right)}{M_{X_f}(\hat{s})}dt. \tag{44}$$

The limiting inversion is determined by finding the pointwise limit of its integrand and applying the dominated convergence theorem.

There are 4 ratio terms in the M_{X_f}-ratio of (44) derived from the 4 terms in (10). The last ratio depends directly upon f and simple computations show it is bounded for sufficently large f as

$$\left|\exp\left\{-\frac{\left(\hat{s}+\frac{it}{\hat{s}\hat{\sigma}}\right)f\theta_2}{1+2\left(\hat{s}+\frac{it}{\hat{s}\hat{\sigma}}\right)f}+\frac{\hat{s}f\theta_2}{1+2\hat{s}f}\right\}\right| \leq (1+\varepsilon)\,e^{\theta_2/2} \tag{45}$$

for some $\varepsilon > 0$. The remaining portion of the integrand is bounded by $(1+\varepsilon)\,h(t)$, where $h(\cdot)$ is the integrable limit derived below.

The pointwise limit of the left side of (45) is 1 so the limiting integrand is based upon the first 3 terms. Since $\hat{s} \to \omega_0$, computations show that the ratio

$$\frac{M_{X_f}\left(\hat{s}+\frac{it}{\hat{s}\hat{\sigma}}\right)}{M_{X_f}(\hat{s})} \sim (1-2it\nu_0)^{-\frac{n_1}{2}}\left(1+\frac{it}{\omega_0\sigma_0}\right)^{-\frac{n_2}{2}}\exp\left\{\frac{\left(\omega_0+\frac{it}{\sigma_0}\right)\frac{n_2}{n_1}\theta_1}{1-2\left(\omega_0+\frac{it}{\sigma_0}\right)\frac{n_2}{n_1}}-\vartheta_1\right\}$$

where

$$\nu_0 = \frac{n_2}{n_1\sigma_0}\left(1-2\frac{n_2}{n_1}\omega_0\right)^{-1}, \tag{46}$$

$$\sigma_0^2 = \lim_{f\to\infty} K''_{X_f}(\hat{s}) = 2n_1n_2^2\frac{n_1+2\theta_1-2\omega_0n_2}{(n_1-2\omega_0n_2)^3}+\frac{n_2}{2\omega_0^2} > 0, \tag{47}$$

and

$$\vartheta_1 = \frac{\omega_0\frac{n_2}{n_1}\theta_1}{1-2\omega_0\frac{n_2}{n_1}}.$$

Writing the exponential term as

$$\frac{\left(\omega_0+\frac{it}{\sigma_0}\right)\frac{n_2}{n_1}\theta_1}{1-2\left(\omega_0+\frac{it}{\sigma_0}\right)\frac{n_2}{n_1}} = -\frac{\theta_1}{2}+\frac{\theta_1/2}{1-2\left(\omega_0+\frac{it}{\sigma_0}\right)\frac{n_2}{n_1}}$$

then

$$\frac{M_{X_f}\left(\hat{s}+\frac{it}{\hat{s}\hat{\sigma}}\right)}{M_{X_f}(\hat{s})} \sim (1-2it\nu_0)^{-\frac{n_1}{2}}\left(1+\frac{it}{\omega_0\sigma_0}\right)^{-\frac{n_2}{2}}e^{-\vartheta_1-\theta_1/2}\exp\left\{\frac{\nu_1}{1-2it\nu_0}\right\}$$

where

$$\nu_1 = \frac{\theta_1}{2}\left(1-2\frac{n_2}{n_1}\omega_0\right)^{-1}. \tag{48}$$

Taylor expansion of the exp function shows that the integrand has the limiting form

$$\left(1+\frac{it}{\hat{s}\hat{\sigma}}\right)^{-1}\frac{M_{X_f}\left(\hat{s}+\frac{it}{\hat{s}\hat{\sigma}}\right)}{M_{X_f}(\hat{s})} \sim \left(1+\frac{it}{\omega_0\sigma_0}\right)^{-\frac{n_2}{2}-1}\sum_{k=0}^{\infty}e^{-\nu_1}\frac{\nu_1^k}{k!}(1-2it\nu_0)^{-\frac{n_1}{2}-k}.$$

This is the product of two characteristic functions (cfs): for $Y_2 = -(2\omega_0\sigma_0)^{-1}\chi^2_{n_2+2}$ and $Y_1 = \nu_0\chi^2(n_1, 2\nu_1)$. Inversion of this product is the density of $Y_1 - Y_2$ at 0 which leads to the asymptotic relative error in (27).

Straightforward computation determines $g_{Y_1-Y_2}(0)$, the density of $Y_1 - Y_2$ at 0, as

$$\frac{(2\nu_0\omega_0\sigma_0)^{\frac{n_2}{2}}\omega_0\sigma_0 e^{-\nu_1}}{(1+2\nu_0\omega_0\sigma_0)^{\frac{1}{2}(n_1+n_2)}, \left(\frac{n_2}{2}+1\right)}\sum_{k=0}^{\infty}\frac{, \left(\frac{n_1+n_2}{2}+k\right)}{, \left(\frac{n_1}{2}+k\right)k!}\left(\frac{\nu_1}{1+2\nu_0\omega_0\sigma_0}\right)^k.$$

Using $\nu_1(1+2\nu_0\omega_0\sigma_0)^{-1} = \theta_1/2$ and expressing the summation in terms of the Taylor expansion of the confluent hypergeometric function given in (13.1.2) of Abramowitz and Stegun (1970), then the summation is

$$\frac{, \left(\frac{n_1+n_2}{2}\right)}{, \left(\frac{n_1}{2}\right)}\,{}_1F_1\left(\frac{n_1+n_2}{2}; \frac{n_1}{2}; \frac{\theta_1}{2}\right)$$

so that

$$g_{Y_1-Y_2}(0) = \frac{(2\nu_0\omega_0\sigma_0)^{\frac{n_2}{2}}\omega_0\sigma_0 e^{-\nu_1}}{(1+2\nu_0\omega_0\sigma_0)^{\frac{1}{2}(n_1+n_2)}B\left(\frac{n_1}{2}, \frac{n_2}{2}\right)\frac{n_2}{2}}\,{}_1F_1\left(\frac{n_1+n_2}{2}; \frac{n_1}{2}; \frac{\theta_1}{2}\right). \tag{49}$$

Relative error of the density approximation

Follwing the same approach used for the cdf,

$$\begin{aligned} g_{W_f}(0) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} M_{W_f}(\hat{s}+it)\,dt \\ &= \frac{M_{W_f}(\hat{s})}{\hat{\sigma}}\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{M_{W_f}(\hat{s}+it/\hat{\sigma})}{M_{W_f}(\hat{s})}dt \\ &= \sqrt{2\pi}\tilde{g}_{W_f}(0)\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{M_{W_f}(\hat{s}+it/\hat{\sigma})}{M_{W_f}(\hat{s})}dt \end{aligned}$$

and

$$\lim_{f\to\infty}\frac{g_F(f)}{\tilde{g}_F(f)} = \lim_{f\to\infty}\frac{g_{W_f}(0)}{\tilde{g}_{W_f}(0)} = \sqrt{2\pi}\lim_{f\to\infty}\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{M_{W_f}(\hat{s}+it/\hat{\sigma})}{M_{W_f}(\hat{s})}dt.$$

The inversion is much the same as with the cdf except as concerns the extra factor (24) that accompanies (23). Denoting this extra factor as $h(s)$, then

$$\frac{h(\hat{s}+it/\hat{\sigma})}{h(\hat{s})} \sim \left(1+\frac{it}{\omega_0\sigma_0}\right)^{-1}$$

and

$$\frac{M_{W_f}(\hat{s}+it/\hat{\sigma})}{M_{W_f}(\hat{s})} \sim \left(1+\frac{it}{\omega_0\sigma_0}\right)^{-1}\frac{M_{X_f}(\hat{s}+it/\hat{\sigma})}{M_{X_f}(\hat{s})},$$

the integrand in the cdf setting. Therefore

$$\lim_{f\to\infty}\frac{g_F(f)}{\tilde{g}_F(f)} = \lim_{f\to\infty}\frac{\Pr(X_f>0)}{\widehat{\Pr}(X_f>0)}. \tag{50}$$

In the central setting,

$$2\nu_0\omega_0\sigma_0 = \frac{n_2}{n_1} \qquad \omega_0\sigma_0 = \sqrt{\frac{(n_1+n_2)\,n_2}{2n_1}}$$

so that the limiting value in (50) is $\hat{B}(\frac{n_1}{2},\frac{n_2}{2})/B(\frac{n_1}{2},\frac{n_2}{2})$.

References

Abramowitz, M. and Stegun, I. A. (1972). *Handbook of Mathematical Functions*. Dover, New York.

Barndorff-Nielsen, O. and Cox, D. R. (1990). *Asymptotic Techniques for Use in Statistics*. Chapman & Hall, London.

Bulgren, W. G. (1971). On Representations of the Doubly non-central F Distribution. *Journal of the American Statistical Association*, 66:184–186.

Chow, S. C. and Shao, J. (1990). On the Difference Between the Classical and Inverse Methods of Calibration. *Applied Statistics*, 39:219–228.

Daniels, H. E. (1954). Saddlepoint Approximation in Statistics. *Ann. Math. Statist.*, 25:631–650.

Daniels, H. E. (1987). Tail Probability Approximation. *International Statistical Review*, 55:37–48.

DeBenedictis, L. F. and Giles, D. E. A. (1998). Diagnostic Testing in Econometrics, Variable Addition, RESET and Fourier Approximations. In Ullah, A. and Giles, D. E. A., editors, *Handbook of Applied Economic Statistics*, pages 383–417. Marcel Dekker, New York.

Ennis, D. M. and Johnson, N. L. (1993). Noncentral and Central Chi-Square, F and Beta Distribution Functions as Special Cases of the Distribution Function of an Indefinite Quadratic Form. *Communications in Statistics – Theory and Methods*, 22(3):897–905.

Farebrother, R. W. (1980a). The Distribution of a Quadratic Form in Normal Variables. *Applied Statistics*, 39:294–309.

Farebrother, R. W. (1980b). The Durbin-Watson Test for Serial Correlation when there is no Intercept in the Regression. *Econometrica*, 48(6):1553–1555. Correction, 1981, 49:277.

Farebrother, R. W. (1984). Remark AS R53. A Remark on Algorithms AS 106, AS 153 and AS 155. The Distribution of a Linear Combination of χ^2 Random Variables. *Applied Statistics*, 33:366–369.

Field, C. and Ronchetti, E. (1990). *Small Sample Asymptotics*. Institute of Mathematical Statistics, Hayward.

Geary, R. C. (1944). Extension of a Theorem by Harald Cramér on the Frequency Distribution of the Quotient of Two Variables. *Journal of the Royal Statistical Society*, 17:56–57.

Helstrom, C. W. (1996). Calculating the Distribution of the Serial Correlation Estimator by Saddlepoint Integration. *Econometric Theory*, 12:458–480.

Helstrom, C. W. and Ritcey, J. A. (1985). Evaluation of the Noncentral F-Distribution by Numerical Contour Integration. *SIAM Journal on Scientific and Statistical Computing*, 8(3):505–514.

Imhof, P. (1961). Computing the Distribution of Quadratic Forms in Normal Variates. *Biometrika*, 48:419–426.

Jensen (1988). Uniform Saddlepoint Approximations. *Adv. Appl. Probab*, 20:622–634.

Jensen, J. L. (1994). *Saddlepoint Approximations*. Oxford University Press, Oxford.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1995). *Continuous Univariate Distributions*, volume 2. John Wiley & Sons, New York, Second edition.

Knüsel, L. (1995). On the Accuracy of the Statistical Distributions in GAUSS. *Computational Statistics and Data Analysis: Statistical Software Newsletter*, 20(6):699–702.

Koerts, J. and Abrahamse, A. P. J. (1969). *On the Theory and Application of the General Linear Model*. University Press, Rotterdam.

Kurumai, H. and Ohtani, K. (1998). The Exact Distribution and Density Functions of a Pre-Test Estimator of the Error Variance in a Linear Regression Model with Proxy Variables. *Statistical Papers*, 39:163–177.

Lenth, R. V. (1987). Computing Noncentral Beta Probabilities. *Applied Statistics*, 36:241–244.

Lieberman, O. (1994). Saddlepoint Approximation for the Distribution of a Ratio of Quadratic Forms in Normal Variables. *Journal of the American Statistical Association*, 89(427):924–928.

Lugannani, R. and Rice, S. O. (1980). Saddlepoint Approximations for the Distribution of Sums of Independent Random Variables. *Adv. Appl. Prob.*, 12:475–490.

Morin-Wahhab, D. (1985). Moments of a Ratio of Two Quadratic Forms. *Communications in Statistics – Theory and Methods*, 14(2):499–508.

Price, R. (1964). Some Non-Central F-Distributions Expressed in Closed Form. *Biometrika*, 51(1 and 2):107–122.

Provost, S. B. and Rudiuk, E. M. (1996). The Exact Distribution of Indefinite Quadratic Forms in Noncentral Normal Vectors. *Ann. Inst. Statist. Math.*, 48(2):381–394.

Randall, J. H. (1994). Calculating Central and Non-Central F Probabilities. *South African Journal of Statistics*, 28:67–72.

Reid, N. (1988). Saddlepoint Methods and Statistical Inference (with Discussion). *Statistical Science*, 3:213–238.

Reid, N. (1996). Likelihood and Higher-Order Approximations to Tail Areas: A Review and Annotated Bibliography. *The Canadian Journal of Statistics*, 24(2):141–166.

Schader, M. and Schmid, F. (1986). Distribution Function and Percentage Points for the Central and Noncentral F Distribution. *Statistische Hefte*, 27:67–74.

Scheffé, H. (1959). *The Analysis of Variance*. John Wiley & Sons, New York.

Stuart, A. and Ord, J. (1994). *Kendall's Advanced Theory of Statistics*, volume 1. Edward Arnold, London, sixth edition.

Tang, P. C. (1938). The Power Function of the Analysis of Variance Tests with Tables and Illustrations of their use. *Statistical Research Memoirs*, 2:126–150.

Tiku, M. L. (1964). Approximating to the General Non-Normal Sampling Distributions. *Biometrika*, 51:83–95.

Tiku, M. L. (1972). A Note on the Distribution of the Doubly Noncentral F Distribution. *Australian Journal of Statistics*, 14(1):37–40.

Tiku, M. L. and Yip, D. Y. N. (1978). A Four Moment Approximation Based on the F Distribution. *Australian Journal of Statistics*, 20(3):257–261.

Table 1: Relative computing times for the cdf of $F_{n_1,n_2}(\theta_1,\theta_2)$ [a]

θ_2		Outward Sum of (35)			Imhof		
	$n_1 \diagdown n_2$	1	10	100	1	10	100
1	1	800	600	800	FAIL	4000	16800
1	10	600	600	600	68000	2000	6000
1	100	1000	600	800	540000	9600	2400
10	1	720	640	760	FAIL	2400	18000
10	10	700	700	800	4000	1200	4800
10	100	1100	800	1000	16000	6000	2400
100	1	3200	3200	4000	600	1200	20000
100	10	2400	2400	2400	600	500	6000
100	100	4000	3600	3600	3200	2800	6000

[a] For each entry, $\theta_1 = 10$. The cutoff value f was chosen to approximate the expected value plus one standard deviation, using the well-known approximation $F^{(2)} \approx rF^{(0)}$, where $r = (1+\theta_1/n_1)/(1+\theta_2/n_2)$ and $F^{(0)} \sim F_{a,b}$ is central F with

$$a = \frac{(n_1+\theta_1)^2}{n_1+2\theta_1}, \quad b = \frac{(n_2+\theta_2)^2}{n_2+2\theta_2}.$$

We take $f = r(\mu+\sigma)$, where $\mu = \dfrac{b}{b-2}$ and $\sigma^2 = \dfrac{2b^2(a+b-2)}{a(b-2)^2(b-4)}$ are the mean and variance of $F_{a,b}$. When $b \le 4$, we set $f = 1.1\theta_1/\theta_2$.

For all parameter values, all computing times for the SPA were virtually the same and, on a 300Mz Pentium-II PC, took 4.9e-5 seconds.

Table 2: Relative percentage error for approximating $F_{n_1,n_2}(\theta_1,\theta_2)$[a]

ϕ_2	0.0		0.5		1.0		1.5		2.0		3.0	
ϕ_1	Tiku	SPA	Tiku	SPA	Tiku	SPA	Tiku	SPA	Tiku	SPA	Tiku	SPA
	$n_1 = 4, \quad n_2 = 8$											
0	0.0000	-0.0601	2.0916	-0.0431	8.1594	-0.0284	18.2517	-0.0250	32.8246	-0.0298	79.2336	-0.0511
1	0.0427	0.0745	0.9718	0.0925	3.9490	0.1011	9.0822	0.1004	16.5503	0.0957	39.7361	0.0839
2	0.0216	0.0130	0.0571	0.0179	0.3636	0.0173	1.2002	0.0150	2.7549	0.0126	8.5760	0.0089
3	-0.0037	0.0006	-0.0747	0.0010	-0.2952	0.0007	-0.6332	0.0004	-1.0172	0.0001	-1.5606	-0.0004
	$n_1 = 4, \quad n_2 = 24$											
0	0.0000	-0.1139	0.4495	-0.1156	1.8101	-0.1167	4.1091	-0.1179	7.3903	-0.1192	17.1768	-0.1222
1	0.0494	0.0260	0.0318	0.0288	0.2839	0.0312	0.8621	0.0331	1.8126	0.0345	4.9901	0.0363
2	0.0040	0.0018	0.0113	0.0022	-0.0116	0.0027	-0.0525	0.0030	-0.0943	0.0033	-0.0991	0.0038
3	-0.0041	0.0000	-0.0065	0.0000	-0.0124	0.0000	-0.0241	0.0000	-0.0448	0.0000	-0.1262	0.0001
	$n_1 = 4, \quad n_2 = 60$											
0	0.0000	-0.1193	0.0894	-0.1215	0.3629	-0.1236	0.8274	-0.1254	1.4898	-0.1270	3.4348	-0.1299
1	0.0776	0.0170	-0.0252	0.0180	-0.0910	0.0190	-0.1081	0.0198	-0.0664	0.0205	0.2290	0.0217
2	-0.0263	0.0007	-0.0084	0.0008	0.0052	0.0009	0.0139	0.0010	0.0176	0.0011	0.0113	0.0013
3	-0.0034	0.0000	-0.0044	0.0000	-0.0056	0.0000	-0.0072	0.0000	-0.0093	0.0000	-0.0154	0.0000

[a]Tiku refers to the location-scale central F approximation based on equating the first three moments; SPA refers to the saddlepoint approximation (5); $\phi_1^2 = \theta_1/(n_1+1)$ and $\phi_2 = \theta_2/\sqrt{n_2}$.

Table 3: Relative percentage error for approximating $F_{n_1,n_2}(\theta_1,\theta_2)$[a]

ϕ_2	0.0		0.5		1.0		1.5		2.0		3.0	
ϕ_1	T3	T4	T3	T4	T3	T4	T3	T4	T3	T4	T3	T4
	$n_1 = 4, \quad n_2 = 10$											
0	0	0	0.1892	0.2304	0.7685	1.0598	1.8140	2.7124	3.5098	5.4441	10.0699	14.9699
1	0.0398	-0.0655	0.0477	-0.1893	0.3189	-0.2364	0.8777	-0.0035	N.A.	0.7218	N.A.	4.7536
2	0.0231	0.0179	-0.1623	-0.2711	N.A.	-1.0636	N.A.	-2.2635	N.A.	-3.7539	N.A.	-6.9103
3	-0.0044	-0.0040	N.A.	-0.1007	N.A.	N.A.	N.A.	N.A.	N.A.	-2.3182	N.A.	-5.6116
	$n_1 = 4, \quad n_2 = 24$											
0	0.0000	0.0000	0.0970	0.0960	0.4393	0.4591	1.1019	1.1523	2.1577	2.1909	5.7360	5.2270
1	0.0494	-0.8954	-0.0894	-1.1096	-0.2053	-1.4545	-0.2473	-1.7689	-0.1717	-1.9320	0.4868	-1.5053
2	0.0040	-0.0710	-0.0094	-0.1714	-0.0943	-0.4941	-0.2390	-1.0568	-0.4284	-1.8417	-0.8766	-3.8865
3	-0.0041	-0.0032	-0.0080	-0.0037	-0.0202	-0.0043	-0.0458	-0.0089	-0.0902	-0.0252	N.A.	-0.1506
	$n_1 = 4, \quad n_2 = 60$											
0	0.0000	0.0000	0.0232	0.0164	0.1032	0.0796	0.2543	0.2032	0.4897	0.3936	1.2614	0.9771
1	0.0776	-1.8052	-0.0455	-1.9282	-0.1732	-2.0925	-0.2949	-2.2703	-0.4008	-2.4391	-0.5304	-2.6823
2	-0.0263	-0.1927	-0.0104	-0.2299	-0.0034	-0.3027	-0.0064	-0.4165	-0.0203	-0.5744	-0.0822	-1.0246
3	-0.0034	-0.0024	-0.0044	-0.0030	-0.0058	-0.0036	-0.0077	-0.0041	-0.0103	-0.0048	-0.0185	-0.0079

[a] $T3$ and $T4$ refer to the three and four moment location-scale central F approximations but using the exact ${}_1F_1$ instead of (38). Entries with N.A. indicate that the method failed.

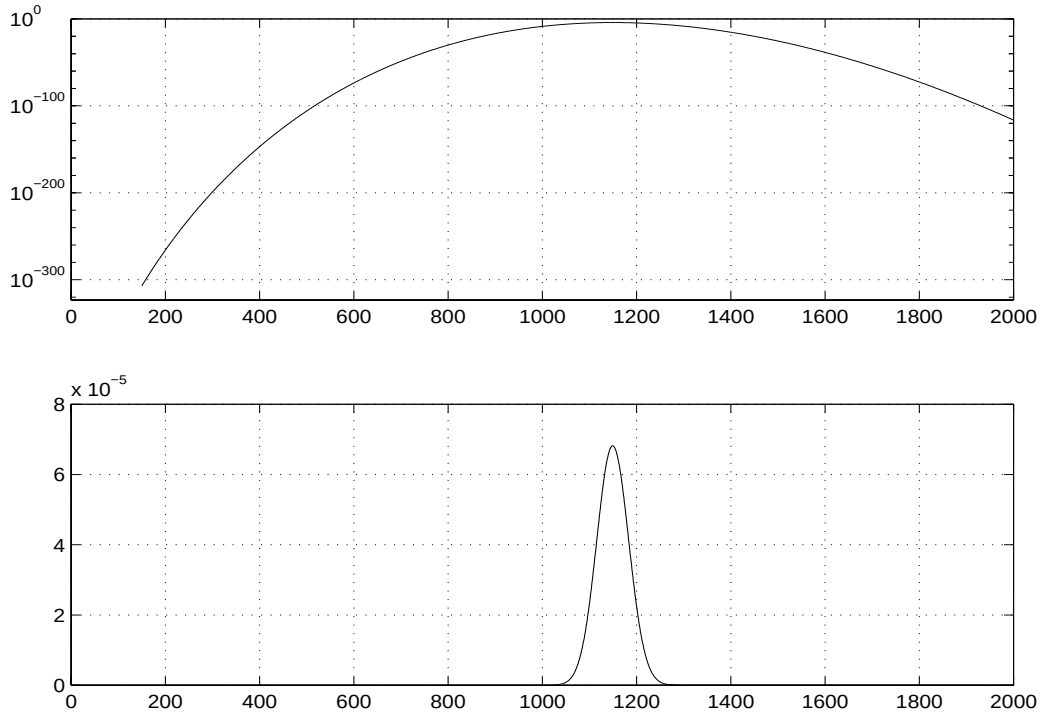

Figure 1: The summands c_j in (30) for $n_1 = 1$, $n_2 = 12$, $\theta_1 = 2316$ and $f = 990$ (top panel uses $\log_{10}$ scale)

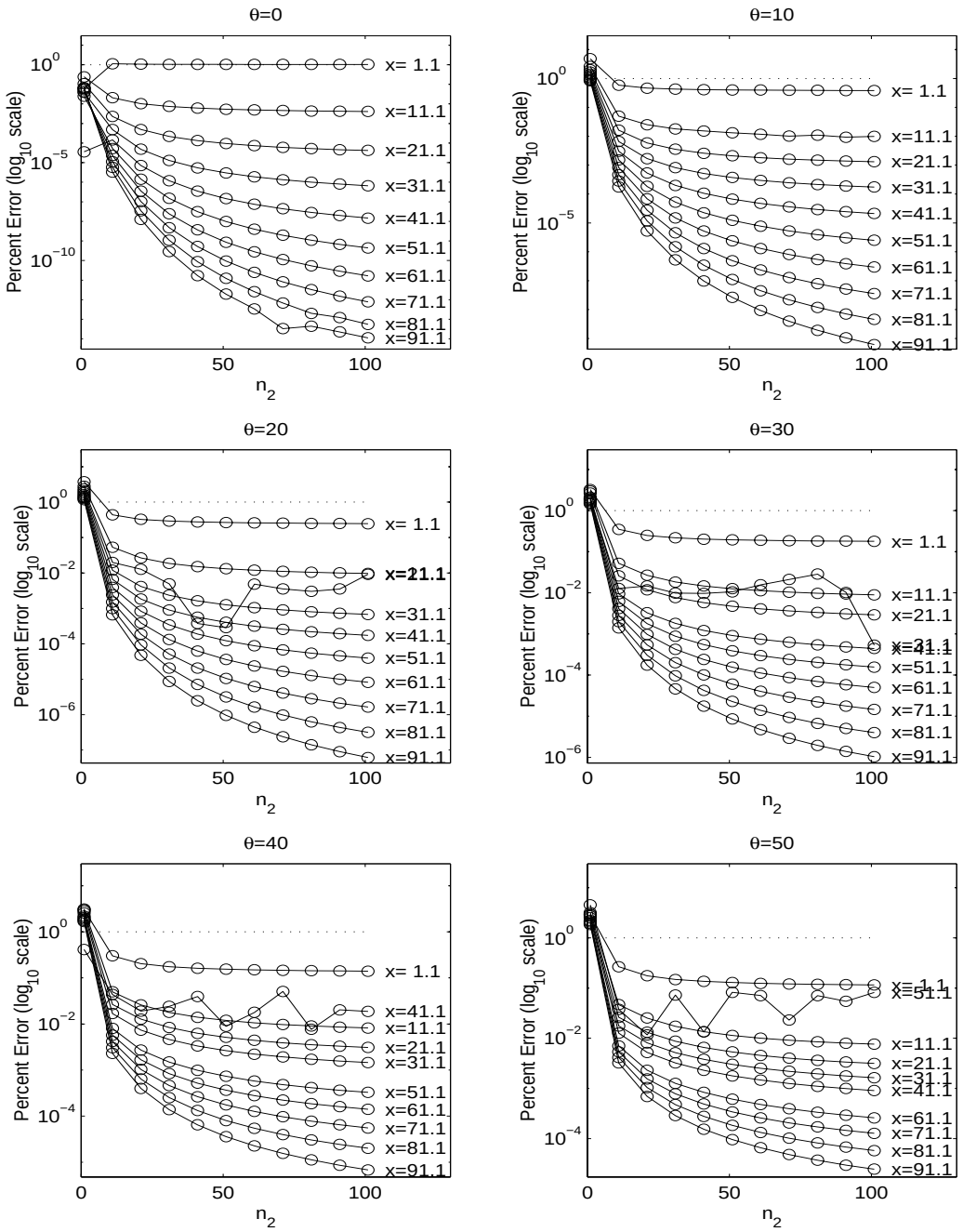

Figure 2: Percentage error of the saddlepoint method for the CDF of the singly noncentral F distribution with parameter $n_1 = 1$. Dashed line indicates 1% error.

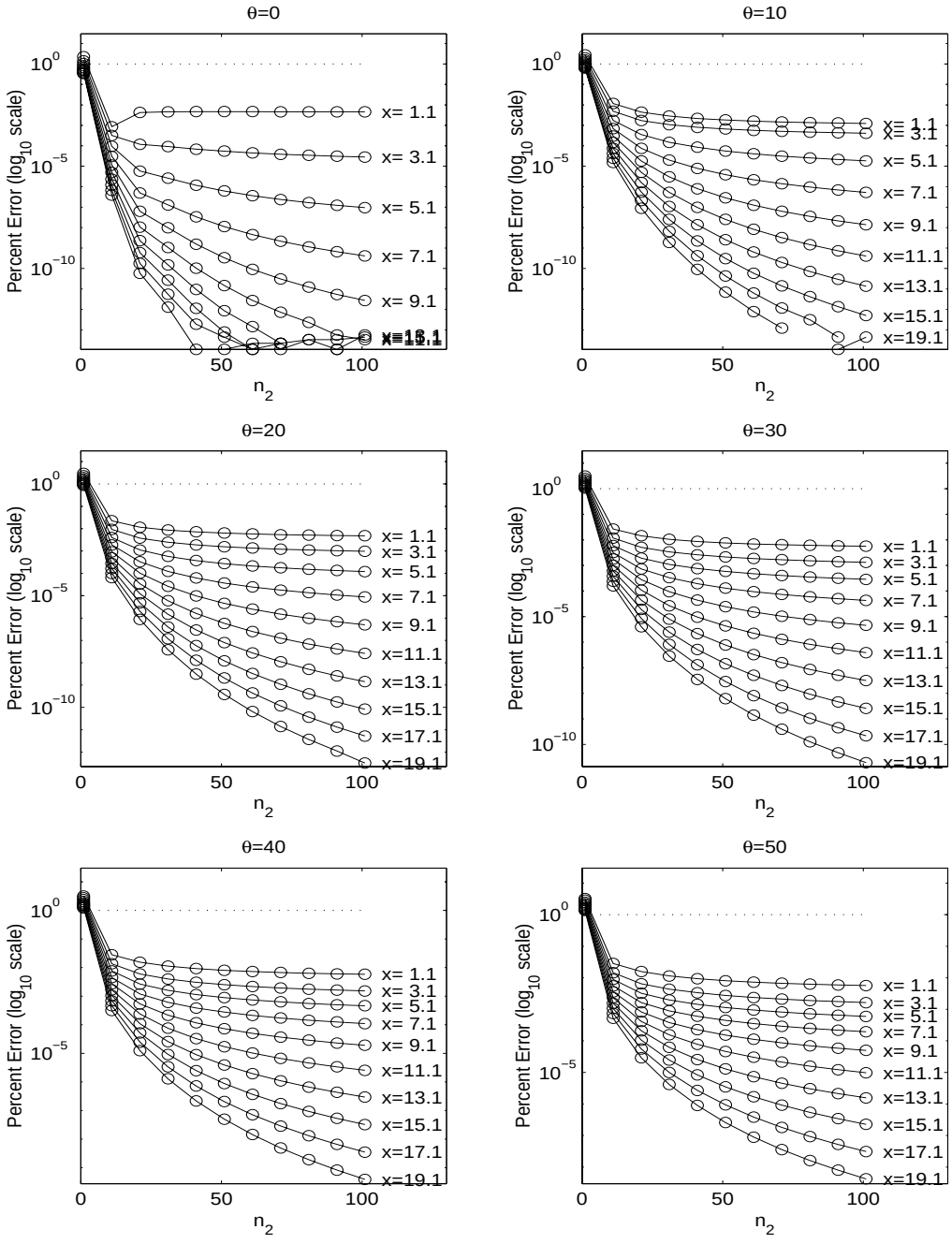

Figure 3: Same as Figure 2 but with $n_1 = 10$

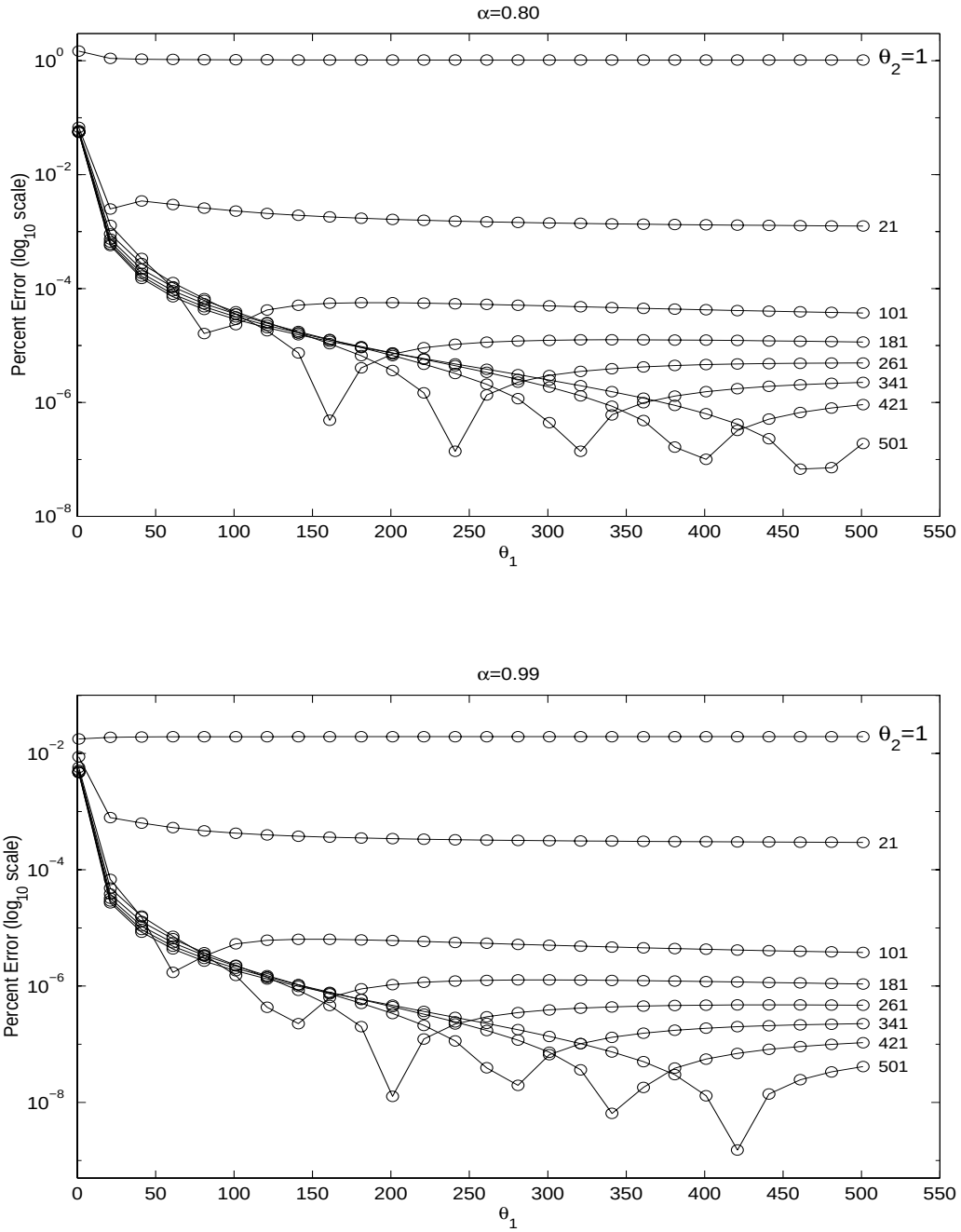

Figure 4: Accuracy of cdf saddlepoint approximation of $F_{n_1,n_2}(\theta_1,\theta_2)$ for fixed values $n_1 = n_2 = 1$ with f chosen such that $\Pr\left(F^{(2)} < f\right) = \alpha$

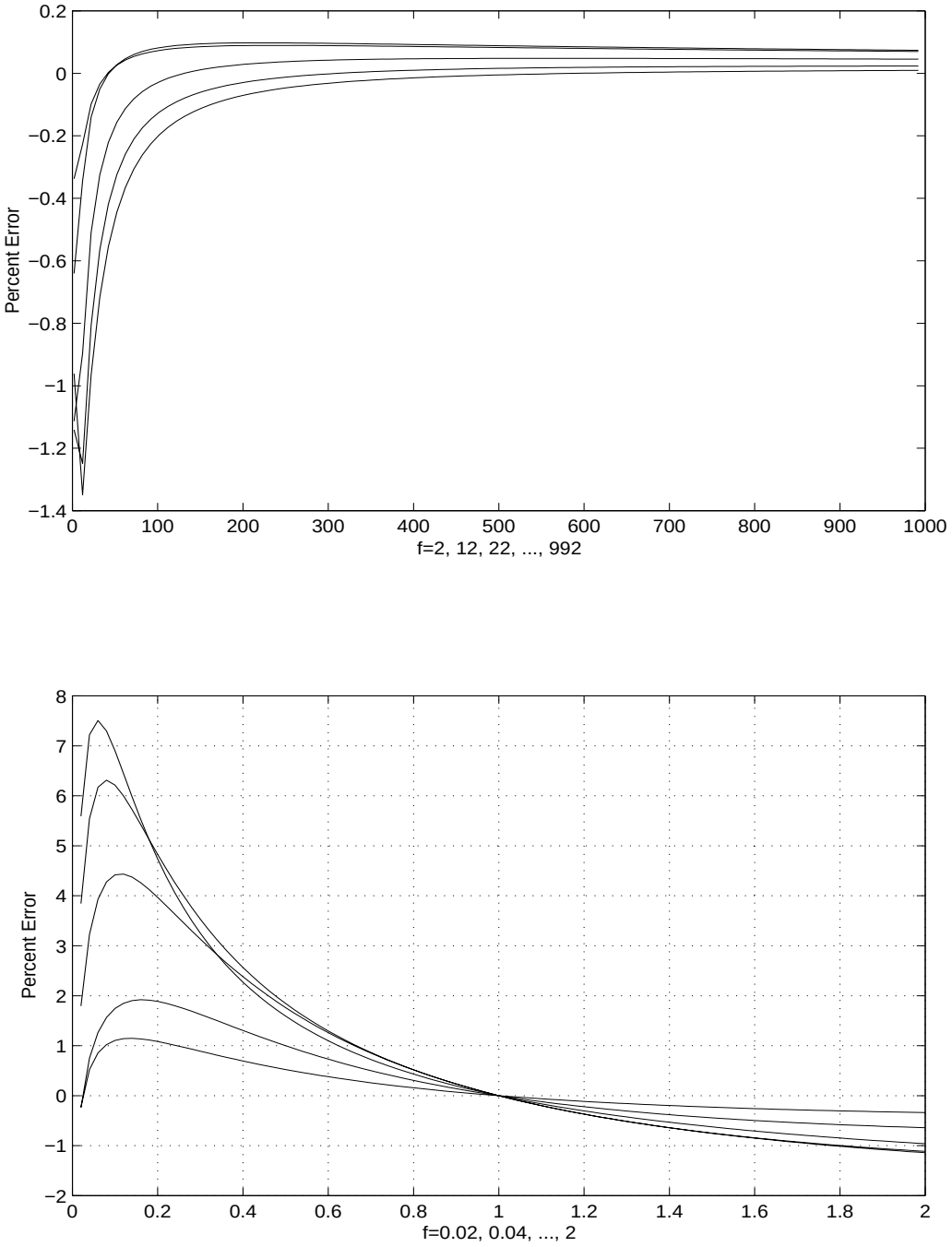

Figure 5: Saddlepoint accuracy for cdf of $F_{n_1,n_2}(\theta_1,\theta_2)$ with $n_1 = n_2 = 1$. Upper panel, from top to bottom, $\theta_1 = \theta_2 = 0, 0.3, 0.6, 0.9$ and 1.2. Lower panel, left side, ordering of θ_2 is reversed.

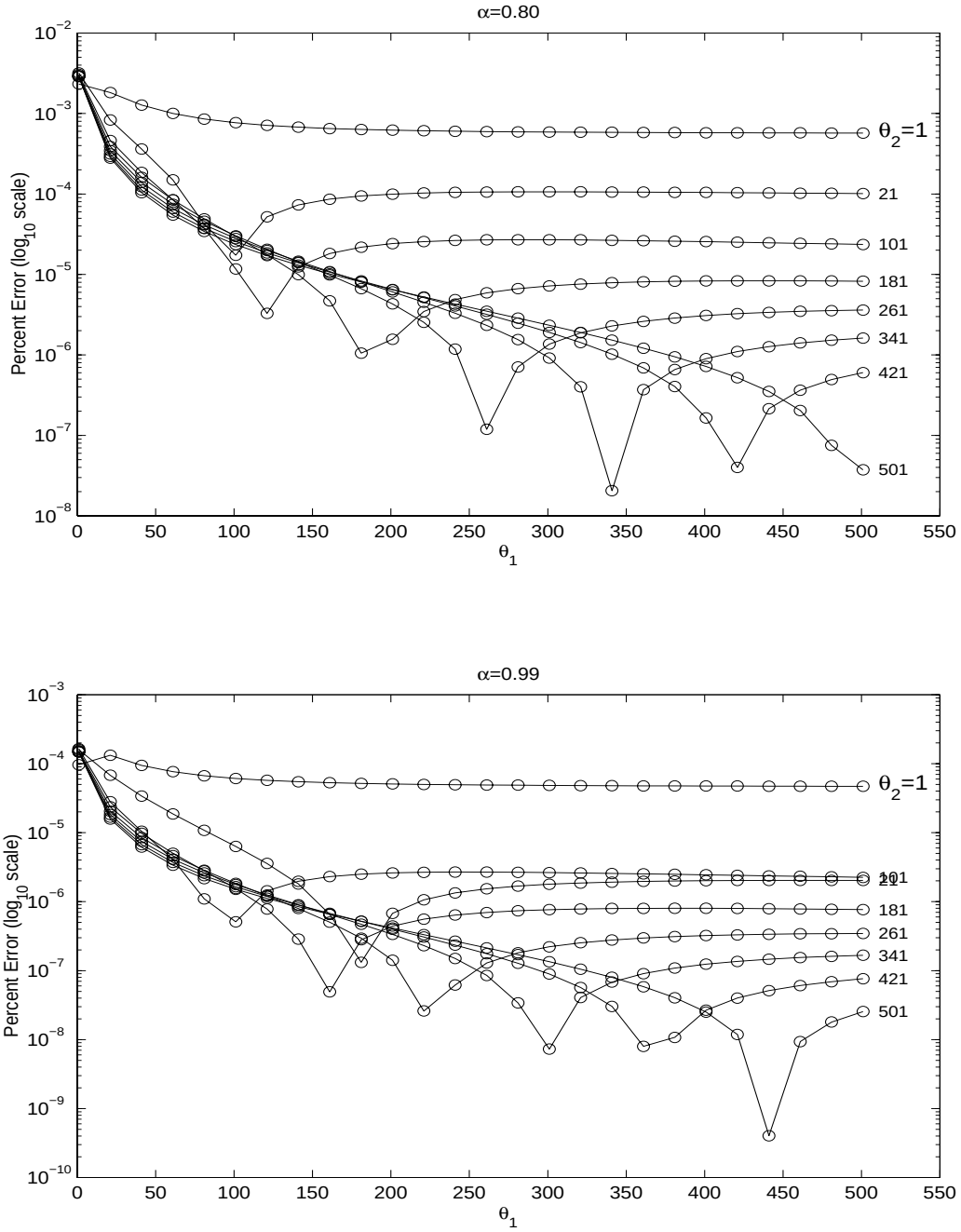

Figure 6: Similar to Figure 4 but with $n_1 = 10$ and $n_2 = 20$