Laplace Approximations to Hypergeometric Functions of Two Matrix Arguments

Ronald W. Butler and Andrew T.A. Wood
Colorado State University and University of Nottingham
August 24, 2001

Abstract
We present a unified approach to Laplace approximation of hypergeometric functions with two matrix arguments. The general form of the approximation is designed to exploit the Laplace approximations to hypergeometric functions of a single matrix argument presented in Butler and Wood (2000, 2001) which have proved to be very accurate in a variety of settings. All but one of the approximations presented here appear to be new. Numerical accuracy is investigated in a number of statistical applications.

1 Introduction

1.1 Background
In this paper we present Laplace approximations to hypergeometric functions of two matrix arguments. In all but one case (that of ${}_5F_4^{(m)}$) the resulting approximation appears to be new. Our principal reason for revisiting this topic is that recent work in Butler and Wood (2000) has shown that certain new Laplace approximations to the single-matrix-argument hypergeometric functions are often very accurate. The approximations to two-matrix-argument hypergeometric functions presented here are designed to exploit the accuracy of the one-matrix-argument approximations in Butler and Wood (2000) as far as possible.

A number of authors derived asymptotic approximations to hypergeometric functions of two matrix arguments in the 1970s and 1980s. See, in particular, Chattopadhyay and Pillai (1973), Chattopadhyay, Pillai and Li (1976), Muirhead (1978), Glynn and Muirhead (1978), Srivastava and Carter (1980a and 1980b), Glynn (1980) and Muirhead (1982). In most of the above work, the emphasis is on obtaining asymptotic representations of these hypergeometric functions to enhance theoretical understanding, rather than for numerical approximation purposes.

In contrast, our principal motivation is to provide practical approximations and give some indication of their numerical accuracy. Frequently, it has been
found that Laplace's method yields approximations in sub-asymptotic situations which are more accurate than one has a right to expect. In practice, however, care is required in the implementation of Laplace's approximation. In particular, care is needed in deciding which terms (if any) should be left out of the maximization, and care is also needed in deciding how to calibrate (i.e. normalize) the approximation (if at all). [In fact, we have not attempted to calibrate the approximations given below as there does not seem to be a natural way to do it in the case of hypergeometric functions with two matrix arguments.] In general, no clear rules seem to be available for deciding on these issues. However, see Butler and Wood (2000, Section 3) for further discussion in the case of hypergeometric functions with one matrix argument.

1.2 Definition of hypergeometric functions

The single-matrix-argument hypergeometric function \( _p F_q \) can be defined in terms of a zonal polynomial expansion as follows:

\[
_p F_q(\alpha_1, \ldots, \alpha_p; \beta_1, \ldots, \beta_q; X) = \sum_{\eta} \frac{(\alpha_1)_\eta \cdots (\alpha_p)_\eta}{(\beta_1)_\eta \cdots (\beta_q)_\eta} \frac{C_\eta(X)}{|\eta|!}
\]

where \( \alpha_1, \ldots, \alpha_p \) and \( \beta_1, \ldots, \beta_q \) are real numbers, and \( X \) is a real symmetric matrix; \( \eta = (\eta_1, \ldots, \eta_k) \), with \( \eta_1 \geq \eta_2 \geq \ldots \eta_k > 0 \), is a partition of the integer \(|\eta| = \eta_1 + \ldots + \eta_k \); \( C_\eta(X) \) is the zonal polynomial associated with the partition \( \eta \) (see e.g. Muirhead, 1982); and, for a real number \( \alpha \) and integer partition \( \eta \), we define

\[
(\alpha)_\eta = \prod_{i=1}^{k} \prod_{j=0}^{\eta_i} \left\{ \alpha - \left( i - 1 \right)/2 - j \right\}.
\]

There are various integral representations for the \( _p F_q \), and also various integral formulae linking different \( _p F_q \); see Muirhead (1982) for further details.

The hypergeometric function of two matrix arguments, written \( _p F_q^{(m)} \), may be written as an integral of the corresponding one-argument function, \( _p F_q \), over the orthogonal group, as follows:

\[
_p F_q^{(m)}(\alpha_1, \ldots, \alpha_p; \beta_1, \ldots, \beta_q; A, B) = \int_{Q \in O(m)} _p F_q(\alpha_1, \ldots, \alpha_p; \beta_1, \ldots, \beta_q; AQBQ^T)(dQ) \quad (1)
\]

where \( O(m) = \{ Q(m \times m) : Q^TQ = QQ^T = I_m \} \) is the group of orthogonal \( m \times m \) matrices, and \( (dQ) \) is normalised Lebesgue measure on \( O(m) \) chosen so that \( \int_{O(m)} (dQ) = 1 \).

1.3 Relevance to eigenvalue distributions

For statistical applications, the important hypergeometric functions with two matrix arguments are given by \( (p, q) = (0, 0), (1, 0), (0, 1), (1, 1) \) and \( (2, 1) \). This
is because, in each case, the joint density of the eigenvalues of a random matrix of statistical importance is determined by \( pF_q^{(m)} \). [When we say is determined by, we mean is equal to the relevant \( pF_q^{(m)} \) multiplied by elementary functions.]

- \((p, q) = (0, 0)\). The function \( 0F_0^{(m)} \) determines the joint density function of the eigenvalues of a \( m \times m \) central Wishart matrix; see Muirhead (1982, p. 106).

- \((p, q) = (1, 0)\). The function \( 1F_0^{(m)} \) determines the joint density of the eigenvalues of \( A_1A_2^{-1} \) where \( A_1 \) and \( A_2 \) are independent central \( m \times m \) Wishart matrices; see Muirhead (1982, p. 312).

- \((p, q) = (0, 1)\). If \( A \) \((m \times m)\) is a non-central Wishart matrix, then \( 0F_1^{(m)} \) determines the joint density of the eigenvalues of \( A \); see Muirhead (1982, p. 442).

- \((p, q) = (1, 1)\). If \( A \) \((m \times m)\) is a non-central Wishart matrix and \( B \) \((m \times m)\) is an independent central Wishart matrix, then \( 1F_1^{(m)} \) determines the joint distribution of the eigenvalues of \( AB^{-1} \); see Muirhead (1982, p. 450).

- \((p, q) = (2, 1)\). Suppose

\[
A = \begin{pmatrix}
A_{11} & A_{21} \\
A_{21} & A_{22}
\end{pmatrix}
\]

is a central Wishart matrix with blocks \( A_{ij} \) of dimension \( m_i \times m_j \), where \( m = m_1 \leq m_2 \). Then \( 2F_1^{(m)} \) determines the joint density of the eigenvalues of \( A_{11}^{-1}A_{12}A_{22}^{-1}A_{21} \); see Muirhead (1982, p. 539).

Further details of applications of these eigenvalue distributions to statistical inference are given by Muirhead (1982). Some of these applications are considered later in this paper.

### 1.4 Outline of paper

In Section 2 we derive the general form of our Laplace approximation, and then briefly discuss implementation in particular cases. Numerical investigations and statistical applications are then considered in Section 3.

### 2 The Approximation

Consider matrices \( A = \text{diag}\{a_1, \ldots, a_m\} \) and

\[
B = \text{diag}\{b_1, \ldots, b_1, b_2, \ldots, b_r, \ldots, b_r\}
\]
where \( a_1 > a_2 > \ldots > a_m > 0 \), \( b_1 > \ldots > b_r > 0 \), and \( b_i \) has multiplicity \( m_i \) where \( m = \sum_{i=1}^{r} m_i \). Write \( Y = Y(Q) = A^{1/2}QBQ^TA^{1/2} \) where \( Q \in \mathcal{O}(m) \), the space of \( m \times m \) orthogonal matrices. Consider

\[
F(Q) \equiv G(Y(Q)) \equiv f(\lambda_1, \ldots, \lambda_m),
\]

where \( F \) depends on \( Q \) only through the (necessarily real) eigenvalues \( \lambda_1, \ldots, \lambda_m \) of \( Y = Y(Q) \). Suppose that \( F \) is maximized at any block diagonal \( Q \in \mathcal{O}(m) \) of the form

\[
Q = \text{diag}(Q_1, \ldots, Q_r), \quad \text{where} \quad Q_i \in \mathcal{O}(m_i), \quad 1 \leq i \leq r,
\]
as will be the case for all \( F \) that we consider; see Chattopadhyay, Pillai and Li (1976) for relevant optimization results. Let \( \omega_m = \frac{2^m \pi^{m^2/2}}{\Gamma_m(m/2)} \) denote the volume of \( \mathcal{O}(m) \), \( m \geq 1 \), and write

\[
\Omega(m_1, \ldots, m_r) = \omega_m^{-1} \prod_{i=1}^{r} \omega_{m_i}.
\]

Define

\[
J = \prod_{\{(i,j) : 1 \leq i < j \leq m \wedge \sigma(i) < \sigma(j)\}} \left\{ (a_i - a_j)(b_{\sigma(i)} - b_{\sigma(j)}) \Psi_{ij} \right\}
\]

where

\[
\Psi_{ij} = 2 \frac{f_i a_i b_{\sigma(i)} - f_j a_j b_{\sigma(j)}}{a_i b_{\sigma(i)} - a_j b_{\sigma(j)}},
\]

\[
f_i = \frac{\partial f(\lambda_1, \ldots, \lambda_m)}{\partial \lambda_i},
\]

and \( \sigma(i) \) is the smallest integer \( k \) such that \( i \leq \sum_{j=1}^{k} m_i \). Note that \( J \) is a product of

\[
s = \sum_{i=1}^{r-1} \sum_{j=i+1}^{r} m_i m_j
\]
terms, and that \( s = m(m - 1)/2 \) when \( m_i = 1 \) for all \( i \).

The Laplace approximation to \( {}_pF_q^{(m)}(\alpha; \beta; A, B) \) is then given by

\[
{}_pF_q^{(m)}(\alpha; \beta; A, B) = (2\pi)^{s/2} \Omega(m_1, \ldots, m_r) J^{-1/2} {}_pF_q^{(m)}(\alpha; \beta; AB)
\]

where \( \alpha = (\alpha_1, \ldots, \alpha_p)^T \), \( \beta = (\beta_1, \ldots, \beta_q)^T \), and \( {}_pF_q^{(m)} \) is the corresponding one-argument matrix hypergeometric function. The evaluation or approximation of \( {}_pF_q^{m} \) is discussed in subsection 2.2.3.

Note that, in this approximation, we should take

\[
f_i = \frac{\partial}{\partial \lambda_i} f(\lambda_1, \ldots, \lambda_m) = \frac{\partial}{\partial \lambda_i} \log {}_pF_q^{(m)}(\alpha; \beta; \text{diag}(\lambda_1, \ldots, \lambda_m)).
\]
We now derive the Hessian $J$ in two cases. In subsection 2.1, $B$ has distinct eigenvalues, i.e. each eigenvalue $b_i$ has multiplicity $m_i = 1$; and in subsection 2.2, $B$ has repeated eigenvalues, with eigenvalue $b_i$ having multiplicity $m_i$, $i = 1, \ldots, r$. Finally, in subsection 2.3, we discuss the calculation (or approximation) of $p F_q$ and $f_i$ in the various cases.

### 2.1 Hessian Case I: distinct eigenvalues

Here we assume that $B$ has distinct eigenvalues $b_1 > b_2 > \ldots > b_m \geq 0$.

The first differential of $F(Q)$ is given by

$$dF(Q) = \sum_{i=1}^{m} f_i d\lambda_i(Q)$$

where

$$d\lambda_i = u_i^T(Y)(dY)u_i(Y)$$

and $u_i(Y)$ is the unit eigenvector of $Y = Y(Q)$ associated with the eigenvalue $\lambda_i$; see e.g. Magnus and Neudecker (1988). The second differential is given by

$$d^2 F(Q) = \sum_{i,j=1}^{m} f_{ij} d\lambda_i d\lambda_j + \sum_{i=1}^{m} f_i d^2 \lambda_i.$$

At $Q = \text{diag}\{\pm 1, \ldots, \pm 1\}$, $d\lambda_i = 0$ for all $i$, so the first term on the right hand side makes no contribution to the Hessian at such $Q$, and so may be ignored here.

Now

$$d^2 \lambda_i = u_i(Y)^T(d^2 Y)u_i(Y) + \text{tr}
\left[(dY)d\{u_i(Y)u_i(Y)^T\}\right].$$

> From Magnus and Neudecker (1988, p.159),

$$du_i(Y) = (\lambda_i I - Y)^+(dY)u_i(Y),$$

where $A^+$ denotes the Moore-Penrose inverse of a matrix $A$. Therefore

$$d\{u_i(Y)u_i(Y)^T\} = \{du_i(Y)\}u_i(Y)^T + u_i(Y)\{du_i(Y)\}^T$$

$$= (\lambda_i I - Y)^+(dY)u_i(Y)u_i(Y)^T + u_i(Y)u_i(Y)^T(dY)^T(\lambda_i I - Y)^+. $$

Now transform from $Q$ to $H$ where $Q = \exp(H)$ and $H$ is skew-symmetric. Then

$$Y = A^{1/2}\left(I + H + \frac{1}{2}H^2 + \ldots\right)B\left(I + H^T + \frac{1}{2}(H^T)^2 + \ldots\right)A^{1/2}$$

so

$$(dY)_{H=0} = A^{1/2}(dH)BA^{1/2} + A^{1/2}B(dH)^T A^{1/2}$$
and
\[
(d^2Y)_{H=0} = A^{1/2}(dH)BA^{1/2} + A^{1/2}B(dH)^T(dH)^T A^{1/2} \\
+ 2A^{1/2}(dH)B(dH)^T A^{1/2}.
\]

Define
\[
U(H) = \sum_{i=1}^m f_i \text{ tr } \left\{ (d^2Y)u_i(Y)u_i(Y)^T \right\},
\]
and
\[
V(H) = \sum_{i=1}^m f_i \text{ tr } \left\{ (dY)d\{u_i(Y)u_i(Y)^T\} \right\}.
\]

At $H = 0$, $Q = I$ and $Y = A^{1/2}BA^{1/2}$ so that $\{u_i(Y)\}$ is the standard orthonormal basis, and
\[
\sum_{i=1}^m f_i u_i(Y)u_i(Y)^T = \text{ diag} \{f_1, \ldots, f_m\} := \Phi.
\]

Let $\tilde{D}_m$ and $\tilde{\nu}(A)$ be defined as in Magnus (1998, p. 94). In particular, $\tilde{D}_m$ and $\tilde{\nu}$ are chosen so that for any skew-symmetric matrix $A$, $\tilde{D}_m \tilde{\nu}(A) = \text{ vec}(A)$. Then
\[
U(0) = \text{ tr } \left\{ \Phi(d^2Y) \right\} \\
= \text{ tr } \left\{ \Phi A^{1/2}(dH)(dH)BA^{1/2} \right\} + \text{ tr } \left\{ \Phi A^{1/2}B(dH)^T(dH)^T A^{1/2} \right\} \\
+ 2 \text{ tr } \left\{ \Phi A^{1/2}(dH)B(dH)^T A^{1/2} \right\} \\
= 2 \text{ tr } \left\{ (\Phi AB)(dH)I(dH) \right\} + 2 \text{ tr } \left\{ (\Phi A)(dH)B(dH)^T \right\} \\
= 2\{ \text{ vec}(dH) \}^T \left[ B \otimes \Phi A - I \otimes \Phi AB \right] \text{ vec}(dH) \\
= 2\tilde{\nu}(dH)^T \tilde{D}_m^T \left[ B \otimes \Phi A - B \otimes \Phi AB \otimes I \right] \tilde{D}_m \tilde{\nu}(dH)
\]
since $(dH)^T = -(dH)$ by skew-symmetry, and
\[
\text{tr}(ABCD) = \text{ vec}(D^T)^T(C^T \otimes A) \text{ vec}(B).
\]

But from Magnus (1988, p.101, Example 6.11),
\[
2\tilde{D}_m^T \left( (\Phi A) \otimes B - (\Phi AB) \otimes I \right) \tilde{D}_m
\]
is a diagonal matrix with diagonal elements
\[
2(f_i a_i b_j + f_j a_j b_i - f_i a_j b_i - f_j a_i b_j) \\
= -2(b_i - b_j)(f_i a_i - f_j a_j), \quad 1 \leq i < j \leq m.
\]
Now consider $V(H)$. We have

$$V(H) = \sum_{i=1}^{m} f_i \text{tr} \left[ (dy) d \{u_i(Y)u_i(Y)^T \} \right]$$

$$= \sum_{i=1}^{m} f_i \text{tr} \left\{ (\lambda_i I - Y)^+ (dy)u_i(Y)u_i(Y)^T(dy) \right\}$$

$$+ u_i(Y)u_i(Y)^T(dY)(\lambda_i I - Y)^+(dY)$$

$$= 2 \sum_{i=1}^{m} f_i \{ \text{vec}(dY) \}^T [u_iu_i^T \otimes (\lambda_i I - Y)^+] \text{vec}(dY).$$

Using the identity $\text{vec}(ABC) = (C^T \otimes A) \text{vec}(B)$, we obtain

$$\text{vec}(dY)_{H=0} = \left( (BA^{1/2}) \otimes A^{1/2} - A^{1/2} \otimes (A^{1/2}B) \right) \text{vec}(dH)$$

$$= (A^{1/2} \otimes A^{1/2}) (B \otimes I - I \otimes B) \text{vec}(dH).$$

Therefore, since all the matrices concerned are diagonal and so commute, and writing $P_i = u_iu_i^T$ and $C_i = (\lambda_i I - AB)^+$, we have

$$V(0) = 2 \sum_{i=1}^{m} f_i \tilde{v}(dH)^T \tilde{D}_m \left( A \otimes A \right) (B \otimes I - I \otimes B)^2 (P_i \otimes C_i) \tilde{D}_m \tilde{v}(dH)$$

Then, using the fact that $AP_i = a_i P_i$ and $BP_i = b_i P_i$, we obtain

$$(A \otimes A) (B \otimes I - I \otimes B)^2 (P_i \otimes C_i)$$

$$= (A \otimes A) \left( b_i^2 P_i \otimes C_i + P_i \otimes (B^2 C_i) - 2b_i P_i \otimes (BC_i) \right)$$

$$= (A \otimes A) \left\{ P_i \otimes (b_i^2 I + B^2 - 2b_i B) C_i \right\}$$

$$= a_i P_i \otimes A(b_i^2 I + B^2 - 2b_i B) C_i.$$ 

Thus

$$V(0) = 2 \tilde{v}(dH)^T \tilde{D}_m \left\{ \sum_{i=1}^{m} f_i a_i P_i \otimes A(b_i^2 I + B^2 - 2b_i B) C_i \right\} \tilde{D}_m \tilde{v}(dH)$$

Using Magnus (1988, p.101, Example 6.11) again, we see that

$$2 \tilde{D}_m \left\{ \sum_{i=1}^{m} f_i a_i P_i \otimes A(b_i^2 I + B^2 - 2b_i B) C_i \right\} \tilde{D}_m$$

is a diagonal matrix with diagonal elements

$$\frac{2 f_i a_i a_j (b_i - b_j)^2}{a_i b_i - a_j b_j} + \frac{2 f_j a_i a_j (b_i - b_j)^2}{a_j b_j - a_i b_i}$$

$$= 2 \frac{(f_i - f_j)a_i a_j (b_i - b_j)^2}{a_i b_i - a_j b_j}, \quad 1 \leq i < j \leq m.$$
Finally, we see that

$$U(0) + V(0) = \nu(dH)^T \Xi \nu(dH)$$

where

$$\Xi = \text{Hessian} \{ F(Q) \}_{Q=I}$$

is a diagonal matrix with diagonal elements

$$-2(b_i - b_j)(f_i a_i - f_j a_j) + 2\frac{(f_i - f_j)a_i a_j (b_i - b_j)^2}{a_i b_i - a_j b_j},$$

which simplifies to

$$-(a_i - a_j)(b_i - b_j) \Psi_{ij}, \quad 1 \leq i < j \leq m,$$

where $\Psi_{ij}$ is defined in (2).

### 2.2 Hessian Case II: repeated eigenvalues

Here we assume that $B$ has distinct eigenvalues $b_1 > b_2 > ... > b_r$ where $b_i$ has multiplicity $m_i$. The function on $O(m)$ given by

$$Y(Q) = AQ^T$$

may be viewed as a function on the product space $M \times N$ where

$$M = \frac{O(m)}{O(m_1) \times ... \times O(m_r)}, \quad \text{and} \quad N = O(m_1) \times ... \times O(m_r),$$

and $Y(Q)$ is constant on $N$. The space $M$ is not a group, but it does have nice structure as a quotient space. In the terminology of O'Neill (1966, p. 466), $M$ is a Riemannian homogeneous space, and in the terminology of O'Neill (1983, p. 212), the natural mapping $\pi : O(m) \rightarrow M$ is a semi-Riemannian submersion. A key point is that geodesics in $O(m)$ which are "horizontal" with respect to this submersion map to geodesics in the quotient space $M$ (see the Corollary on p. 212 of O'Neill, 1983). In practical terms, this means that at each point in $M$, the tangent space can be represented as a linear subspace, $T$ say, of the vector space of skew-symmetric matrices. The linear subspace $T$ may be described as follows:

$$T = \{ A : \quad A^T = -A; \quad a_{ij} = 0 \text{ if } \sigma(i) = \sigma(j) \}$$

where $\sigma$ is the map defined at the beginning of Section 2. Note that $T$ has dimension $s$ given in (3).

In the repeated eigenvalue case, the derivation of $J$ is essentially the same as that given in subsection 2.1, except that the skew-symmetric matrix $H$ which appears there is now restricted to the linear subspace $T$. 
2.3 Implementation in various cases

We briefly run through the implementation of the approximation in the cases considered in subsection 1.3.

- \((p, q) = (0, 0)\). In this case we use the fact that \(\text{etr}(X) = e^{\text{tr}(X)}.\) Thus \(f_i = 1\) in (5).

- \((p, q) = (1, 0)\). From Muirhead (1982, p.262), \(\text{etr}(X) = |X - \bar{X}|^{-\alpha},\)
  and \(f_i\) in (5) is given by \(f_i = a/(1 - \lambda_i)\).

In the remaining cases \(pF_q\) can not be expressed in terms of elementary functions, and we make use of the Laplace approximations given in Butler and Wood (2001). Without loss of generality we assume that \(X = \text{diag}(x_1, \ldots, x_m)\).

- \((p, q) = (0, 1)\). Here, we approximation \(\text{etr}^{0}_1\) using the following Laplace approximation given in Butler and Wood (2001):

\[
\text{etr}^{0}_1(n/2; XX^T/(4)) = R_{0, 1}^{-1/2} \prod_{i=1}^{m} \left\{ (1 - \hat{y}_i^2)^{(n-2m-1)/2} e^{x_i \hat{y}_i} \right\}
\]

where

\[
R_{0, 1} = \prod_{i=1}^{m} \prod_{j=1}^{m} \frac{1 + \hat{y}_i \hat{y}_j}{(1 - \hat{y}_i^2)(1 - \hat{y}_j^2)};
\]

and \(\hat{y}_i = y_i(x_i)\) is given by

\[
\hat{y}(x) = u/\sqrt{u^2 + 1 + 1}
\]

where \(u = 2x/(n - 1)\). The derivatives \(f_i\) in (5) may be approximated by numerical differentiation of \(\text{etr}^0_1\).

- \((p, q) = (1, 1)\). From Butler and Wood (2000) we have

\[
\text{etr}^1_1(a; b; X) = b^{\text{etr}(m-m+1)/4} R_{1, 1}^{-1/2} \prod_{i=1}^{m} \left\{ \left( \frac{\hat{y}_i}{a} \right)^a \left( 1 - \hat{y}_i \right)^{b-a} e^{x_i \hat{y}_i} \right\}
\]

where

\[
R_{1, 1} = \prod_{i=1}^{m} \prod_{j=1}^{m} \left\{ \frac{\hat{y}_i \hat{y}_j}{a} + \frac{(1 - \hat{y}_i)(1 - \hat{y}_j)}{b - a} \right\};
\]

and \(\hat{y}_i = y_i(x_i)\) is given by

\[
\hat{y}(x) = 2a/(b - x + \sqrt{(x - b)^2 + 4ax}).
\]

Similarly, \(f_i\) in (5) may be approximated by numerical differentiation of \(\text{etr}^1_1\).
(p, q) = (2, 1). From Butler and Wood (2000), we have

\[
2 \tilde{F}_1(a; b; c; X) = CR_{2,1}^{-1/2} \prod_{i=1}^{m} \left\{ \frac{\tilde{y}_i \tilde{y}_j}{a} \left( \frac{1 - \tilde{y}_i}{c - a} \right)^{c-a} (1 - x_i \tilde{y}_i)^{-b} \right\}
\]

where \( C = c^{m-m(m+1)/4} \) and

\[
R_{2,1} = \prod_{i=1}^{m} \prod_{j=1}^{m} \left\{ \frac{\tilde{y}_i \tilde{y}_j}{a} + \frac{(1 - \tilde{y}_i)(1 - \tilde{y}_j)}{c - a} - \frac{b x_i x_j \tilde{y}_i \tilde{y}_j (1 - \tilde{y}_i)(1 - \tilde{y}_j)}{(1 - x_i \tilde{y}_i)(1 - x_j \tilde{y}_j)a(c - a)} \right\},
\]

and \( \tilde{y}_i = \tilde{y}(x_i) \) is given by

\[
\tilde{y}(x) = 2a/(\sqrt{\tau^2 - 4ax(c - b)} - \tau)
\]

where \( \tau = x(b - a) - c \). As in the previous two cases, \( f_i \) in (5) can be approximated by numerical differentiation, this time of the function \( \log 2 \tilde{F}_1 \).

3 Numerical Examples and Applications

The two argument \( pF_q^{(m)} \) function has been defined in (1) and our Laplace approximation in (4) provides approximate values for this integral. The oscillatory nature of the integrand in (1) makes this an exceedingly difficult problem. However, we find that our Laplace approximation reproduces values of this integral with quite reasonable accuracy. This accuracy is an improvement over that of the Laplace approximations suggested in Glynn and Muirhead (1978) and Glynn (1980). Even further improvement is achieved when approximating relative values of (1). These aspects of its accuracy are discussed in subsection 3.1.

The good relative accuracy of this approximation has practical importance when applied in some importance sampling applications. Such a setting occurs when testing the rank of the noncentrality matrix in MANOVA. Here, importance sampling can be used to compute the power function of the likelihood ratio test using importance weights that are ratios of \( 1F_1^{(m)} \)-values. This application is considered in subsection 3.2. A second related example considers the power computation of the likelihood ratio test for the number of nonzero canonical correlations relating two vectors. The power of this test involves ratios of \( 2F_1^{(m)} \)-values and is considered in subsection 3.3.

3.1 Absolute and relative error

Consider the accuracy of approximation (4) when determining \( 1F_1^{(m)}(\alpha; \beta; A, B) \). To evaluate (4), we use the single argument approximation \( \tilde{F}_1(\alpha; \beta; AB) \) given in (6). This particular single argument approximation has exhibited relative errors of no more than 0.1% over a wide selection of values for \( \alpha, \beta, \) and \( AB \).
including quite small values of $\alpha$ and $\beta$ near 1 that represent small degrees of freedom in practical applications. Thus, in practice, it is likely that most of the error in estimating the two argument function $\mathbf{1}_{F_1}^{(m)}(\alpha; \beta; A, B)$ will be due to the Laplace approximation in (4), as opposed to the error in approximating the one argument function $\mathbf{1}_{F_1}$.

Table 1 compares Laplace approximation (4) to an "Exact" computation determined from simulation, and the Laplace approximation for $\mathbf{1}_{F_1}^{(m)}$ developed in Glynn (1980, §6). Except for the case $m = 8$, values for the parameters $m, \alpha, \beta, A$, and $B$ have been chosen which are directly relevant to applications for testing the dimensionality of the noncentrality matrix in a 1-way MANOVA; see the next subsection for details. The "Exact" entry was determined by simulation as

$$\mathbf{1}_{F_1}^{(m)}(\alpha; \beta; A, B) \approx 10^{-6} \sum_{i=1}^{10^6} \mathbf{1}_{F_1}(\alpha; \beta; AQ_iBQ_i^T)$$

where $\{Q_i\}$ are i.i.d. determinations of orthogonal matrices that are uniform on $O(m)$. Values for $Q_i$ are conveniently generated by extracting the eigenvectors from i.i.d. sequences of Wishart $m (2m, I_m)$ matrices. For each choice of parameters $m, \alpha, \beta, B$, two choices for the matrix $A_1$ and $A_2$ are listed in the table so that we might assess the relative accuracy of approximation (4) as matrix argument $A$ varies. The ± range for each of the simulations determines a 95% confidence interval of the function.

>From column 3, we see that the Laplace approximation is able to track the correct order of the result from simulation, as would be expected from the asymptotic development. It shows consistently better accuracy than the approximation of Glynn (1980) in column 4 both in the table as well as in numerous other examples. The first case of Table 1 presents a sub-asymptotic example that corresponds to 5 degrees of freedom for error in a 3 dimensional MANOVA setting. Here, approximation (4) maintains respectable accuracy but the Glynn (1980) approximation cannot. In the fourth and last case, Glynn (1980) is undefined; it contains the factor $\Gamma_k(\beta)$ with $k = \text{rank}(A)$ and its value here $\Gamma_8(2)$ is undefined.

The relative accuracy of approximation (4) is better than its absolute accuracy as may be seen from the last column. Here the top entry is the ratio of simulated values whereas the bottom entry is the ratio of Laplace approximations. We see very good agreement in these ratios whose accuracy is used when implementing the importance sampling of the next subsection.
<table>
<thead>
<tr>
<th>$m$</th>
<th>$\alpha; \beta$</th>
<th>$B$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>4; $1\frac{1}{2}$</td>
<td>2.5 (2, 1, $\frac{3}{2}$)</td>
</tr>
<tr>
<td></td>
<td>$(\frac{3}{4}, \frac{1}{2}, \frac{1}{4})$</td>
<td>8569 ± 9.8</td>
</tr>
<tr>
<td></td>
<td>$(\frac{1}{2}, \frac{3}{10}, \frac{1}{10})$</td>
<td>338.7 ± .35</td>
</tr>
<tr>
<td>3</td>
<td>30; 5</td>
<td>25 (\frac{31}{3}, \frac{1}{3}, \frac{5}{60})</td>
</tr>
<tr>
<td></td>
<td>$(\frac{3}{4}, \frac{3}{5}, \frac{1}{5})$</td>
<td>(4.931 ± 0.022)$\times 10^{14}$</td>
</tr>
<tr>
<td></td>
<td>$(\frac{1}{2}, \frac{4}{10}, \frac{2}{10})$</td>
<td>(4.500 ± 0.025)$\times 10^{11}$</td>
</tr>
<tr>
<td>5</td>
<td>18; 3</td>
<td>15 (\frac{6}{5}, \frac{7}{5}, \frac{1}{5}, \frac{6}{5}, \frac{2}{5})</td>
</tr>
<tr>
<td></td>
<td>$\frac{1}{10}$ (8, 7, 6, 4, 2)</td>
<td>(1.201 ± 0.015)$\times 10^{49}$</td>
</tr>
<tr>
<td></td>
<td>$\frac{1}{10}$ (6, 5, 4, 2, 1)</td>
<td>(2.592 ± 0.028)$\times 10^{36}$</td>
</tr>
<tr>
<td>8</td>
<td>19; 2</td>
<td>17 (\frac{9}{5}, \frac{8}{5}, \frac{7}{5}, \frac{6}{5}, \frac{1}{5}, \frac{4}{5}, \frac{3}{5})</td>
</tr>
<tr>
<td></td>
<td>$\frac{1}{10}$ (9, 1, 2, 3, 1)</td>
<td>(1.895 ± 0.060)$\times 10^{114}$</td>
</tr>
<tr>
<td></td>
<td>$\frac{1}{30}$ (17, 2, 3, 1)</td>
<td>(8.831 ± 0.305)$\times 10^{106}$</td>
</tr>
</tbody>
</table>

Table 1. Assessment of the absolute and relative accuracy of the Laplace approximation in (4).

Similar computations to determine values for $2F_1(m)$ ($\alpha, \beta; \gamma; A, B$) came to virtually the same conclusions as for $1F_1(m)$. Relative accuracy is better than absolute accuracy when the two argument approximation is based upon the single argument approximation $2F_1(\alpha, \beta; \gamma; AB)$ given in (7). Furthermore, the approximation in Glynn and Muirhead (1978) is consistently less accurate particularly in the sub-asymptotic settings.

### 3.2 Testing Noncentrality Rank in MANOVA

We suppose a MANOVA setting with an error matrix $SSE \sim \text{Wishart}_m(n, I)$ independent of treatment error measured by $SSB \sim \text{Noncentral-Wishart}_m(n_1, I, \Omega)$ with $\Omega = \text{diag}(\omega_1, \ldots, \omega_m)$ as the noncentrality matrix and $\omega_1 \geq \cdots \geq \omega_m \geq 0$.

The standard likelihood ratio test for $\Omega = 0$, or equivalently $H_0 : \omega_1 = \omega_2 = \cdots = \omega_m = 0$, rejects for small values of $T_0 = \prod_{i=1}^{m} (1 - u_i)$ where $1 \geq u_1 \geq u_2 \geq \cdots \geq u_m \geq 0$ are the sample eigenvalues of $(SSE + SSB)^{-1} SSB$.

The likelihood ratio test for $H_k : \text{rank}(\Omega) \leq k$ for $k = 1, \ldots, m$, or

$$H_k : \omega_{k+1} = \cdots = \omega_m = 0,$$  \hspace{1cm} (8)
rejects for small values of
\[ T_k = \prod_{i=k+1}^{m} (1 - u_i) \]  
(9)
as discussed in Muirhead (1981, §10.7.4). Following his conditional approach, we
condition \( T_k \) on the observed values for \( u_1, \ldots, u_k \) and use importance sampling
to determine the power function for the test in (8). The conditional p-value
of this test is specified in Theorem 10.7.5 of Muirhead and is based upon the
asymptotic approximation
\[ -\left( n_2 - k + \frac{1}{2} (n_1 - m - 1) + \sum_{i=1}^{k} u_i^{-1}\right) \ln T_k \overset{\text{H}_0}{\sim} \chi^2_{(m-k)(n_1-k)} \]  
(10)
given \( u_1, \ldots, u_k \). The m.l.e. of \( \Omega \), or \( \hat{\Omega} = \text{diag}(\hat{\omega}_1, \ldots, \hat{\omega}_m) \), consists of the eigenvalues of the matrix \( n_2 SSE^{-1}SSB \) and has \( \hat{\omega}_i = n_2 u_i/(1 - u_i) \).

3.2.1 Importance Sampling

The m.l.e. for the p-value of the conditional test, as an alternative to (10),
is approximated using simulation. We simulated \( N = 10^6 \) values of \( T_k \) under
hypothesis \( H_0 \) taking \( \hat{\omega}_1, \ldots, \hat{\omega}_k \) as the true values of the larger unknown eigenvalues of \( \Omega \). With these values of \( T_k \) denoted as \( \{t_{(i)} : i = 1, \ldots, N\} \), then the
m.l.e of the p-value is
\[ \hat{p}_k := N^{-1}\sum_{i=1}^{N} 1 \{ \ln t_{(i)} < \ln t_k^0 \} \],  
(11)
where \( t_k^0 \) is the observed value of \( T_k \).

The attained power function of this test may be determined using impor-
tance sampling. The importance weights make use of the joint density of \( U = \text{diag}(u_1, \ldots, u_m) \) under the alternative hypothesis. This has been given in Muir-
head (1982, §10.7.1) as
\[ f(U; \Omega) \propto 1 \rho_1^{(m)} \left( \frac{n_1 + n_2}{2}; \frac{n_1}{2}; \frac{1}{2} \Omega; U \right) g(U) h(\Omega) , \]
for suitable functions \( g \) and \( h \). We use \( f(U; \hat{\Omega}) \) as our importance distribution to simulate \( N \) values of \( U \), denoted as \( \{U_{(i)}\} \), under alternative \( \hat{\Omega} \) from the two original Wishart matrices \( SSE \) and \( SSB \). These importance samples approximate
the power of the likelihood ratio test at alternative \( \Omega \) as
\[ \hat{p}_{k, \Omega} := \frac{\sum_{i=1}^{N} \hat{\omega}_{(i)} \mathbb{1} \{ \ln t_{(i)} < \ln t_k^0 \}}{\sum_{i=1}^{N} \hat{\omega}_{(i)}} \]  
(12)
where \( t_{(i)} \) is determined from \( U^{(i)} \) as in (9), and the estimated importance weights \( \{ \hat{w}_{(i)} \} \) are the relative values of \( \hat{F}_{1}^{(m)} \) given as

\[
\hat{w}_{(i)} = \frac{\hat{F}_{1}^{(m)} \left( \frac{n_1+n_2}{2}, \frac{n_1}{2}; \frac{1}{2}; \frac{1}{2} \Omega; U_{(i)} \right)}{\hat{F}_{1}^{(m)} \left( \frac{n_1+n_2}{2}, \frac{n_1}{2}; \frac{1}{2}; \frac{1}{2} \Omega; U_{(i)} \right)}.
\]

### 3.2.2 Example

Suppose a 1-way MANOVA with \( m = 4 \), six levels so \( n_1 = 5 \), and 20 replicates/level so \( n_2 = 114 \). Suppose that \( \Omega = \text{diag}(114, 57, 11.4, 0) = n_2 \Theta \) where \( \Theta = \text{diag}(1.0, 0.5, 0.1, 0.0) \). This replicates the approach to asymptotics applied in the 1-way setting in Muirhead (1982, §10.7.3). A value of \( U \) was simulated from this model to serve as the data and gave the value \( U = \text{diag}(0.57, 0.31, 0.08, 0.038) \). From this, \( \hat{\Omega} = \text{diag}(151, 51, 9.9, 4.5) \).

With \( k = 3 \), we test that \( \Omega \) has a rank of at most 3. The Muirhead p-value is 0.0830 and the m.l.e. from (11) has \( \hat{p}_3 = 0.0681 \).

With \( k = 2 \), the Muirhead p-value is 0.0266 and the m.l.e. from (11) has \( \hat{p}_3 = 0.0255 \) again showing good agreement. Importance sampling was used to determine the attained power function of the test against various sorts of alternatives. These alternatives assume that \( (\omega_1, \omega_2) = (\hat{\omega}_1, \hat{\omega}_2) \) and specify the other values as (a) \( (\omega_3, \omega_4) = (t, 0) \); (b) \( (\omega_3, \omega_4) = (t, t/2) \); and (c) \( (\omega_3, \omega_4) = (t, t) \) for \( t = 1(2.5)40 \). Plots for these three power curves versus \( t \) are given in Figure 1 as the solid, dotted and dashed lines respectively. To check their accuracy, direct simulation of these same power functions has been implemented at selected points. These values are shown as circles, diamonds, and crosses. Overall, we see that the determination of power from importance sampling has been quite successful presumably because of the relative accuracy of the importance weights as ratios of \( \hat{F}_{1}^{(m)} \)-values.

### 3.3 Testing the Number of Nonzero Canonical Correlations

The development of tests for the number of useful canonical correlations closely parallels that for tests of noncentrality rank. The main exception is that the densities of sample eigenvalues now involve the two argument \( F_{1}^{(p_0)} \) function instead of \( \hat{F}_{1}^{(m)} \). Suppose that

\[
S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \sim \text{Wishart}_{m} (n, \Sigma)
\]

with \( S_{11} \) as \( p_1 \times p_1 \), \( S_{12} \) as \( p_1 \times p_2 \), etc., \( p_1 + p_2 = m \), and \( p_1 \leq p_2 \). Partition \( \Sigma \) similarly and let

\[
\Omega^2 = \text{diag} (\omega_1^2, \ldots, \omega_{p_1}^2)
\]

contain the population squared canonical correlations as the eigenvalues of \( \Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \) with \( 1 \geq \omega_1^2 \geq \cdots \geq \omega_{p_1}^2 \geq 0 \). The likelihood ratio test of
the null hypothesis that rank(Ω) ≤ k or

\[ H_k : \omega_{k+1} = \omega_{k+2} = \ldots = \omega_{p_1} = 0 \]

for k = 0, ..., p_1 - 1, rejects for small values of

\[ T_k = \prod_{i=k+1}^{p_1} (1 - r_i^2), \quad (13) \]

where 1 ≥ r_1^2 ≥ r_2^2 ≥ ... ≥ r_{p_1}^2 ≥ 0 are the sample eigenvalues of S_{12}^{-1} S_{12} S_{22}^{-1} S_{22}.

Under the null \( H_k \), the asymptotic distribution of \( T \) given \( r_1^2, ..., r_k^2 \) is such that

\[ -\left( n - k - \frac{1}{2} (m + 1) + \sum_{i=1}^{k} r_i^{-2} \right) \ln T \sim \chi_{(p_1-k)(p_2-k)}^2, \quad (14) \]

as given in Theorem 11.3.9 of Muirhead (1982).

### 3.3.1 Importance Sampling

A m.l.e. for the p-value of the conditional test may be found as an alternative to (14). This and all the importance sampling proceeds in exactly the same manner as described with the previous example. The joint density of \( R^2 = \text{diag}(r_1^2, ..., r_{p_1}^2) \) has been given in Muirhead (1982, §11.3.4) as

\[ f(R^2; \Omega^2) \propto 2 F_{1}^{(p_1)} \left( \frac{n}{2}; \frac{n}{2}; \frac{p_1}{2}; \Omega^2; R^2 \right) g(R^2) h(\Omega^2), \]

for suitable functions \( g \) and \( h \). We use \( f(R^2; \hat{\Omega}^2) \) as our importance distribution where \( \hat{\Omega}^2 \) is the observed value of \( R^2 \). To implement the importance sampling, \( N = 10^5 \) values of \( S \) are simulated assuming population canonical correlations given by \( \hat{\Omega} \). From this, squared sample canonical correlations \( \{ R_{(i)}^2 : i = 1, \ldots, N \} \) along with their associated test values of \( T_k \) are computed. The estimated importance weights, used in the power computations of (12) at alternative \( \Omega \), are now

\[ \hat{w}_{(i)} = \frac{2 F_{1}^{(p_1)} \left( \frac{n}{2}; \frac{n}{2}; \frac{p_1}{2}; \hat{\Omega}^2; R_{(i)}^2 \right)}{2 F_{1}^{(p_1)} \left( \frac{n}{2}; \frac{n}{2}; \frac{p_1}{2}; \Omega^2; R_{(i)}^2 \right)}. \]

### 3.3.2 Example

Suppose \( m = 8, p_1 = 4 = p_2, n = 300 \) and let \( \Omega = \text{diag}(0.6, 0.3, 0.1, 0.) \). With this value of \( \Omega \), we simulated a data set and determined its m.l.e. as

\[ R = \text{diag}(0.559, 0.267, 0.1236, 0.07755). \]

With \( k = 3 \), we test that \( \Omega \) has a rank of at most 3. The Muirhead p-value is 0.1330 and the m.l.e. from (11) is \( \hat{p}_3 = 0.1049 \).
With $k = 2$, the Muirhead p-value is 0.1534 and the m.l.e. from (11) is $\hat{p}_2 = 0.1462$. Importance sampling was used to determine the attained power function of the test against the same sort of alternatives as previously considered. Plots for these three power curves are given in Figure 2. The accuracy of the importance sampling was checked using results from direct power simulations and are given as the circles, diamonds, and crosses. For this setting, the determination of power from importance sampling has not shown quite the accuracy seen in Figure 1. Overall, however, it has still been quite successful.

4 Acknowledgements

The authors are grateful to Huiling Le (University of Nottingham) for providing the O’Neill (1966) reference, and to the NSF for financial support under grant DMS-9970785.

References


Figure 1. Attained power functions for the test that the noncentrality rank is two or less.

Figure 2. Attained power functions for the test of two or less nonzero canonical correlations.