The Annals of Statistics
2002, Vol. 30, No. 4, 1155–1177

LAPLACE APPROXIMATIONS FOR HYPERGEOMETRIC FUNCTIONS WITH MATRIX ARGUMENT[1]

BY RONALD W. BUTLER AND ANDREW T. A. WOOD

Colorado State University and University of Nottingham

In this paper we present Laplace approximations for two functions of matrix argument: the Type I confluent hypergeometric function and the Gauss hypergeometric function. Both of these functions play an important role in distribution theory in multivariate analysis, but from a practical point of view they have proved challenging, and they have acquired a reputation for being difficult to approximate. Appealing features of the approximations we present are: (i) they are fully explicit (and simple to evaluate in practice); and (ii) typically, they have excellent numerical accuracy. The excellent numerical accuracy is demonstrated in the calculation of noncentral moments of Wilks' Λ and the likelihood ratio statistic for testing block independence, and in the calculation of the CDF of the noncentral distribution of Wilks' Λ via a sequential saddlepoint approximation. Relative error properties of these approximations are also studied, and it is noted that the approximations have uniformly bounded relative errors in important cases.

1. Introduction. Laplace approximations are presented for two hypergeometric functions of matrix argument: the confluent hypergeometric ${}_1F_1(a; b; Z)$ and the Gauss hypergeometric ${}_2F_1(a, b; c; Z)$ where Z is a $p \times p$ symmetric matrix argument. The approximations are shown to be extremely accurate by comparison with simulated values in the statistical applications we consider.

These functions have power series expansions in terms of zonal polynomials as described in Muirhead [(1982), Chapter 7] or Mathai [(1993), Section 4.6]. Such series generally converge extremely slowly [Muirhead (1978), Section 1] and are very difficult and time consuming to compute even with current computing technology. Computational difficulty is cited in Muirhead (1975, 1978), Sections 1, 2, particularly for large p, large values of a, b and c and large and small eigenvalues of Z. Underflow and overflow difficulties are also cited which are problems we also encountered in our numerical simulations. None of these difficulties occurs with the explicit Laplace approximations. We provide evidence that the approximations are able to maintain high accuracy in all these settings.

Some discretion is required in the general use of Laplace approximations if they are to achieve their greatest accuracy. Specifically, a choice must be made

Received October 2000; revised August 2001.

[1]Supported by NSF Grants DMS-96-25396 and DMS-99-70785.

AMS 2000 subject classifications. Primary 62H10; secondary 62E17.

Key words and phrases. Confluent hypergeometric function, Gauss hypergeometric function, Laplace approximation, likelihood ratio test, matrix-argument hypergeometric function, saddlepoint approximation, sequential saddlepoint approximation, Wilks' Λ.

as to what portion of the integrand is used in the maximization. This issue is reviewed in Section 3. The issue takes on greater importance in settings involving higher dimensional integration, such as occur with these special functions, where the integration is over regions of dimension $p(p+1)/2$. The particular implementation of Laplace's method proposed here leads to a level of accuracy that is not achieved by other implementations of Laplace's method we have tried.

Guidance in this usage of Laplace approximations has been provided by a second approach to approximating ${}_1F_1$. This function is the moment generating function (MGF) for the null distribution of the Bartlett–Nanda–Pillai trace statistic in MANOVA as shown in Muirhead (1982), Section 10.6.3. An indirect approximation to this MGF has been provided in Butler, Huzurbazar and Booth (1992) using a sequential saddlepoint approach based upon a conditional distributional characterization of the test statistic. Their indirect approximation proved to be extremely accurate in so far as it allowed the second stage of saddlepoint approximation to proceed with extreme accuracy. In Section 9 we present a theorem which states that this sequential saddlepoint approximation to ${}_1F_1$ is analytically equivalent to a Laplace approximation applied by making a particular choice for the maximization factor. Because this particular approximation proved to be so accurate, we have used this agreement as partial motivation for our choice of the maximization factor to be used with Laplace's approximation for both ${}_1F_1$ and ${}_2F_1$, though alternative motivation is also given in Section 3.

The scalar and matrix cases of the ${}_1F_1$ approximation are discussed in Section 4 and ${}_2F_1$ is considered in Section 5. Applications of the Laplace approximations when arguments are complex and/or when the integral representation fails to hold are discussed in Section 6.

Numerical calculations are presented in Section 7. One application is concerned with computation of the noncentral moments of the likelihood ratio statistic in MANOVA, which can be expressed as a product of ${}_1F_1$ and an elementary function. A second application is concerned with noncentral moments of the likelihood ratio statistic for testing block independence, which can be expressed as a product of ${}_2F_1$ and an elementary function. Comparison with values obtained by simulation shows that our Laplace approximations provide extremely accurate approximations to these moments in both cases.

A number of statistical applications of our Laplace approximations are mentioned in Section 8. The most important of these applications involves the construction of a sequential saddlepoint approximation to the noncentral distribution of Wilks' Λ in MANOVA, and numerical results are provided in this case. When applied over a wide range of MANOVA settings, these power computations incur relative errors of less than 1% with even smaller typical errors.

We now briefly mention some other related work. The hypergeometric function ${}_2F_1(a, b; c; zI_p)$ of scalar matrix argument can be expressed as a Pfaffian of a matrix of scalar argument hypergeometric functions as shown in Gupta and Richards (1985). This relation also follows from the earlier Pfaffian expression

in Roy (1939) coupled with Sugiyama (1967) and the fact that the incomplete beta entries in Roy (1939) can be written in terms of hypergeometric function values. See Johnson and Kotz (1972), page 184, for details. More recently in Olver (1991, 1993) and Paris (1992), "exponentially improved" asymptotic expansions have been developed in the scalar argument case of ${}_1F_1$. Unfortunately, this interesting approach does not appear to be tractable in the matrix argument case which is our principal concern here.

We now summarize the principal novel contributions of this paper. First, the insights about how to implement Laplace's approximation for the class of problems considered in the paper (Section 3), resulting in successful Laplace approximations for ${}_1F_1$ and ${}_2F_1$ (Sections 4 and 5); the realization that these approximations perform well even in regions in which the integral representation (from which the approximations were derived) does not hold (Section 6); a substantial amount of numerical evidence that the approximations perform extremely well in practice (Sections 7 and 8); clarification of the relationship with an alternative approach (Section 9); and some new theoretical results on the relative error properties of the approximations (Section 10).

Finally, we note that the use of Laplace's approximation in classical multivariate analysis goes back a number of years; see, for instance, Glynn and Muirhead (1978), Muirhead (1978, 1982) and Srivastava and Carter (1980). However, these papers are concerned with hypergeometric functions of two matrix arguments, in contrast to the single matrix argument functions considered here. Moreover, these earlier papers do not address the question of numerical accuracy of the Laplace approximations they consider.

2. Review of ${}_1F_1$ and ${}_2F_1$. The classical Type I confluent hypergeometric function and Gauss hypergeometric function have well-known integral representations; see Abramowitz and Stegun (1972). In the case of ${}_1F_1$, this is given by

$$(1) \qquad {}_1F_1(a;b;z) = B(a,b-a)^{-1}\int_0^1 y^{a-1}(1-y)^{b-a-1}e^{zy}\,dy,$$

which is valid for all $z \in \mathbf{C}$ and all $a, b \in \mathbf{C}$ satisfying $\Re(a) > 0$ and $\Re(b-a) > 0$. In the above, $B(\cdot,\cdot)$ is the beta function, defined by $B(\alpha,\beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$, where $\Gamma(\cdot)$ is the gamma function.

In the case of ${}_2F_1$, the integral representation is given by

$$(2) \qquad {}_2F_1(a,b;c;z) = B(a,c-a)^{-1}\int_0^1 y^{a-1}(1-y)^{c-a-1}(1-zy)^{-b}\,dy,$$

which is valid whenever $\Re(z) < 1$, $\Re(a) > 0$ and $\Re(c-a) > 0$.

These, and other, classical functions have natural generalizations to the case in which $z \in \mathbf{C}$ is replaced by a symmetric $p \times p$ matrix Z with complex entries. It is a remarkable fact that these generalizations inherit many of the

important properties of their classical versions. We shall write these functions as ${}_1F_1(a;b;Z)$ and ${}_2F_1(a,b;c;Z)$, with Z understood to be a $p\times p$ matrix argument. These functions can be expressed as infinite power series in the scalar argument case, and infinite series of zonal polynomials in the matrix argument case; see Muirhead (1982), page 258. In this paper, however, we shall focus on integral representations of ${}_1F_1(a;b;Z)$ and ${}_2F_1(a,b;c;Z)$ which are natural generalizations of (1) and (2).

We first introduce some notation. Let (dY) denote Lebesgue measure on the space of $p\times p$ positive definite matrices. Given $p\times p$ matrices A and B, we say that $A>B$ $(A\geq B)$ if $A-B$ is positive (nonnegative) definite. Write $\mathrm{tr}(A)$ and $|A|$ for the trace and determinant, respectively, of a square matrix A and denote $\exp\{\mathrm{tr}(A)\}$ by $\mathrm{etr}(A)$. The $p\times p$ identity and zero matrix are denoted by I_p and 0_p, respectively.

For integer $p\geq 1$, the multivariate gamma and beta functions, $\Gamma_p(a)$ and $B_p(\alpha,\beta)$, are defined equivalently by

$$\Gamma_p(a)=\int_{Y>0_p}\mathrm{etr}(-Y)|Y|^{a-(p+1)/2}(dY)=\pi^{p(p-1)/4}\prod_{i=1}^{p}\Gamma\{a-(i-1)/2\}$$

and

$$B_p(\alpha,\beta)=\frac{\Gamma_p(\alpha)\Gamma_p(\beta)}{\Gamma_p(\alpha+\beta)}=\int_{0_p<Y<I_p}|Y|^{\alpha-(p+1)/2}|I_p-Y|^{\beta-(p+1)/2}(dY);$$

see, for example, Muirhead (1982). Note that when $p=1$ the classical gamma and beta functions are recovered.

The integral representation for ${}_1F_1(a;b;Z)$ is given by

$$\begin{aligned}{}_1F_1(a;b;Z)&=B_p(a,b-a)^{-1}\\&\quad\times\int_{0_p<Y<I_p}\mathrm{etr}(ZY)|Y|^{a-(p+1)/2}|I_p-Y|^{b-a-(p+1)/2}(dY)\end{aligned}\tag{3}$$

and is valid under the following conditions: $Z\in\mathbf{C}^{p\times p}$ is symmetric; $\Re(a)>(p-1)/2$; and $\Re(b-a)>(p-1)/2$.

The integral representation for ${}_2F_1(a,b;c;Z)$ is given by

$$\begin{aligned}&{}_2F_1(a,b;c;Z)\\&\quad=B_p(a,c-a)^{-1}\\&\qquad\times\int_{0_p<Y<I_p}|Y|^{a-(p+1)/2}|I_p-Y|^{c-a-(p+1)/2}|I_p-ZY|^{-b}(dY)\end{aligned}\tag{4}$$

and is valid under the following conditions: $Z\in\mathbf{C}^{p\times p}$ is symmetric and satisfies $\Re(Z)<I_p$; $\Re(a)>(p-1)/2$; and $\Re(c-a)>(p-1)/2$. See Muirhead (1982) for further details of (3) and (4).

Our principal purpose in this paper is to present Laplace approximations to ${}_1F_1$ and ${}_2F_1$ based on the integral representations (3) and (4).

We shall mainly be concerned with real Z here (the only exception being in Section 6), and to emphasize this point we will, from now on, write X for Z, on the understanding that X is a real symmetric matrix. The scalar quantities a, b and c may also be assumed real from now on (except in Section 6).

Finally, we recall a key invariance property of ${}_1F_1$ and ${}_2F_1$: for any symmetric $p \times p$ matrix X, ${}_1F_1(a; b; X)$ and ${}_2F_1(a, b; c; X)$ depend only on the eigenvalues of X, so that there is no loss of generality in assuming that X is a diagonal matrix.

3. Review of Laplace's approximation. Consider the integral

$$I = \int_{y \in D} h(y) e^{-\lambda g(y)}\, dy$$

where $D \subseteq \mathbf{R}^d$ is an open set and λ is a real parameter. If $g(y)$ has a unique minimum over the closure of D, and this minimum occurs at stationary point $\hat{y} \in D$ of $g(y)$, then Laplace's approximation to I is given by

$$\tilde{I} = (2\pi)^{d/2} \lambda^{-d/2} |g''(\hat{y})|^{-1/2} h(\hat{y}) e^{-\lambda g(\hat{y})} \tag{5}$$

where

$$g''(y) = \frac{\partial^2 g(y)}{\partial y \partial y^T}$$

is the Hessian of g.

Under mild conditions, the accuracy of the approximation is given by

$$I = \tilde{I}\{1 + O(\lambda^{-1})\} \qquad \text{as } \lambda \to \infty.$$

However, there are many examples which show that Laplace's approximation is often quite accurate even in subasymptotic situations (i.e., when λ is not large). In this paper, the concern will be with subasymptotic settings, so for simplicity we shall usually take $\lambda = 1$.

When implementing Laplace's approximation there are two important points to consider that can affect accuracy: the choice of $(g; h)$ representation, and the possibility of calibration. These are discussed in Sections 3.1 and 3.2.

3.1. *The* $(g; h)$ *representation.* In practice, we need to decide on a $(g; h)$ representation for the integrand, the point being that h does not play a role in the determination of $\hat{y}$. For example, if h is everywhere-positive, and we take $\lambda = 1$, the integrand $f(y)$ may be written in the equivalent forms

$$f(y) = h(y) e^{-g(y)} = h_0(y) e^{-g_0(y)}$$

where $h_0(y) = 1$ and $g_0(y) = g(y) - \log h(y)$, and we may approximate I by

$$\tilde{I}_0 = (2\pi)^{d/2} |g_0''(\hat{y}_0)|^{-1/2} h_0(\hat{y}_0) e^{-g_0(\hat{y}_0)}$$

instead of $\tilde{I}$ in (5), where $\hat{y}_0$ is the minimum of g_0 in D. Clearly $\tilde{I}$ and $\tilde{I}_0$ will be different in general, so we do need to give some consideration to the choice of $(g; h)$ representation for the integrand.

However, once we have chosen a $(g; h)$ representation for the integrand, Laplace's approximation is invariant with respect to 1–1 transformations of y in the following sense. Write $u = q(y)$, where $q : D \to \mathbf{R}^d$ is smooth and 1–1, and so has an inverse q^{-1} say. Now consider the transformed integral

$$I = \int_{u \in q(D)} h_1(u) e^{-g_1(u)}\, du$$

where

$$h_1(u) = h\{q^{-1}(u)\} \left| \frac{\partial y(u)}{\partial u^T} \right| \quad \text{and} \quad g_1(u) = g\{q^{-1}(u)\}.$$

Then by elementary calculus, I is invariant with respect to the change of variables. Interestingly, it turns out that Laplace's approximation $\tilde{I}$ is also invariant in this sense. See, for example, Efstathiou, Gutiérrez–Peña and Smith (1998) for a proof. For further discussion of Laplace and related saddlepoint approximations, see Jensen (1995) and Barndorff-Nielsen and Wood (1998).

An example directly relevant to present concerns is the beta integral

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)} = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\, dx = \int_0^1 h(x) e^{-g(x)}\, dx$$

for some choice of g and h. A family of possible choices is given by

$$h_\gamma(x) = x^{\gamma-1}(1-x)^{\gamma-1} \quad \text{and} \quad g_\gamma(x) = -(\alpha - \gamma)\log x - (\beta - \gamma)\log(1-x),$$

with $\gamma = 0$ and $\gamma = 1$ especially plausible contenders. We present four reasons why $\gamma = 0$ is to be preferred.

1. *The Stirling connection.* Taking $g(x)$ to be $g_\gamma(x)$ with $\gamma = 0$ yields a unique minimum at $\hat{x} = \alpha/(\alpha + \beta)$. An application of (5) with $p = 1$ and $\lambda = 1$ yields the approximation

$$\begin{aligned} \hat{B}(\alpha, \beta) &= (2\pi)^{1/2} \left[\frac{(\alpha+\beta)^2}{\alpha} + \frac{(\alpha+\beta)^2}{\beta} \right]^{-1/2} \left(\frac{\alpha}{\alpha+\beta} \right)^{\alpha-1} \left(\frac{\beta}{\alpha+\beta} \right)^{\beta-1} \\ &= \frac{\hat{\Gamma}(\alpha)\hat{\Gamma}(\beta)}{\hat{\Gamma}(\alpha+\beta)} \end{aligned}$$

where $\hat{\Gamma}(\alpha) = (2\pi)^{1/2}\alpha^{\alpha-(1/2)}e^{-\alpha}$ is Stirling's approximation to Γ. No other choice of γ leads to Stirling's approximation in this way.

2. *Normalizing transformation.* In view of the invariance property of Laplace's approximation mentioned above, taking $h(x) = h_\gamma(x)$ with $\gamma = 0$ is equivalent to applying Fisher's [(1990), Chapter 6] "normalizing" z-transformation $u = \log\{x/(1-x)\}$ in the original integral, and then applying Laplace with the new variable u, taking $h(u) \equiv 1$. Thus taking $\gamma = 0$ has a transparent interpretation, and this interpretation suggests that $\gamma = 0$ is a good choice.

3. *Convenience.* If we take $\gamma > 0$ then problems occur with Laplace's approximation if $\alpha \in (0, \gamma)$ and/or $\beta \in (0, \gamma)$, because then $g_\gamma(x)$ does not have a minimum in $(0, 1)$. This problem does not arise if we choose $\gamma \leq 0$; so from this point of view $\gamma = 0$ is preferable to $\gamma = 1$.

4. *Invariant distribution.* The choice $\gamma = 0$ can also be motivated by consideration of invariant distributions.

The above discussion provides motivation for taking $h(y)$ proportional to $y^{-1}(1-y)^{-1}$ in (1) and (2) [equivalently, (3) and (4) with $p = 1$]. In the matrix-argument case ($p > 1$) we recommend taking $h(Y)$ proportional to $|Y|^{-(p+1)/2}|I_p - Y|^{-(p+1)/2}$, a natural generalization. Justification of this choice rests principally on Theorem 1 below. Theorem 1 states that, with this choice of $h(Y)$, the implied Laplace approximation is identical to an approximation derived by saddlepoint methods which is known to be very accurate.

3.2. *Calibration.* Given the "raw" Laplace approximation in (5), it may be beneficial to employ some form of calibration. Consider, for example, ${}_1F_1(a; b; X)$. From the first definition of the multivariate beta function given in (2), it follows that ${}_1F_1(a; b; 0_p) = 1$. So, if we are interested in approximating ${}_1F_1(a; b; X)$ when X is not too different from 0_p, and we write ${}_1\tilde{F}_1(a; b; X)$ for the raw Laplace approximation to ${}_1F_1(a; b; X)$ obtained via (5), it would make sense to use an approximation ${}_1\hat{F}_1$ which is "calibrated at $X = 0_p$" as follows:

$$ {}_1\hat{F}_1(a; b; X) = {}_1\tilde{F}_1(a; b; X)/{}_1\tilde{F}_1(a; b; 0_p), \tag{6}$$

the point being that ${}_1\hat{F}_1(a; b; X) = {}_1F_1(a; b; X)$ when $X = 0_p$. The function ${}_2F_1(a, b; c; X)$ may be calibrated at $X = 0_p$ in the same manner. Other forms of calibration are possible but we shall only consider calibration at 0_p.

Calibration at 0_p also has subtler benefits: in particular, it leads to convenient and accurate approximations to ${}_1F_1$ and ${}_2F_1$ which maintain high accuracy outside the domain of each integral representation. See Section 6 for further discussion.

4. Laplace approximation to ${}_1F_1$.

We shall first consider the case in which $p = 1$, so that X is a real scalar quantity, x say. Then we shall consider the general case in which X is a real symmetric $p \times p$ matrix.

4.1. *The scalar case.* In view of the discussion in the previous section, we shall adopt the following $(g; h)$ representation of the integrand in (1): $h(y) = B(a, b-a)^{-1} y^{-1}(1-y)^{-1}$ and

$$g(y) = -\{a \log(y) + (b-a)\log(1-y) + xy\}.$$

We shall see later that, with this particular choice of $(g; h)$, the calibrated Laplace approximation agrees with a sequential saddlepoint approximation for ${}_1F_1$. As the first and second derivatives of g are given by

$$g'(y) = \frac{b-a}{1-y} - \frac{a}{y} - x \quad \text{and} \quad g''(y) = \frac{b-a}{(1-y)^2} + \frac{a}{y^2},$$

it is seen that g is minimized over $y \in [0, 1]$ at $y = \hat{y}$, say, where $\hat{y}$ is the appropriate solution of the quadratic equation

$$P(y; a, b, x) \equiv xy^2 - y(x-b) - a = 0. \tag{7}$$

It is straightforward to check that, for $0 < a < b$ and $x \in \mathbf{R}$, $P(y; a, b, x)$ has a unique solution $\hat{y} \in (0, 1)$ given by

$$\hat{y} = \frac{2a}{b - x + \sqrt{(x-b)^2 + 4ax}} \tag{8}$$

and that this solution is the value of y at which g is minimized. Therefore, using (5), we obtain the raw Laplace approximation

$${}_1\tilde{F}_1(a; b; x) = (2\pi)^{1/2} B(a, b-a)^{-1} j_{1,1}^{-1/2} \hat{y}^a (1-\hat{y})^{b-a} e^{x\hat{y}} \tag{9}$$

where

$$j_{1,1} \equiv j_{1,1}(a, b, x) = a(1-\hat{y})^2 + (b-a)\hat{y}^2. \tag{10}$$

The calibrated approximation ${}_1\hat{F}_1(a; b; x)$ is given by

$$\begin{aligned} {}_1\hat{F}_1(a; b; x) &= \frac{{}_1\tilde{F}_1(a; b; x)}{{}_1\tilde{F}_1(a; b; 0)} \\ &= b^{b-1/2} r_{1,1}^{-1/2} \left(\frac{\hat{y}}{a}\right)^a \left(\frac{1-\hat{y}}{b-a}\right)^{b-a} e^{x\hat{y}} \end{aligned} \tag{11}$$

where

$$r_{1,1} = \frac{\hat{y}^2}{a} + \frac{(1-\hat{y})^2}{b-a}. \tag{12}$$

4.2. *The matrix case.* In view of the invariance property mentioned at the end of Section 2, we may without loss of generality take $X = \operatorname{diag}\{x_1, \ldots, x_p\}$. Motivated by the scalar case $p = 1$ and our intent to have the calibrated Laplace approximation agree with the sequential saddlepoint approximation considered in Theorem 1, we adopt the following $(g; h)$ representation for the integrand in (3):

$$g(Y) = -\operatorname{tr}(XY) - a \log|Y| - (b-a)\log|I_p - Y| \tag{13}$$

and

$$h(Y) = B_p(a, b-a)^{-1}|Y|^{-(p+1)/2}|I_p - Y|^{-(p+1)/2}. \tag{14}$$

Using (5) again, it may be shown that the raw Laplace approximation to ${}_1F_1(a; b; X)$ is given by

$${}_1\tilde{F}_1(a; b; X) = 2^{p/2}\pi^{p(p+1)/4}B_p(a, b-a)^{-1}J_{1,1}^{-1/2}\prod_{i=1}^{p}\{\hat{y}_i^a(1-\hat{y}_i)^{b-a}e^{x_i\hat{y}_i}\} \tag{15}$$

where, for $i = 1, \ldots, p$, $\hat{y}_i$ is the solution to $P(y; a, b, x_i) = 0$ given by (8), with P given in (7); and

$$J_{1,1} = \prod_{i=1}^{p}\prod_{j=i}^{p}\{a(1-\hat{y}_i)(1-\hat{y}_j) + (b-a)\hat{y}_i\hat{y}_j\}. \tag{16}$$

Note that (15) and (16) reduce to (9) and (10) when $p = 1$.

The calibrated approximation ${}_1\hat{F}_1(a; b; X)$ is given by

$$\begin{aligned} {}_1\hat{F}_1(a; b; X) &= \frac{{}_1\tilde{F}_1(a; b; X)}{{}_1\tilde{F}_1(a, b; 0_p)} \\ &= b^{bp-p(p+1)/4}R_{1,1}^{-1/2}\prod_{i=1}^{p}\left\{\left(\frac{\hat{y}_i}{a}\right)^a\left(\frac{1-\hat{y}_i}{b-a}\right)^{b-a}e^{x_i\hat{y}_i}\right\} \end{aligned} \tag{17}$$

where

$$R_{1,1} = \prod_{i=1}^{p}\prod_{j=i}^{p}\left\{\frac{\hat{y}_i\hat{y}_j}{a} + \frac{(1-\hat{y}_i)(1-\hat{y}_j)}{b-a}\right\}. \tag{18}$$

The above formulas are derived in Butler and Wood (2000). As noted in Butler and Wood (2000), certain important properties of ${}_1F_1$, such as the Kummer relation given in Muirhead [(1982), Theorem 7.4.3] are inherited by the approximations ${}_1\tilde{F}_1$ and ${}_1\hat{F}_1$.

5. Laplace approximation to ${}_2F_1$. As in the previous section, we shall first consider the scalar case and then proceed to the matrix case.

5.1. *The scalar case.* We use a similar $(g; h)$ representation of the integrand in (1), with $h(y)$ defined as in Section 4.1 and

$$g(y) = -\{a\log(y) + (c-a)\log(1-y) - b\log(1-xy)\}.$$

The first and second derivatives of g are given by

$$g'(y) = \frac{c-a}{1-y} - \frac{a}{y} - \frac{bx}{1-xy} \quad \text{and} \quad g''(y) = \frac{c-a}{(1-y)^2} + \frac{a}{y^2} - \frac{bx^2}{(1-xy)^2}.$$

If we set $g'(y) = 0$, the appropriate solution of the resulting quadratic equation is

$$\hat{y} = \frac{2a}{\sqrt{\tau^2 - 4ax(c-b)} - \tau} \tag{19}$$

where $\tau = x(b-a) - c$.

It can be shown that, provided $0 < a < c$, $b \geq 0$ and $x \in [0, 1)$, $\hat{y}$ yields the unique minimum of $g(y)$ on $[0, 1]$. Moreover, given that $0 < a < c$, there is no effective restriction in taking $b \geq 0$ and $x \in [0, 1)$ because the Euler relations [see (15.3.3)–(15.3.5) in Abramowitz and Stegun (1972)]

$$ {}_2F_1(a, b; c; x) = (1-x)^{-b}\,{}_2F_1\left(c-a, b; c; -\frac{x}{1-x}\right) \tag{20}$$

$$ = (1-x)^{c-a-b}\,{}_2F_1(c-a, c-b; c; x) \tag{21}$$

allow us to extend consideration to $x < 0$ [using (20)] and/or $b < 0$ [using (21)].

It follows that the raw Laplace approximation ${}_2\tilde{F}_1(a, b; c; x)$ is given by

$$ {}_2\tilde{F}_1(a, b; c; x) = (2\pi)^{1/2} B(a, c-a)^{-1} j_{2,1}^{-1/2} \hat{y}^a (1-\hat{y})^{c-a} (1-x\hat{y})^{-b} \tag{22}$$

where $j_{2,1} \equiv j_{2,1}(a, b, c, x)$ is given by

$$j_{2,1} = a(1-\hat{y})^2 + (c-a)\hat{y}^2 - bx^2\hat{y}^2(1-\hat{y})^2/(1-x\hat{y})^2. \tag{23}$$

The calibrated approximation ${}_2\hat{F}_1(a, b; c; x)$ is given by

$$\begin{aligned} {}_2\hat{F}_1(a, b; c; x) &= \frac{{}_2\tilde{F}_1(a, b; c; x)}{{}_2\tilde{F}_1(a, b; c; 0)} \\ &= c^{c-1/2} r_{2,1}^{-1/2} \left(\frac{\hat{y}}{a}\right)^a \left(\frac{1-\hat{y}}{c-a}\right)^{c-a} (1-x\hat{y})^{-b} \end{aligned} \tag{24}$$

where

$$r_{2,1} = \frac{\hat{y}^2}{a} + \frac{(1-\hat{y})^2}{c-a} - \frac{bx^2}{(1-x\hat{y})^2}\frac{\hat{y}^2}{a}\frac{(1-\hat{y})^2}{c-a}. \tag{25}$$

5.2. *The matrix case.* Given that X is symmetric then, as noted at the end of Section 2, we may without loss of generality take $X = \operatorname{diag}\{x_1, \ldots, x_p\}$. Following from the ${}_1F_1$ approximation, we make use of the following $(g; h)$ representation:

$$g(Y) = -a \log|Y| - (c-a)\log|I_p - Y| + b\log|I_p - XY| \tag{26}$$

and

$$h(Y) = B_p(a, c-a)^{-1}|Y|^{-(p+1)/2}|I_p - Y|^{-(p+1)/2}. \tag{27}$$

Using (5) again, it may be shown that the raw Laplace approximation to ${}_2F_1(a, b; c; X)$ is given by

$${}_2\tilde{F}_1(a, b; c; X) = \frac{2^{p/2}\pi^{p(p+1)/4}}{B_p(a, c-a)} J_{2,1}^{-1/2} \prod_{i=1}^{p} \{\hat{y}_i^a (1-\hat{y}_i)^{c-a}(1 - x_i\hat{y}_i)^{-b}\} \tag{28}$$

where $\hat{y}_i$ is defined by (19) with $x = x_i$, $J_{2,1} \equiv J_{2,1}(a, b, c, X)$ is given by

$$J_{2,1} = \prod_{i=1}^{p}\prod_{j=i}^{p} \{a(1-\hat{y}_i)(1-\hat{y}_j) + (c-a)\hat{y}_i\hat{y}_j - bL_iL_j\} \tag{29}$$

and

$$L_i = x_i\hat{y}_i(1-\hat{y}_i)/(1 - x_i\hat{y}_i). \tag{30}$$

The calibrated approximation ${}_2\hat{F}_1(a, b; c; X)$ is given by

$$\begin{aligned} {}_2\hat{F}_1(a, b; c; X) &= {}_2\tilde{F}_1(a, b; c; X)/{}_2\tilde{F}_1(a, b; c; 0_p) \\ &= c^{cp - p(p+1)/4} R_{2,1}^{-1/2} \prod_{i=1}^{p} \left\{ \left(\frac{\hat{y}_i}{a}\right)^a \left(\frac{1-\hat{y}_i}{c-a}\right)^{c-a} (1 - x_i\hat{y}_i)^{-b} \right\} \end{aligned} \tag{31}$$

where

$$R_{2,1} = \prod_{i=1}^{p}\prod_{j=i}^{p} \left\{ \frac{\hat{y}_i\hat{y}_j}{a} + \frac{(1-\hat{y}_i)(1-\hat{y}_j)}{c-a} - \frac{bx_ix_j\hat{y}_i\hat{y}_j(1-\hat{y}_i)(1-\hat{y}_j)}{(1-x_i\hat{y}_i)(1-x_j\hat{y}_j)a(c-a)} \right\}. \tag{32}$$

Note that when $p = 1$, (28), (29) and (30) reduce to (22) and (23). These formulas are derived in Butler and Wood (2000) where it is also noted that certain important properties of ${}_2F_1$, such as the Euler relations given in Muirhead (1982), Theorem 7.4.3, are shared by the approximations ${}_2\tilde{F}_1$ and ${}_2\hat{F}_1$.

6. Extended interpretation of the approximations. In Sections 4 and 5 we presented Laplace approximations to ${}_1F_1$ and ${}_2F_1$ using the integral representations (1) and (2) in the scalar argument case, and (3) and (4) in the matrix argument case. These derivations assumed that all arguments are real. In this section we discuss these approximations when some arguments are complex and/or the integral representations do not hold. The discussion of complex arguments in Butler and Wood (2000) provides additional insight into the accuracy achieved with our particular $(g; h)$ representation using Laplace's approximation.

6.1. *Complex arguments.* In some statistical applications we may wish to evaluate ${}_1F_1$ or ${}_2F_1$ when some arguments are complex. For example, the characteristic functions of the multivariate test statistics considered in Sections 8.1–8.3 depend on ${}_1F_1$ and ${}_2F_1$ evaluated at complex arguments. With such arguments, the integral representations of these functions must now be regarded as contour integrals to which Laplace's approximation is no longer applicable. As contour integrals however, the method of steepest descents becomes applicable [see Bleistein and Handelsman (1975) for a description of this method]. The main points to note with this method are that: (1) $\hat{y}$-values (8) and (19) are the appropriate complex-valued saddlepoints if we just substitute in our complex arguments; and (2) the Laplace approximation, also evaluated at our complex arguments, is the leading and dominant term in the expansion resulting from the method of steepest descents. Further discussion and numerical results are given in Butler and Wood (2000).

6.2. *Integral representation not valid.* In some statistical applications involving ${}_1F_1$ and ${}_2F_1$ (see, e.g., Section 8) it is necessary to approximate these functions when the integral representations do not hold. In the case of ${}_1F_1$, the integral representation (3) is not valid if $a < (p-1)/2$ or $b - a < (p-1)/2$. However, it can be seen after careful study of the formulae in Section 4 that the calibrated Laplace approximation ${}_1\hat{F}_1$ is well behaved as a passes through 0 or b, in that the apparent singularities turn out to be removable. This is very important from a practical point of view. Presumably, the Laplace approximation is able to retain high accuracy when the integral representation does not hold because, through analytic continuation, ${}_1\hat{F}_1$ continues to "track" ${}_1F_1$ outside the domain of the integral representation. It is difficult to make this argument precise, but this does seem a plausible explanation for the excellent numerical accuracy in Section 8 when the integral representation does not hold.

The story is essentially the same with ${}_2F_1$ (except that, in this case, c plays the role of b).

7. Numerical accuracy. For certain values of a, b and c, the accuracy of ${}_1\hat{F}_1(a; b; X)$ and ${}_2\hat{F}_1(a, b; c; X)$ may be checked using simulation as shown below. The simulations demonstrate extremely high accuracy for the Laplace approximations even in the difficult setting in which there are $p = 32$ dimensions.

7.1. *Accuracy of* ${}_1\hat{F}_1(s; b+s; X)$. Function ${}_1F_1(s; b+s; X)$ appears as a factor in the sth noncentral moment of Wilks' Λ. Thus simulated estimates of this function may be computed by averaging 10^6 noncentral Λ^s-values in order to check the accuracy of our Laplace approximation.

Suppose E is a $p \times p$ matrix error sums of squares with a central Wishart$_p(n, \Sigma)$ distribution. Let $T = V^T V$ be the treatment sum of squares with a noncentral Wishart$_p(m, \Omega, \Sigma)$ distribution with m degrees of freedom

and noncentrality matrix $\Omega = \Sigma^{-1} M^T M$. This results when V is $(m \times p)$ with a Normal$_{m\times p}(M, \Sigma)$ distribution, with mean $E(V) = M$ and columns that are independent with common covariance Σ. The likelihood ratio statistic is $\Lambda = W^{n/2}$ where $W = |E|/|T + E|$. The noncentral moments of W are specified in Theorem 10.5.1 of Muirhead (1982) as

$$E(W^s) = \frac{\Gamma_p(n/2+s)\Gamma_p((n+m)/2)}{\Gamma_p(n/2)\Gamma_p((n+m)/2+s)}\, {}_1F_1\left(s; \frac{n+m}{2}+s; -\frac{1}{2}\Omega\right). \tag{33}$$

We approximate these moments by replacing ${}_1F_1$ with ${}_1\hat{F}_1$, where ${}_1\hat{F}_1$ is given in (17).

Table 1 displays simulation-based and ${}_1\hat{F}_1$-based approximations to the first four moments of W for various choices of p, n, m and noncentrality matrix Ω. There is remarkable agreement in Table 1 between the simulated values of the

TABLE 1

The first four moments about zero for the noncentral distribution of the likelihood ratio test in MANOVA ("*Sim.*" *denotes simulated values based upon averaging* 10^6 *values,* ${}_1\hat{F}_1$ *denotes the approximation, and the percentage relative error is listed as* "*% rel. err*")

		Moment			
(p, n, m)		**1**	**2**	**3**	**4**
(2, 10, 3)		$\Omega = \text{diag}\{\frac{1}{2}, 1\}$			
	Sim. 10^6	0.52304	0.30478	0.19201	0.12829
	${}_1\hat{F}_1$	0.52303	0.30480	0.19205	0.12835
	% rel. err.	$-0.0^2223\%$	$0.0^2793\%$	0.0244%	0.0453%
(5, 20, 5)		$\Omega = \text{diag}\{\frac{1}{4}\,(\frac{1}{4})\,1, 1\frac{1}{2}\}$			
	Sim. 10^6	0.25194	0.072381	0.023191	0.0^281460
	${}_1\hat{F}_1$	0.25216	0.072492	0.023234	0.0^281616
	% rel. err.	0.0874%	0.152%	0.188%	−0.191%
(8, 40, 7)		$\Omega = \text{diag}\{0, \frac{1}{4}, \frac{1}{2}, 1\,(\frac{1}{2})\,3\}$			
	Sim. 10^6	0.19717	0.042299	0.0^297948	0.0^224312
	${}_1\hat{F}_1$	0.19715	0.042283	0.0^297861	0.0^224271
	% rel. err.	$-0.0^2688\%$	−0.0358%	−0.0889%	−0.166%
(16, 40, 14)		$\Omega_{16} = \text{diag}\{2\,(0), \frac{1}{4}, \frac{1}{2}\,(\frac{1}{2})\,3, 3\frac{1}{4}\,(\frac{1}{4})\,4\frac{3}{4}\}$			
	Sim. 10^6	0.0^215154	0.0^532231	0.0^893082	$0.0^{10}35496$
	${}_1\hat{F}_1$	0.0^215158	0.0^532242	0.0^893101	$0.0^{10}35468$
	% rel. err.	0.0230%	0.0356%	0.0204%	−0.0785%
(32, 60, 28)		$\Omega_{32} = I_2 \otimes \Omega_{16}$			
	Sim. 10^6	0.0^744805	$0.0^{14}37381$	$0.0^{21}55062$	$0.0^{27}13461$
	${}_1\hat{F}_1$	0.0^744761	$0.0^{14}37267$	$0.0^{21}55161$	$0.0^{27}13942$
	% rel. err.	−0.0991%	−0.303%	0.179%	3.57%

moments and those based on ${}_1\hat{F}_1$. The percentage relative errors in estimating ${}_1F_1$ are exactly those for the moment estimates and are listed in Table 1. Note that very high accuracy is maintained even when the dimension is quite large (see $p = 16$, 32). In many of the cases considered in Table 1 [in fact, all those cases in which $s \le (p-1)/2$], ${}_1F_1$ does not have the integral representation (3), because the requirement $a > (p-1)/2$ does not hold; yet the very high level of accuracy is maintained in these cases.

A zonal polynomial expansion for the ${}_1F_1$ term in (33) with large Ω has been derived in Constantine and Muirhead [(1976), Theorem 3.2] and Muirhead [(1978), equation 7.2 and below]. The leading term of this expansion, when used to approximate the mean for the $p = 5$ example of Table 1, gives the value 1.323×10^7 which is quite far from the simulated value 0.25194.

7.2. *Accuracy of* ${}_2\hat{F}_1(a, a; a+s; X)$. The function ${}_2F_1$ is a factor in the noncentral moments of the likelihood ratio test for block independence and we may check its accuracy in the same manner.

Suppose blocks of variables with dimensions p_1 and p_2 where $p_1 + p_2 = p$. Let A be the matrix of sample covariances with a Wishart$_p$ (n, Σ) distribution. Specify A in block form as

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

where A_{ij} is $p_i \times p_j$. The likelihood ratio test for block independence rejects for small values of $W = |A|/(|A_{11}||A_{22}|)$. The noncentral moments of W are specified in Theorem 11.2.6 of Muirhead (1982) as

$$E(W^s) = \frac{\Gamma_{p_1}(n/2)\Gamma_{p_1}((n-p_2)/2+s)}{\Gamma_{p_1}(n/2+s)\Gamma_{p_1}((n-p_2)/2)} \times |I_{p_1} - P^2|^{n/2}\, {}_2F_1\left(\frac{n}{2}, \frac{n}{2}; \frac{n}{2}+s; P^2\right), \tag{34}$$

where $P = \operatorname{diag}\{\rho_1, \ldots, \rho_{p_1}\}$ contains the population canonical correlations or the eigenvalues of $\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ where $\{\Sigma_{ij}\}$ form the conformable subblocks of Σ. We approximate these moments by substituting ${}_2\hat{F}_1$ into (34). Such approximate moments are given in Table 2 and are compared with empirical moments based on simulating 10^6 values of W. There is remarkable agreement between the explicit analytical approximations using ${}_2\hat{F}_1$ and those obtained through long and time-consuming simulations.

Theorem 1 in Glynn and Muirhead (1978) provides a different Laplace approximation to ${}_2F_1(\frac{n}{2}, \frac{n}{2}; \frac{n}{2}+s; P^2)$ based upon representing the ${}_2F_1$ function as an integral over a different space of considerably higher dimension. Numerical computation of this approximation for the mean of the $p_1 = 5$ example gives a mean of 4.061×10^{-8} which is not close to the simulated value of 0.00743.

TABLE 2
The first four moments about zero for the noncentral distribution of the likelihood ratio test for block independence

(p_1, p_2, n)		Moment 1	2	3	4
(2, 3, 10)			$P = \text{diag}\{0.3, 0.8\}$		
	Sim. 10^6	0.17602	0.045659	0.015584	$0.0^2$64957
	${}_2\hat{F}_1$	0.17606	0.045722	0.015628	$0.0^2$65209
	% rel. err.	0.0234%	0.140%	0.282%	0.387%
(5, 7, 20)			$P = \text{diag}\{0.4\,(0.1)\,0.8\}$		
	Sim. 10^6	$0.0^2$74312	$0.0^3$10603	$0.0^5$255842	$0.0^7$95614
	${}_2\hat{F}_1$	$0.0^2$74248	$0.0^3$10588	$0.0^5$25496	$0.0^7$94246
	% rel. err.	−0.0869%	−0.143%	−0.344%	−1.43%
(10, 13, 40)			$P = \text{diag}\{0.1\,(0.1)\,0.9, 0.95\}$		
	Sim. 10^6	$0.0^4$14332	$0.0^9$40854	0.0^{13}21700	0.0^{17}20176
	${}_2\hat{F}_1$	$0.0^4$14328	$0.0^9$40727	0.0^{13}21577	0.0^{17}20214
	% rel. err.	−0.0284%	−0.312%	−0.566%	0.187%
(25, 28, 70)		$P = \text{diag}\{2\,(0.1), 2\,(0.2), 3\,(0.3), \ldots, 3\,(0.7), 2\,(0.8), 2\,(0.9), 2\,(0.95)\}$			
	Sim. 10^6	0.0^{13}13552	0.0^{27}72427	0.0^{39}13915	0.0^{52}73407
	${}_2\hat{F}_1$	0.0^{13}13552	0.0^{27}72142	0.0^{39}13553	0.0^{52}81928
	% rel. err.	$-0.0^3$156%	−0.393%	−2.60%	11.6%

We suggest two reasons for this inaccuracy. First the integral representation for ${}_2F_1$ is over a much larger space. Secondly, their approximation is undefined at $P = \text{diag}\{0, \ldots, 0\}$ and therefore cannot be calibrated, an adjustment that leads to considerable improvement in accuracy.

8. Statistical applications. There are many statistical inference applications in which these Laplace approximations could serve as surrogates for the true values of ${}_1F_1$ and ${}_2F_1$. A summary of such applications is given below. Prominent among these applications are the proposed sequential saddlepoint applications for the noncentral distributions of the various test statistics listed in item 1. In each of these 5 settings, ${}_1F_1$ or ${}_2F_1$ is a factor in the Mellin transform of the test statistic. Replacement of the true function by its Laplace approximation provides an approximate Mellin transform that is subsequently inverted using the Lugannani and Rice (1980) saddlepoint procedure. The net result is an extremely accurate CDF approximation for each of the noncentral distributions that is listed. Application to the noncentral CDF of Wilks' Λ is shown in Section 8.1.

The functions ${}_1F_1$ and ${}_2F_1$ are needed to determine the following transforms, densities and CDFs:

1. The Mellin transform for the following distributions. Sequential saddlepoint approximations may be used to approximate their CDFs using ${}_1\hat{F}_1$ for (a)–(c) and ${}_2\hat{F}_1$ for (d) and (e).
   (a) The noncentral distribution of the likelihood ratio test or Wilks' Lambda in MANOVA [Muirhead (1982), Theorem 10.5.1, and Constantine (1963)]. See Section 8.1 for numerical examples.
   (b) The distribution of the generalized variance for a noncentral Wishart [Muirhead (1982), Theorem 10.3.7; Herz (1955); and Constantine (1963)].
   (c) The null distribution of Bartlett–Nanda–Pillai trace statistic in MANOVA [Muirhead (1982), Section 10.6.3, and James (1964)]. Numerical work is given in Butler, Huzurbazar and Booth (1992).
   (d) The noncentral distribution of the likelihood ratio test statistic for block independence with two block groups [Muirhead (1982), Theorem 11.2.6, and Sugiura and Fujikoshi (1969)].
   (e) The noncentral distribution of the likelihood ratio test statistic for equality of two covariance matrices [Muirhead (1982), Theorem 8.2.11, and Sugiura (1969)].
2. The CDF of the largest eigenvalue of a Wishart [Muirhead (1982) Theorem 9.7.1, and Constantine (1963)].
3. The CDF of the largest eigenvalue of a matrix Beta [Mathai (1993), page 223, and Constantine (1963)].
4. The density for the $p \times 1$ instrumental variable regression estimator used in econometric simultaneous equation models with $p + 1$ endogenous variables [Phillips (1983)].
5. Applications to shape analysis including the Bingham distribution [Bingham, Chang and Richards (1992)].
6. Bayesian posterior analysis for distributions on circles, cylinders, and hypercylinders [Bagchi and Kadane (1991)].
7. The matrix F density [Muirhead (1982), Theorem 10.4.1, and James (1964)] and the Studentized Wishart density [Muirhead (1982), Theorem 10.4.4, and James (1964)].
8. Kernel density estimation on the Stiefel manifold [Chikuse (1998)].

8.1. *Sequential saddlepoint approximation to noncentral Wilks'* Λ. Expression (33) determines the MGF for the noncentral distribution of $\log W$ in terms of ${}_1F_1$. The substitution of ${}_1\hat{F}_1$ as its surrogate, followed by saddlepoint inversion using the Lugannani and Rice (1980) approximation, produces a sequential saddlepoint approximation, as in Fraser, Reid and Wong (1991), for the nonnull distribution function which we compute below. In our implementation, derivatives are calculated numerically.

Table 3 displays sequential saddlepoint (seq. sad.) percentages whose accuracy is measured in terms of how close the entries are to the true percentages listed in the top row. The approximations were computed for dimensions $p = 2, \ldots, 64$

TABLE 3

Sequential saddlepoint approximations (Seq. sad.), the $O(n^{-3})$ expansions of Sugiura and Fujikoshi (1969) *and the $O(n^{-3/2})$ expansions of Sugiura* (1973) *for percentages of the noncentral CDF of Wilks'* $\log W$ *in the general linear MANOVA hypothesis. (The table entries are the three CDF approximations evaluated at the simulated empirical percentiles listed in Table* 4*. Accuracy may be judged by comparing the table entries to the true percentile levels given in the top row.)*

(p, n, m)	1	5	10	30	50	70	90	95	99
$(2, 10, 3)$				$\Omega = \text{diag}\{\frac{1}{2}, 1\}$					
Seq. sad.	1.011	5.021	9.987	30.07	50.06	70.00	89.99	94.99	98.99
$O(n^{-3})$	1.332	5.393	10.32	30.10	49.99	69.96	89.98	94.99	99.00
$O(n^{-3/2})$	$0.0^{7}132$	$0.0^{2}193$	0.1272	14.31	57.63	101.1	129.1	132.2	129.4
$(5, 20, 5)$				$\Omega = \text{diag}\{\frac{1}{4}\,(\frac{1}{4})\,1, 1\frac{1}{2}\}$					
Seq. sad.	1.003	4.974	9.944	29.90	49.88	69.90	89.98	94.98	99.00
$O(n^{-3})$	1.283	5.488	10.53	30.31	50.03	69.96	89.95	94.93	98.99
$(8, 40, 7)$				$\Omega = \text{diag}\{0, \frac{1}{4}, \frac{1}{2}, 1\,(\frac{1}{2})\,3\}$					
Seq. sad.	0.9919	4.996	9.992	29.96	49.93	70.00	90.04	95.05	99.01
$O(n^{-3})$	1.276	5.492	10.53	30.31	50.03	69.96	89.95	94.93	98.99
$(16, 40, 14)$				$\Omega_{16} = \text{diag}\{2\,(0), \frac{1}{4}, \frac{1}{2}\,(\frac{1}{2})\,3, 3\frac{1}{4}\,(\frac{1}{4})\,4\frac{3}{4}\}$					
Seq. sad.	0.9969	5.001	9.987	29.98	49.95	69.98	89.99	94.99	98.99
$O(n^{-3})$	4.817	12.10	17.66	33.76	49.26	66.63	87.26	93.26	98.52
$(32, 60, 28)$				$\Omega_{32} = I_2 \otimes \Omega_{16}$					
Seq. sad.	1.006	5.034	10.00	29.98	49.97	69.98	90.04	95.03	99.01
$O(n^{-3})$	2.254	9.708	17.55	41.22	57.32	69.69	82.71	88.24	95.71
$(64, 75, 56)$				$\Omega_{64} = I_4 \otimes \Omega_{16}$					
Seq. sad.	0.9908	4.984	9.966	29.95	49.90	69.94	89.96	94.99	99.00
$O(n^{-3})$	$0.0^{14}20$	$0.0^{11}20$	$0.0^{10}58$	$0.0^{7}34$	$0.0^{5}18$	$0.0^{4}63$	$0.0^{2}56$	0.364	0.7270

with error and hypothesis degrees of freedom n and m and sample size assumed to be $n + 1$. The entries show two and often three significant digit accuracy and the largest percentage absolute relative error for the whole table is $\leq 1\%$. This is astonishing accuracy considering the range of dimensionalities and quite small degrees of freedom used. Also included are computations for the expansions of Sugiura and Fujikoshi (1969) in terms of noncentral chi-square distributions that admit errors of $O(n^{-3})$, and the Edgeworth-type expansions of Sugiura (1973) with error $O(n^{-3/2})$ for the case $p = 2$.

The former expansion provides good accuracy in dimensions up to $p = 8$ but its accuracy deteriorates for $p \geq 16$. The latter approximation was considerably less accurate. These findings are partially explained by the following points: in our examples, Ω has entries of modest size (corresponding to an alternative hypothesis not very far from the null hypothesis); Sugiura and Fujikoshi (1969) assumes that

TABLE 4

Empirical percentiles (top) and saddlepoint percentiles (bottom; not shown when identical to table accuracy) for noncentral $\log W$. *(The empirical percentiles were determined from* 10^6 *independent simulations of* $\log W$. *The cases in which* $p = 32$ *and* 64 *resulted in equal percentiles to table accuracy. The approximate means and standard deviations* $\hat{\mu}$ *and* $\hat{\sigma}$ *were determined from the approximated CGF while those in parentheses were determined from simulation.)*

1	**5**	**10**	**30**	**50**	**70**	**90**	**95**	**99**
$(p, n, m) = (2, 10, 3)$			$\hat{\mu} = -0.7166\,(-0.7164)$			$\hat{\sigma} = 0.3945\,(0.3940)$		
−1.919	−1.461	−1.247	−0.8644	−0.6485	−0.4705	−0.2756	−0.2059	−0.1108
−1.916	−1.460	−1.247	−0.8635	−0.6480	−0.4704	−0.2758	−0.2061	−0.1111
$(p, n, m) = (5, 20, 5)$			$\hat{\mu} = -1.452\,(-1.453)$			$\hat{\sigma} = 0.4000\,(0.4001)$		
−2.536	−2.165	−1.983	−1.634	−1.417	−1.220	−0.9681	−0.8608	−0.6820
	−2.167	−1.984	−1.635	−1.418	−1.221	−0.9684	−0.8613	
$(p, n, m) = (8, 40, 7)$			$\hat{\mu} = -1.669\,(-1.669)$			$\hat{\sigma} = 0.3056\,(0.3059)$		
−2.458	−2.200	−2.070	−1.815	−1.651	−1.496	−1.290	−1.199	−1.038
−2.459		−2.071	−1.816				−1.197	−1.037
$(p, n, m) = (16, 40, 14)$			$\hat{\mu} = -6.674\,(-6.674)$			$\hat{\sigma} = 0.6142\,(0.6144)$		
−8.184	−7.715	−7.472	−6.982	−6.655	−6.339	−5.900	−5.696	−5.327
−8.185				−6.656			−5.697	−5.330
$(p, n, m) = (32, 60, 28)$			$\hat{\mu} = -17.25\,(-17.25)$			$\hat{\sigma} = 0.8193\,(0.8196)$		
−19.21	−18.62	−18.31	−17.67	−17.24	−16.81	−16.21	−15.92	−15.40
$(p, n, m) = (64, 75, 56)$			$\hat{\mu} = -61.29\,(-61.30)$			$\hat{\sigma} = 1.613\,(1.613)$		
−65.14	−63.98	−63.37	−62.13	−61.28	−60.44	−59.24	−58.67	−57.62

Ω has $O(1)$ entries as $n \to \infty$; and Sugiura (1973) assumes that Ω has $O(n)$ entries as $n \to \infty$.

On the computational side, the noncentral chi-square approximations, upon which the Sugiura and Fujikoshi (1969) approximation is based, were quite difficult to perform as a result of register underflow problems in floating point computation. By contrast the sequential saddlepoint approximations were straightforward and instantaneous.

All the table entries were determined in the following way. Empirical quantiles for the noncentral distribution of $\log W$ associated with the probabilities in the top row were determined by simulating 10^6 independent values of $\log W$. These empirical quantiles are displayed in Table 4. Sequential saddlepoint approximations and other expansions were then evaluated at these empirical percentiles so that Table 1 shows the accuracy of the saddlepoint approximations were the empirical percentiles regarded as exact.

Table 4 displays empirical percentiles along with saddlepoint percentiles (when different to table accuracy) associated with the computations of Table 3. It also

shows the approximate mean $\hat{\mu}$ and standard deviation $\hat{\sigma}$ for the noncentral distribution of $\log W$ as computed from the approximated cumulant generating function (CGF) $\hat{K}(s)$ of $\log W$. These cumulants were computed as $\hat{\mu} = \hat{K}'(0)$ and $\hat{\sigma} = \sqrt{\hat{K}''(0)}$. For comparison, the empirical averages and standard deviations of $\log W$ from the 10^6 simulated values are given in parentheses and show virtually no differences.

9. Sequential saddlepoint connections. The integral expression for ${}_1F_1(a; b; Z)$ in (3) specifies that it is the MGF of a Matrix Beta$_p(a, b)$ density when Z is taken to have (i, i)th component z_{ii} and (i, j)th component $z_{ij}/2$ for $i \neq j$. This MGF is also a conditional MGF when based upon the following construction. Suppose that $A \sim \text{Wishart}_p(a, \Sigma)$ independent of $B \sim \text{Wishart}_p(b, \Sigma)$. Then A given $S = A + B = I_p$ has a Matrix Beta$_p(a, b)$ density and the conditional MGF of $A|S = I_p$ is ${}_1F_1(a; b; Z)$.

The sequential saddlepoint method discussed in Fraser, Reid and Wong (1991) and Butler, Huzurbazar and Booth (1992) is concerned with approximating this conditional MGF using the double saddlepoint density approximation for the tilted conditional density of $A|S = I_p$.

THEOREM 1. *The calibrated conditional MGF approximation for $A|S = I_p$, obtained through the double saddlepoint density approximation, is analytically equivalent to ${}_1\hat{F}_1(a; b; Z)$, the calibrated Laplace approximation.*

A proof of this result is given in Butler and Wood (2000).

10. Relative error properties. In this section we briefly discuss the relative error properties of (i) the calibrated Laplace approximations for ${}_1F_1$ and ${}_2F_1$, and (ii) the sequential saddlepoint approximation referred to in Section 8.1.

10.1. *Relative error of ${}_1\hat{F}_1$ and ${}_2\hat{F}_1$.* As noted in Butler and Wood (2000), Section 7, there are many asymptotic regimes we may wish to consider for ${}_1F_1$ and ${}_2F_1$, and for this reason it seems difficult to give a complete summary of the global relative error properties of ${}_1\hat{F}_1$ and ${}_2\hat{F}_1$ in the form of a theorem. However, in each of the long list of cases reported in Butler and Wood (2000), it turned out that the relative error of the approximation remained uniformly bounded in the limit. Given the simplicity of ${}_1\hat{F}_1$ and ${}_2\hat{F}_1$ and the complexity of ${}_1F_1$ and ${}_2F_1$, these findings are remarkable. All the theoretical evidence we have points towards the following speculative conclusion: for fixed p, $b_0 > (p-1)/2$ and $c_0 > (p-1)/2$, and symmetric X,

$$\sup_{b \geq b_0, a \in \mathbf{R}, X \in \mathbf{R}^{p \times p}} \left| \log {}_1\hat{F}_1(a; b; X) - \log {}_1F_1(a; b; X) \right| < \infty$$

and for any $\varepsilon \in (0, 1)$,

$$\sup_{c \geq c_0, a, b \in \mathbf{R}, 0_p \leq X < (1-\varepsilon) I_p} \left| \log {}_2\hat{F}_1(a, b; c; X) - \log {}_2F_1(a, b; c; X) \right| < \infty.$$

The conditions $b \geq b_0$ and $c \geq c_0$ are required to avoid the regions in which the singularities of ${}_1F_1$ and ${}_2F_1$ occur (these regions are easily identified using the zonal polynomial expansions of ${}_1F_1$ and ${}_2F_1$). Of course bounded relative error does not in itself guarantee good numerical accuracy in practice, but these relative error results provide important insight into the excellent numerical accuracy seen in our examples.

10.2. *Relative error of the sequential saddlepoint approximation.* Let $Y = -\log W$ be the random variable whose moment generating function is given by the right-hand side of (33), but with s replaced by $-s$. Let $\hat{F}_{\mathrm{SSA}}$ denote the sequential saddlepoint approximation to the CDF of Y, obtained as indicated in Section 8.1. Note that, since $W \in (0, 1)$, we have $Y \in (0, \infty)$.

THEOREM 2. *For fixed p, m, n and Ω, we have the following limits*:

$$\lim_{y \to 0} \frac{P(Y \leq y)}{\hat{F}_{\mathrm{SSA}}(y)} = \frac{\hat{\Gamma}(mp/2 + 1)}{\Gamma(mp/2 + 1)}$$

and

$$\lim_{y \to \infty} \frac{P(Y > y)}{1 - \hat{F}_{\mathrm{SSA}}(y)} = \sqrt{2\pi} e^{-1} = \frac{\hat{\Gamma}(1)}{\Gamma(1)}$$

where $\hat{\Gamma}(a) = \sqrt{2\pi} a^{a-1/2} e^{-a}$ is Stirling's approximation to the gamma function.

Thus the relative error in the extreme upper and lower tails is actually rather small. This does not, of course, fully explain the excellent numerical results in Tables 3 and 4, but it does provide some insight into why the approximation works so well.

It should be noted that these limits are the same as in the null case. See, for example, Booth, Butler, Huzurbazar and Wood (1995) for the upper limit, and the lower limit is derived similarly. A general discussion of relative error of saddlepoint approximations is given by Jensen (1995).

To prove Theorem 2, all we need to do is show that: ${}_1F_1$ makes a negligible theoretical contribution to the extreme upper and lower tail probability; and that, in the extremes, ${}_1\hat{F}_1$ makes a negligible contribution to the Lugannani and Rice (1980) approximation. We now briefly sketch the proof.

Using the Kummer relation [see Theorem 7.4.3 of Muirhead (1982)], we may write the cumulant generating function of Y as

$$K_Y(s) = K_0(s) + K_1(s) + tr(\Omega/2)$$

where $K_1(s) = \log {}_1F_1((n+m)/2; (n+m)/2 - s; \Omega/2)$ and

$$K_0(s) = \log\left[\frac{\Gamma_p(n/2+s)\Gamma_p((n+m)/2)}{\Gamma_p(n/2)\Gamma_p((n+m)/2+s)}\right].$$

By expressing Γ_p as a product of univariate gamma functions, and using the fact that the latter has a simple pole at zero, we find that $K_0(s)$ is finite for $s \in (-\infty, (n-p+1)/2)$, but has a simple pole at $s = (n-p+1)/2$. Also, using the zonal polynomial expansion for all ${}_1F_1$ [see, e.g., Muirhead (1982)], we find that $K_1(s)$ is finite for all $s \in (-\infty, (n+m-p+1)/2)$ and has a simple pole at $s = (n+m-p+1)/2$. Consequently, since m is positive, the relevant domain of K_Y is $s \in (-\infty, (n-p+1)/2)$. Moreover, the upper tail of Y is determined by the behavior of K_Y around $s = (n-p+1)/2$, while the lower tail of Y is determined by the behavior of K_Y at $s = -\infty$. Now ${}_1F_1$ and ${}_1\hat{F}_1$ make a negligible contribution to the extreme upper tail because they (and their derivatives) remain bounded while $K_0(s)$ and its derivatives blow up at $s = (n-p+1)/2$. They also make a negligible contribution to the lower tail because, as $s \to -\infty$,

$$K_0(s) \sim -\tfrac{1}{2}mp\log(-s), \quad K_0'(s) \sim -\tfrac{1}{2}mp/s \quad \text{and} \quad K_0''(s) \sim \tfrac{1}{2}mp/s^2$$

while [writing $\hat{K}_1(s)$ when ${}_1\hat{F}_1$ replaces ${}_1F_1$ in K_1]

$$K_1^{(r)}(s) = O(s^{-r-1}) \quad \text{and} \quad \hat{K}_1^{(r)}(s) = O(s^{-r-1}), \qquad r = 0, 1, 2.$$

Of course this is just a sketch of the proof; the omitted details are straightforward but lengthy.

REFERENCES

ABRAMOWITZ, M. and STEGUN, I. A. (1972). *Handbook of Mathematical Functions*, 9th ed. Dover, New York.

BAGCHI, P. and KADANE, J. B. (1991). Laplace approximations to posterior moments and marginal distributions on circles, spheres, and cylinders. *Canad. J. Statist.* **19** 67–77.

BARNDORFF-NIELSEN, O. E. and WOOD, A. T. A. (1998). On large deviations and choice of ancillary for p^* and r^*. *Bernoulli* **4** 35–63.

BINGHAM, C., CHANG, T. and RICHARDS, D. (1992). Approximating the matrix Fisher and Bingham distributions: applications to spherical regression and Procrustes analysis. *J. Multivariate Anal.* **41** 314–337.

BLEISTEIN, N. and HANDELSMAN, R. A. (1975). *Asymptotic Expansions of Integrals*. Dover, New York.

BOOTH, J. G., BUTLER, R. W., HUZURBAZAR, S. and WOOD, A. T. A. (1995). Saddlepoint approximations for p-values of some tests of covariance matrices. *J. Statist. Comput. Simulation* **53** 165–180.

BUTLER, R. W., HUZURBAZAR, S. and BOOTH, J. G. (1992). Saddlepoint approximations for the Bartlett–Nanda–Pillai trace statistic in multivariate analysis. *Biometrika* **79** 705–715.

BUTLER, R. W. and WOOD, A. T. A. (2000). Laplace approximations for hypergeometric functions of matrix argument. Technical Report 00-05, Div. Statistics, Univ. Nottingham. Available at www.maths.nottingham.ac.uk/statsdiv/research/reports.html.

CHIKUSE, Y. (1998). Density estimation on the Stiefel manifold. *J. Multivariate Anal.* **66** 188–206.

CONSTANTINE, A. G. (1963). Some noncentral distribution problems in multivariate analysis. *Ann. Math. Statist.* **34** 1270–1285.

CONSTANTINE, A. G. and MUIRHEAD, R. J. (1976). Asymptotic expansions for distributions of latent roots in multivariate analysis. *J. Multivariate Anal.* **6** 369–391.

EFSTATHIOU, M., GUTIÉRREZ-PEÑA, E. and SMITH, A. F. M. (1998). Laplace approximations for natural exponential families with cuts. *Scand. J. Statist.* **25** 77–92.

FISHER, R. A. (1990). *Statistical Methods, Experimental Design, and Scientific Inference*. Oxford Univ. Press.

FRASER, D. A. S., REID, N. and WONG, A. (1991). Exponential linear models: a two-pass procedure for saddlepoint approximation. *J. Roy. Statist. Soc. Ser. B* **53** 483–492.

GLYNN, W. J. and MUIRHEAD, R. J. (1978). Inference in canonical correlation analysis. *J. Multivariate Anal.* **8** 468–478.

GUPTA, R. D. and RICHARDS, D. ST. P. (1985). Hypergeometric functions of scalar matrix argument are expressible in terms of classical hypergeometric functions. *SIAM J. Math. Anal.* **16** 852–858.

HERZ, C. S. (1955). Bessel functions of matrix argument. *Ann. Math.* **61** 474–523.

JAMES, A. T. (1964). Distributions of matrix variates and latent roots derived from normal samples. *Ann. Math. Statist.* **35** 475–501.

JENSEN, J. L. (1995). *Saddlepoint Approximations*. Oxford Univ. Press.

JOHNSON, N. L. and KOTZ, S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions*. Wiley, New York.

LUGANNANI, R. and RICE, S. O. (1980). Saddlepoint approximations for the distribution of the sum of independent random variables. *Adv. Appl. Probab.* **12** 475–490.

MAGNUS, J. R. and NEUDECKER, H. (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley, New York.

MATHAI, A. M. (1993). *A Handbook of Generalized Special Functions for Statistical and Physical Sciences*. Clarendon Press, Oxford.

MUIRHEAD, R. J. (1975). Expressions for some hypergeometric functions of matrix argument with applications. *J. Multivariate Anal.* **5** 283–293.

MUIRHEAD, R. J. (1978). Latent roots and matrix variates: a review of some asymptotic results. *Ann. Statist.* **6** 5–33.

MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley, New York.

OLVER, F. W. J. (1991). Uniform, exponentially improved, asymptotic expansions for the confluent hypergeometric function and other integral transforms. *SIAM J. Math. Anal.* **22** 1475–1489.

OLVER, F. W. J. (1993). Exponentially-improved asymptotic solutions of ordinary differential equations. I. The confluent hypergeometric function. *SIAM J. Math. Anal.* **24** 756–767.

PARIS, R. B. (1992). Smoothing of the Stokes phenomenon using Mellin–Barnes integrals. Asymptotic methods in analysis and combinatorics. *J. Comput. Appl. Math.* **41** 117–133.

PHILLIPS, P. C. B. (1983). Exact small sample theory in the simultaneous equations model. In *Handbook of Econometrics* **1** (Z. Griliches, and M. D. Intriligator, eds.) Chapter 8. North-Holland, Amsterdam.

ROY, S. N. (1939). p-statistics, or some generalizations in analysis of variance appropriate to multivariate problems. *Sankhyā* **4** 381–396.

SRIVASTAVA, M. S. and CARTER, E. M. (1980). Asymptotic expansions for hypergeometric functions. In *Multivariate Analysis V* (P. Krishnaiah, ed.) 337–347. North-Holland, Amsterdam.

SUGIURA, N. (1969). Asymptotic non-null distributions of the likelihood ratio criteria for covariance matrix under local alternatives. Mimeo Ser. 609, Inst. Statistics, Univ. North Carolina, Chapel Hill.

SUGIURA, N. (1973). Further asymptotic formulas for the non-null distributions of three statistics for multivariate linear hypothesis. *Ann. Inst. Statist. Math.* **25** 153–163.

SUGIURA, N. and FUJIKOSHI, Y. (1969). Asymptotic expansions of the non-null distributions of the likelihood ratio criteria for multivariate linear hypothesis and independence. *Ann. Math. Statist.* **40** 942–952.

SUGIYAMA, T. (1967). On the distribution of the largest latent root of the covariance matrix. *Ann. Math. Statist.* **38** 1148–1151.

DEPARTMENT OF STATISTICS
COLORADO STATE UNIVERSITY
FORT COLLINS, COLORADO 80523
E-MAIL: walrus@stat.colostate.edu

SCHOOL OF MATHEMATICAL SCIENCES
UNIVERSITY OF NOTTINGHAM
NOTTINGHAM NG7 2RD
UNITED KINGDOM
E-MAIL: atw@maths.nott.ac.uk