Approximation of Power in Multivariate Analysis

Ronald W. Butler and Andrew T.A. Wood
Colorado State University and University of Nottingham

January 4, 2002

Abstract

We consider the calculation of power functions, and the closely related issue of sample size determination, in classical multivariate analysis. In this context, power can be expressed in terms of tail probabilities of certain noncentral distributions. The necessary noncentral distribution theory was developed between the 1940s and 1970s. However, tractable methods for calculating the relevant probabilities have been lacking. In this paper we present simple yet extremely accurate saddlepoint approximations to power functions associated with the following classical test statistics: the likelihood ratio statistic for testing the general linear hypothesis in MANOVA; the likelihood ratio statistic for testing block independence; and Bartlett’s modified likelihood ratio statistic for testing equality of covariance matrices. We also present theoretical results which explain why the approximation is accurate in high dimensions.

1 Introduction.

In this paper we present simple yet extremely accurate approximations to power functions which arise in multivariate analysis. To the best of our knowledge there are no other methods currently available that compute these power functions with the combination of accuracy and speed displayed by the saddlepoint computations presented below.

To set the scene, we review a very elementary example. Consider the standard one-way ANOVA ("analysis of variance") model

\[ y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1, \ldots, I, \quad j = 1, \ldots, J \]  

(1)
where \( \epsilon_{ij} \) are independent and identically distributed (i.i.d.) \( N(0, \sigma^2) \). Consider testing the null hypothesis

\[
H_0 : \quad \alpha_i = 0, \quad i = 1, \ldots, I, \tag{2}
\]

that the \( I \) group means are equal. Under \( H_0 \), the hypothesis sum of squares, \( SS_{\text{hypo}} \), has a \( \sigma^2 \chi^2_{I-1} \) distribution and is independent of the residual sum of squares, \( SS_{\text{res}} \), which has a \( \sigma^2 \chi^2_{(I-1)I} \) distribution. The mean square ratio

\[
F = \frac{SS_{\text{hypo}}/(I-1)}{SS_{\text{res}}/((J-1)I)}
\]

has a theoretical \( F_{I-1,(J-1)I} \) distribution, and the p-value for \( H_0 \) is given by

\[
\Pr \{ F_{I-1,(J-1)I} > f_{\text{obs}} \}
\]

where \( f_{\text{obs}} \) is the observed value of the mean square ratio \( F \). Of course, this probability can be evaluated quickly and easily in most of the widely-used statistical packages.

Calculation of probabilities under the null hypothesis is not the only thing we may wish to do, however. When planning an experiment (and, in particular, when determining a suitable sample size) we may wish to calculate the distribution of the mean square ratio \( F \) under various alternative hypotheses. When \( \alpha_i \neq 0 \), the distribution of \( SS_{\text{res}} \) remains as \( \sigma^2 \chi^2_{(J-1)I} \) and independent of \( SS_{\text{hypo}} \). However, the distribution of \( SS_{\text{hypo}} \) is now \( \sigma^2 \) times a noncentral \( \chi^2_{I-1}(\omega) \) with noncentrality parameter

\[
\omega = J\sigma^{-2} \sum_{i=1}^{I} \alpha_i^2.
\]

The mean square ratio \( F \) is therefore a noncentral \( F_{I-1,(J-1)I}(\omega) \) distribution. Thus the power function of the classical \( F \) test can be expressed in terms of the noncentral \( F \) distribution. However, despite the fundamental importance of this distribution, it has not been particularly easy for applied statisticians to obtain its percentage points. Recently highly accurate saddlepoint approximations have
been made available in Butler and Paolella (2000) to provide tail probabilities for
the noncentral \( F_{j-1,(J-1)J} \). These approximations have relative errors typically
around 0.03% and are rarely much larger than 0.1%.

Model (1) is, of course, just an important special case of the multivariate
linear model. This paper is concerned with power calculations associated with
various hypotheses in the multivariate linear model, including the general linear
hypothesis. The noncentral distribution theory discussed in relation to model
(1) above has a natural (but mathematically sophisticated) generalization to the
multivariate case. This noncentral distribution theory was developed from the
1940s onwards in a sequence of papers by Anderson, James, Constantine, Pillai,
Muirhead and many others. These advances depended on the development on
new tools, especially zonal polynomials (James, 1960; 1961) and hypergeometric
functions of matrix argument (Herz, 1955). An excellent starting point for this
body of work is the monograph by Muirhead (1982); see also Anderson (1984).

Despite these impressive achievements in the theoretical sphere, the mul-
tivariate noncentral distribution theory has unfortunately had relatively little
practical impact. The main reason has been the practical difficulty in actually
computing percentage points from the theoretical expressions that specify the
multivariate noncentral distributions.

We shall use saddlepoint methods to determine these percentage points.
Such methods have already been successfully employed to calculate p-values
from the null distributions of statistics which fall in the Box class (first intro-
duced by Box, 1949). The Box class is characterized as consisting of statistics
whose null moment generating function (MGF) can be expressed as a ratio of
products of gamma functions. A prominent member of this class is the likeli-
hood ratio test statistic for the general linear hypothesis whose null distribution
has been approximated in Srivastava and Yao (1989) and Butler, Huzurbazar
and Booth (1992). For further details of these and other applications in the Box
class see Butler, Booth and Huzurbazar (1993) and Booth, Butler, Huzurbazar

These saddlepoint approximations have proven to be extremely accurate and yet very straightforward to implement, the most complicated step being the solution of the saddlepoint equation (a non-linear equation in a real variable). For practical purposes, the problem of computing the null distribution of these statistics has been solved.

In this paper we discuss approximation of the noncentral distributions of three statistics in the Box class: the likelihood ratio statistic (referred to below as Wilks' statistic) for testing the general linear hypothesis (which is the most important of the three); the likelihood ratio statistic for testing block independence; and Bartlett's M-statistic for testing equality of covariance matrices. The main complication is that the non-null MGF of each of these statistics involves a hypergeometric function of matrix argument. These functions are complicated and have acquired a reputation for being difficult to compute.

Our approach in the non-null case involves two steps: replacing the relevant hypergeometric function of matrix argument in the noncentral MGF with a Laplace approximation, followed by the application of saddlepoint methods to this approximate MGF as in the null case. This two step approach has been referred to as a sequential saddlepoint approximation. The Laplace approximations to the matrix-argument hypergeometric functions $1F_1$ and $2F_1$, and the sequential saddlepoint approximation to Wilks' statistic, were first presented in Butler and Wood (2002a). That paper also contained some numerical results (in Tables 3 and 4) for the sequential saddlepoint approximation to the non-null distribution of Wilks' statistic, focussing on the case in which the eigenvalues of the noncentrality matrix are small to moderate.

We now summarize the main contributions of this paper: further numerical results demonstrating the high accuracy of the approximation for Wilks'
statistic, focussing on cases in which the noncentrality matrix has some large eigenvalues (Section 2); an example illustrating the use of these power calculations in sample size determination (Section 2); comparison with alternative approximations due to Sugiura and Fujikoshi (1969), Sugiura (1973) and Muller et al. (1992) (see Table 1); numerical results which demonstrate the generally high accuracy of the sequential saddlepoint approach applied to the "block independence" statistic and Bartlett $M$ statistic (Section 3); details of how to implement the sequential saddlepoint approach (Section 4; Appendices A and B); and an explanation (in the case of Wilks' statistic) as to why (on theoretical grounds) the approximations may be expected to work well in high dimensions (Section 5).

As mentioned in the previous paragraph, we also give some consideration to the question of how these power calculations can be applied to experimental design in multivariate contexts, and particularly to the problem of sample size determination. See also Muller et al. (1992) for useful discussion of this issue.

It is, of course, the case that all the non-null distributions considered in this paper can be approximated by simulation. Indeed, we have used simulation to check the accuracy of our approximations. However, the saddlepoint approximations can be computed in seconds whereas the times needed to obtain comparable accuracy by simulation ranged from minutes to hours to days (depending on the dimension and sample size). Taking the simulated values (obtained from $10^6$ independent simulations in each case) as the gold standard, the relative errors of the noncentral saddlepoint tail probabilities in the general linear hypothesis were all less that 6% and typically much less. Likewise with block independence the largest relative error was 2% and typically much less. For the Bartlett $M$ test the approximations were somewhat less accurate in the center of the distribution. While we cannot conceivably check all cases, we have performed a sufficient number of comparisons with simulation to conclude that the saddlepoint approach for each of these tests may be trusted, is the pre-
ferred method of computation, and provides more than adequate accuracy for most all practical considerations. Even with very small sample sizes relative to the dimension, the saddlepoint approximations have demonstrated remarkable accuracy when compared with simulation.

Finally, our main practical message is this: the procedures for approximating the power functions considered in this paper have proved extremely accurate and yet straightforward to compute, and we therefore recommend them for general use.

2 Noncentral Distribution for the Likelihood Ratio Statistic in MANOVA

Before going into details, we invite the reader to have an early glance at Table 1 in order to gain an idea of the remarkable accuracy achieved by the approximations proposed in this paper.

2.1 The General Linear Hypothesis

Consider the general linear model

\[ Y = XB + E \]  \hspace{1cm} (3)

where the rows of \( E \) are i.i.d. \( N_p(0, \Sigma) \), \( X \ (N \times q) \) is a known full rank matrix of covariates, \( B \ (q \times p) \) is a matrix of unknown parameters, \( E \ (N \times p) \) is the "error" matrix and \( Y \ (N \times p) \) is the observation matrix. Throughout, we shall assume that \( \Sigma \) is an unknown covariance matrix of full rank.

The general linear hypothesis \( H_{GLM} \) is

\[ H_{GLM} : CB = 0_{m,p} \]  \hspace{1cm} (4)

where \( C \ (m \times q) \) is a rank \( m \) matrix of known constants, and \( 0_{m,p} \) is the \( m \times p \) matrix of zeros. With a suitable choice of \( C \), a wide range of hypotheses can
be represented in the form (4). For example, see Muirhead (1982, p. 432) for a choice of \( C \) which corresponds to the null hypothesis in a one-way MANOVA, which is the multivariate generalization of (2).

It follows from some standard calculations (see e.g. Muirhead, 1982) that the likelihood ratio statistic \( W_{GLM} \) for testing the null hypothesis \( H_{GLM} \), may be expressed as

\[
W_{GLM} = \frac{|SS_{err}|^{N/2}}{|SS_{hypo} + SS_{err}|^{N/2}}
\]

where \( SS_{hypo} = (Y_1^*)^T Y_1^* \) and \( SS_{err} = (Y_3^*)^T Y_3^* \) represent the hypothesis and error sums of squares respectively, as specified in Appendix A in terms of canonical variables \( Y_1^* \) and \( Y_3^* \). Under model (3), the following hold: \( SS_{hypo} \) and \( SS_{err} \) are independent; \( SS_{err} \) has a \( p \)-dimensional central Wishart distribution with \( n = N - q \) degrees of freedom and shape matrix \( \Sigma \), denoted as \( \text{Wishart}_p(n, \Sigma) \); and \( SS_{hypo} \) has a \( p \)-dimensional non-central Wishart distribution with \( m \) degrees of freedom, shape matrix \( \Sigma \), and non-centrality matrix \( \Omega = \Sigma^{-1} M_1^T M_1 \), where \( M_1 = E(Y_1^*) \), and \( M_1 = 0_{m,p} \) if and only if \( CB = 0_{m,p} \).

2.2 Numerical Results for a 1-Way MANOVA

We now consider using the sequential saddlepoint approximation for the noncentral distribution of \( \log \Lambda_{GLM} = \frac{1}{2} \log W_{GLM} \), the log of Wilks' \( \Lambda_{GLM} \) statistic. This approximation is computed by taking a saddlepoint inversion of a Laplace approximation for the MGF of \( \log \Lambda_{GLM} \) as given in Section 4.1 (see formula (10)).

Consider a balanced 1-way MANOVA design in \( p \) dimensions with \( I \) levels and \( J \) repetitions. The model is given by (1), but with all quantities now \( p \)-vectors rather than scalars. We wish to test a hypothesis of the form (4) corresponding to the null hypothesis in a one-way MANOVA (see the comment below (4)), with \( m = I - 1 \) degrees of freedom for hypothesis and \( n = I(J - 1) \) degrees of freedom for error. The noncentrality matrix in Wilks' test is given
<table>
<thead>
<tr>
<th>$(p, n, m)$</th>
<th>$(I, J)$</th>
<th>$\Omega = J\Omega_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(3, 24, 7)</td>
<td>(8, 4)</td>
<td>$\Omega_1 = \text{diag}{2, 4, 7}$</td>
</tr>
<tr>
<td>Lug. &amp; Rice</td>
<td>.9903</td>
<td>4.966</td>
</tr>
<tr>
<td>$O(n^{-3})$</td>
<td>$\chi^2$</td>
<td>24.74</td>
</tr>
<tr>
<td>Noncent. $F$</td>
<td>.7662</td>
<td>4.000</td>
</tr>
<tr>
<td>$O\left(n^{-3/2}\right)$</td>
<td>.0255</td>
<td>.6665</td>
</tr>
<tr>
<td>(7, 24, 7)</td>
<td>(8, 4)</td>
<td>$\Omega_1 = \text{diag}{1(1)}$</td>
</tr>
<tr>
<td>Lug. &amp; Rice</td>
<td>.9633</td>
<td>4.883</td>
</tr>
<tr>
<td>$O(n^{-3})$</td>
<td>$\chi^2$</td>
<td>-60.4</td>
</tr>
<tr>
<td>Noncent. $F$</td>
<td>.1655</td>
<td>1.213</td>
</tr>
<tr>
<td>$O\left(n^{-3/2}\right)$</td>
<td>.03715</td>
<td>.0394</td>
</tr>
<tr>
<td>(7, 24, 7)</td>
<td>(8, 4)</td>
<td>$\Omega_1 = \text{diag}{2(5)}$</td>
</tr>
<tr>
<td>Lug. &amp; Rice</td>
<td>.8940</td>
<td>4.475</td>
</tr>
<tr>
<td>Noncent. $F$</td>
<td>.0220</td>
<td>.2384</td>
</tr>
<tr>
<td>$O\left(n^{-3/2}\right)$</td>
<td>.0242</td>
<td>.4986</td>
</tr>
<tr>
<td>(12, 56, 7)</td>
<td>(8, 8)</td>
<td>$\Omega_1 = {5(0), 1(1)}$</td>
</tr>
<tr>
<td>Lug. &amp; Rice</td>
<td>.9847</td>
<td>4.946</td>
</tr>
<tr>
<td>Noncent. $F$</td>
<td>.0947</td>
<td>.7434</td>
</tr>
<tr>
<td>(25, 56, 7)</td>
<td>(8, 8)</td>
<td>$\Omega_1 = {18(0), 1(1)}$</td>
</tr>
<tr>
<td>Lug. &amp; Rice</td>
<td>1.007</td>
<td>4.979</td>
</tr>
<tr>
<td>(25, 56, 7)</td>
<td>(8, 8)</td>
<td>$\Omega_1 = {18(0), 2(5)}$</td>
</tr>
<tr>
<td>Lug. &amp; Rice</td>
<td>1.007</td>
<td>4.979</td>
</tr>
<tr>
<td>(50, 88, 7)</td>
<td>(8, 12)</td>
<td>$\Omega_1 = {43(0), 1(1)}$</td>
</tr>
<tr>
<td>Lug. &amp; Rice</td>
<td>.9938</td>
<td>5.013</td>
</tr>
</tbody>
</table>

Table 1. One-way MANOVA. Sequential saddlepoint approximations (Lug.
\& Rice), the \( O(n^{-3}) \) expansions of Sugiura and Fujikoshi (1969), the noncentral \( F \) approximations of Muller and Peterson (1984), and the \( O(n^{-3/2}) \) expansions of Sugiura (1973) for the noncentral CDF of Wilks' \( \log \Lambda_{GLM} \) in the general linear MANOVA hypothesis. The table entries are the CDF approximations evaluated at the simulated empirical percentiles listed in Table 2. Accuracy may be judged by comparing the table entries to the true percentile levels given in the top row.

by

\[ \Omega = J \sum^{-1} \sum_{i=1}^{I} \alpha_i \alpha_i^T := J \Omega_1 \]  

(5)

and increases with \( J \) where \( \Omega_1 \) is that portion which is fixed.

The accuracy of the sequential saddlepoint is shown in Table 1 for various settings of the balanced one-way MANOVA. It displays percentages that measure saddlepoint accuracy in terms of how close the entries are to the true percentages listed in the top row. The approximations demonstrate remarkable accuracy with the entries showing two and often three significant digit accuracy. If \( 100 \hat{p}_i \) denotes the \( i \)th table entry with \( 100p_i \) as its target, then the largest percentage absolute relative error is

\[ 100 \% \times \max_i \left\{ \frac{|\hat{p}_i - p_i|}{\min(p_i, 1 - p_i)} \right\} \leq 6\% \]

when maximized over the table. Most of the relative errors are considerably smaller and not much more than 1\%. This is quite remarkable accuracy when the range of percentiles, range of dimensionalities, and the difficulties associated with uninformative data (i.e. cases in which the sample size \( n \) is not much larger than the dimension \( p \)) are all taken into account. A closer examination reveals that the greatest accuracy occurs with the smaller entries in \( \Omega_1 \) which corresponds to power computations that are not widely discrepant from the null hypothesis. The least accuracy is seen when using the larger entries in \( \Omega_1 \) which deal with power considerations that are far from the null hypothesis. Table entries for the saddlepoint approximation were determined in the following way. Empirical quantiles for the noncentral distribution of \( \log \Lambda_{GLM} \) associated with the probabilities in the top row were determined by simulating \( 10^6 \) independent
values of $\log A_{GLM}$. These empirical quantiles are displayed in Table 2. The sequential saddlepoint approximations were evaluated at these empirical percentiles so that Table 1 shows the accuracy of the saddlepoint approximations were the empirical percentiles regarded as exact. Any saddlepoint inaccuracy as seen in Table 1 could easily be attributed to sample variation in the empirical quantiles, so Table 1 represents perhaps the greatest accuracy that could be determined from simulation studies of this sort.

Table 1 also shows the same sort of computations to determine the comparable accuracy for other power approximations discussed in the literature. These include the noncentral $\chi^2$ expansions of Sugiura and Fujikoshi (1969) with error $O(n^{-3})$, the noncentral $F$ approximation introduced in Muller and Peterson (1984) and numerically evaluated in Muller et al. (1992), and the expansions of Sugiura (1973) with error $O(n^{-3/2})$. None of these approximations comes close to the accuracy and consistent reliability of the saddlepoint method when considered across the range of parameters involved. The expansions of Kulp and Nagarsenkar (1984) are not shown but were computed and found to be very inaccurate in the first setting with $p = 3$. Terms in this expansion of noncentral-Beta values to order $O(n^{-2})$ are incorrectly stated in the paper and therefore we were unable to compute it to its greatest "accuracy"; when computed using terms up to order $O(n^{-1})$, all the percentages that should range from 1% to 99% were negative; and when only the leading positive term was computed, all the percentages were less than 0.0102%.

There was a substantial difference in computational time between the saddlepoint approximations and simulation. All the saddlepoint approximations were virtually instantaneous. By contrast, simulation times ranged from 4 minutes for $p = 3$ up to 22 hours for the setting $p = 50$. These simulations were performed in Fortran 90 on a 300 MHz PC.

Table 2 also shows the approximate mean $\bar{\mu}$ and standard deviation $\bar{\sigma}$ for the noncentral distribution of $\log A_{GLM}$ as computed from $\hat{K}(s)$, the Laplace
approximation to its cumulant generating function (CGF). These cumulants were computed as

\[ \hat{\mu} = \hat{K}'(0) \quad \hat{\sigma} = \sqrt{\hat{K}''(0)}. \]

For comparison, the empirical averages and standard deviations of \( \log \Lambda_{GLM} \) from the \( 10^6 \) simulated values are given in parentheses. They show virtually no

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>5</th>
<th>10</th>
<th>30</th>
<th>.50</th>
<th>.70</th>
<th>.90</th>
<th>.95</th>
<th>.99</th>
</tr>
</thead>
<tbody>
<tr>
<td>( (p, n, m) = (3, 24, 7) )</td>
<td>( \hat{\mu} = -2.184 ) (-2.185)</td>
<td>( \hat{\sigma} = .4260 ) (.4264)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-3.271</td>
<td>-2.920</td>
<td>-2.742</td>
<td>-2.391</td>
<td>-2.162</td>
<td>-1.946</td>
<td>-1.654</td>
<td>-1.523</td>
<td>-1.292</td>
<td></td>
</tr>
<tr>
<td>( (p, n, m) = (7, 24, 7) )*</td>
<td>( \hat{\mu} = -5.019 ) (-5.025)</td>
<td>( \hat{\sigma} = .6693 ) (.6718)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( (p, n, m) = (7, 24, 7) )**</td>
<td>( \hat{\mu} = -9.910 ) (-9.946)</td>
<td>( \hat{\sigma} = .7737 ) (.7814)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( (p, n, m) = (12, 56, 7) )</td>
<td>( \hat{\mu} = -4.413 ) (-4.415)</td>
<td>( \hat{\sigma} = .4272 ) (.4272)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( (p, n, m) = (25, 56, 7) )*</td>
<td>( \hat{\mu} = -6.684 ) (-6.685)</td>
<td>( \hat{\sigma} = .5452 ) (.5548)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( (p, n, m) = (25, 56, 7) )**</td>
<td>( \hat{\mu} = -11.42 ) (-11.42)</td>
<td>( \hat{\sigma} = .6061 ) (.6063)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( (p, n, m) = (50, 88, 7) )</td>
<td>( \hat{\mu} = -8.379 ) (-8.379)</td>
<td>( \hat{\sigma} = .5171 ) (.5264)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Empirical percentiles (top) and saddlepoint percentiles (bottom; not shown when identical to table accuracy) for noncentral \( \log \Lambda_{GLM} \). The empirical percentiles were determined from \( 10^6 \) independent simulations of \( \log \Lambda_{GLM} \). 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. Superscripts * and ** denote respectively the settings with the smaller and larger entries in \( \Omega_1 \).
differences but entail much more computational effort. The value of \( \hat{\mu} \) is quite close to the simulated median suggesting that the distributions are not particularly skewed. In fact the mean and median appear to coalesce as \( p \) increases, as predicted by the theoretical results in Section 5.

2.3 Sample Size Determination

Consider a balanced 1-way MANOVA design as previously described in \( p = 5 \) dimensions with \( I = 8 \) levels and \( J \) repetitions per level where value \( J \) has yet to be determined. We wish to test the null hypothesis of equality of means across levels with \( m = 7 \) degrees of freedom for hypothesis and \( n = 8(J - 1) \) degrees of freedom for error. The noncentrality matrix in Wilks' test is given in (5) by \( \Omega = J \Omega_1 \) and increases with \( J \) where \( \Omega_1 \) is the portion which is fixed.

For a given choice of \( \Omega_1 \neq 0_{p,p} \), we are now able to solve the following design question. What is the minimal value of \( J \) for which the 5% level test of the null hypothesis in a one-way MANOVA achieves 90% power at \( \Omega_1 \)? The power of the test depends only upon the eigenvalues of \( \Omega_1 \) and \( J \). We first consider alternatives with equal eigenvalues scaled so that

\[
\Omega_1 = \omega^2 I_5 \quad \omega = 1, \frac{3}{4}, \frac{1}{2}, \frac{1}{4},
\]

where \( \omega \) measures the common distance from the null setting in standard units. Figure 1 plots the percentage power at \( \Omega_1 \) versus \( J \) for the 5% level test. Reading from the raw scores underlying the plots, the design question is solved with replicate sizes of \( J = 8, 13, 27 \), and 105 with the four respective values of \( \Omega_1 \) in decreasing order of \( \omega \). For the first three, the power at \( J = 8, 13, \) and 27 just exceeds 90% as may be seen in the top three plots of Figure 1.

The determination of this plot requires two steps. First, the MGF of \( \log \Lambda_{GLM} \) in the null setting is used to determine a saddlepoint approximation to the null CDF

\[
G_0(\lambda) := \Pr (\log \Lambda_{GLM} \leq \lambda | \Omega = 0_{p,p})
\]
which we denote as $\hat{G}_0(\lambda)$. An IMSL rootfinder approximates the 5th percentile as $\lambda_0 := \hat{G}_0^{-1}(0.05)$. Secondly, an approximation to the noncentral MGF of $\log \Lambda_{\text{GLM}}$ is used in conjunction with a saddlepoint approximation to determine a sequential saddlepoint approximation to the noncentral CDF

$$ G_{\Omega_1}(\lambda, J) := \Pr\left(\log \Lambda_{\text{GLM}} \leq \lambda | \Omega = J\Omega_1 \right) $$

which is denoted as $\hat{G}_{\Omega_1}(\lambda, J)$. Figure 1 contains the plots of

$$ 100\hat{G}_{\Omega_1}(\lambda_0, J) = 100\hat{G}_{\Omega_1}\left(\hat{G}_0^{-1}(0.05), J\right) \text{ vs. } J = 2, 3, \ldots $$

for the four $\Omega_1$-values given in (6).

![Graph](image)

**Fig. 1.** Power of Wilks' test for the 1-way MANOVA versus replicates/level $J$ with $p = 5$ and $I = 8$. The eigenvalues of $\Omega$ are all equal and scaled to be 1 (solid), $(3/4)^2$ (dashed), $(1/2)^2$ (dotted), and $(1/4)^2$ (dashed-dotted).

The equal eigenvalue setting of Figure 1 is the context in which the Bartlett-Nanda-Pillai trace statistic is expected to be most powerful among the four
commonly used tests (Carter, 1989). Therefore the power plots of Wilks' $\Lambda_{GLM}$ in Figure 1 might serve as a lower bound for the power of this alternative test.

The power achieved by Wilks' test against one-dimensional alternatives is considerably less as may be seen in Figure 2.

![Figure 2](image)

**Fig. 2.** Power of Wilks' test for the 1-way MANOVA versus replicates/level $J$ with $p = 5$ and $I = 8$. The eigenvalues of $\Omega$ are all zero except for one which is scaled to the value $2^2$ (solid), 1 (dashed), $(3/4)^2$ (dotted), $(1/2)^2$ (dashed-1-dotted), and $(1/4)^2$ (dashed-2-dotted).

This figure plots the approximate power of the 5% level test as $100\hat{C}_{\Omega_1}(\lambda_0, J)$ versus $J = 2, 3, \ldots$ for the five $\Omega_1$-values

$$\Omega_1 = \omega^2 (1, 0, \ldots, 0) \quad \omega = 2, 1, \frac{3}{4}, \frac{1}{2}, \frac{1}{4}. \quad (7)$$

Replicate values of $J = 19, 35, 60, 131, 518$ are the minimal values of $J$ providing 90% power for the 5% level test at the five alternative values of $\Omega_1$ in (7). Roy's test is expected to provide the greatest power against such alternatives (Carter, 1989) and the plots of Figure 2 provide lower bounds for its power function.
The small irregularity in the plot with $\omega = 1/2$ (dot-dashed) at $J = 43$ occurs because of numerical instability of the non-null saddlepoint approximation near its mean. This is prone to happening when the power curve is very shallow and successive saddlepoints, with increasing $J$, are closely packed together as they pass through zero from negative to positive values.

Figure 3 provides power plots for Wilks’ test at level 5% against $\Omega_1$ matrices with unequal eigenvalues given by

$$\Omega_1 = \omega \left( \left(\frac{5}{4}\right)^2, 1, \left(\frac{3}{4}\right)^2, \left(\frac{1}{2}\right)^2, \left(\frac{1}{4}\right)^2 \right) \quad \omega = 1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}. \quad (8)$$

Replicate values of $J = 11, 20, 39, 77, 152$ assure 90% power for the 5% level Wilks’ test at the five values of $\Omega_1$ in (8).

Fig. 3. Power of Wilks’ test for the 1-way MANOVA versus replicates/level $J$ with $p = 5$ and $I = 8$. The eigenvalues of $\Omega$ are proportional to $(5/4)^2, 1, (3/4)^2, (1/2)^2, (1/4)^2$ and scaled by $1$ (solid), $1/2$ (dashed), $1/4$ (dotted), $1/8$ (dashed-1-dotted), and $1/16$ (dashed-2-dotted).
2.4 One-way MANOVA Example: Ringneck Parrots

The flight calls of ringneck parrots near Perth, Western Australia have been analyzed by Baker (1999). He is interested in determining if there are significant differences in the sonograms of parrots from \( I = 4 \) regions that result from different "dialects" of flight calls. Sonograms of \( J = 10 \) parrots were recorded from each region and \( p = 7 \) distinguishing characteristics for each of the sonograms were measured. These measurements included first and second note characteristics for frequency range, mid-frequency, and duration. A seventh variable is the interval between the two notes. Altogether the data are 40 observations of 7 distinguishing characteristics with 10 per region. The researcher is returning to Perth to collect further data and wishes to know what replicate sizes \( J \) will be required to return 90% power using 5% level tests.

This data may be used to estimate the alternative eigenvalues \( \Omega_1 \). The expected sums of squares are

\[
E(SS_{err}) = 36\Sigma
\]

\[
E(SS_{hyp}) = 3\Sigma + 10 \sum_{i=1}^{4} \alpha_i \alpha_i^T = \Sigma (3I_7 + 10\Omega_1)
\]

and we estimate the nonzero eigenvalues of \( \Omega_1 \) using the positive eigenvalues of

\[
3.6(SS_{err})^{-1} SS_{hyp} = .3I_7.
\]

We may thus take

\[
\hat{\Omega}_1 (7 \times 7) = \text{diag} (25.45, 3.968, 1.177, 0, 0, 0, 0).
\]  \hspace{1cm} (9)

Table 3 lists the percentage power achieved by the 5% level test at the replicate values in the top row for the alternative values \( \omega \hat{\Omega}_1 \) with \( \omega = 1, 1/2, \) and \( 1/4 \). The test is not possible with two replicates per level since then \( n = 4 < 7 = p \) and the estimate for the error matrix is not full rank.

A stepwise discriminant analysis of the 7 variables for region enters three variables in the following order: first note duration, interval between notes, and
first note frequency range. From this analysis, we therefore chose to limit the MANOVA test to these three variables with the hope of further increasing the power. This results in \( p = 3 \) with an estimated noncentrality matrix

\[
\tilde{\Omega}_1 (3 \times 3) = \text{diag}(23.01, 1.672, 3006).
\]

Again the power attained by the 5% level test is listed in Table 3 for alternative values \( \omega \tilde{\Omega}_1 \) with \( \omega = 1, 1/2, \) and \( 1/4 \). The table shows the increased power achieved with the dimension reduction. Based upon this power study, we suggest that six replicates per region would be more than adequate for the researcher's purposes.

<table>
<thead>
<tr>
<th>Eigenvalues</th>
<th>( J = 2 )</th>
<th>( J = 3 )</th>
<th>( J = 4 )</th>
<th>( J = 5 )</th>
<th>( J = 6 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Using all 7 responses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \tilde{\Omega}_1 (7 \times 7) )</td>
<td>—</td>
<td>53.80</td>
<td>98.71</td>
<td>99.99</td>
<td>100.00</td>
</tr>
<tr>
<td>( \frac{1}{2} \tilde{\Omega}_1 (7 \times 7) )</td>
<td>—</td>
<td>29.76</td>
<td>80.83</td>
<td>97.48</td>
<td>99.81</td>
</tr>
<tr>
<td>( \frac{1}{4} \tilde{\Omega}_1 (7 \times 7) )</td>
<td>—</td>
<td>16.85</td>
<td>47.14</td>
<td>73.60</td>
<td>89.43</td>
</tr>
<tr>
<td>Using best 3 discriminating responses</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \tilde{\Omega}_1 (3 \times 3) )</td>
<td>45.32</td>
<td>97.92</td>
<td>99.99</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>( \frac{1}{2} \tilde{\Omega}_1 (3 \times 3) )</td>
<td>26.17</td>
<td>79.10</td>
<td>97.40</td>
<td>99.82</td>
<td>100.00</td>
</tr>
<tr>
<td>( \frac{1}{4} \tilde{\Omega}_1 (3 \times 3) )</td>
<td>15.67</td>
<td>47.26</td>
<td>75.12</td>
<td>90.69</td>
<td>97.10</td>
</tr>
</tbody>
</table>

Table 3. Powers for the 5% level test as they vary with replicates per level \( J \).

3 Two Further Test Statistics

In this section we consider the numerical approximation of power functions for two further multivariate test statistics: the likelihood ratio test for block
independence, and the Bartlett $M$-statistic for testing equality of covariance matrices.

3.1 Noncentral Distribution for the Likelihood Ratio Statistic of Block Independence

Consider a random sample $x_1, \ldots, x_N$ of vectors from a multivariate normal population $N_p(\mu, \Sigma)$. Suppose that the observation vector splits into two components of dimensions $p_1$ and $p_2$, with $p_1 \leq p_2$ and $p_1 + p_2 = p$. Suppose that the two components correspond to the block representation

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

where $\mu_i$ is of dimension $p_i$, and $\Sigma_{ij}$ is of dimension $p_i \times p_j$. The null hypothesis of block independence (BI) may be specified as

$$H_{BI} : \Sigma_{12} (= \Sigma^{T}_{21}) = 0_{p_1,p_2} \quad \text{with} \quad \Sigma_{11}, \Sigma_{22} \quad \text{and} \quad \mu \quad \text{unrestricted},$$

where $0_{p_1,p_2}$ is the $p_1 \times p_2$ matrix of zeros.

Let $A$ denote the sample covariance matrix based on $x_1, \ldots, x_N$ with $n = N - 1$ degrees of freedom. Specify $A$ in block form as

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

where $A_{ij}$ is $p_i \times p_j$. Then it can be shown (see Muirhead, 1982, §11.2) that the likelihood ratio test for block independence rejects the null hypothesis $H_{BI}$ for small values of

$$\Lambda_{BI} = \frac{|A|}{|A_{11}| |A_{22}|}.$$ 

It turns out that the noncentral distribution of $\Lambda_{BI}$ (i.e. the distribution of $\Lambda_{BI}$ when $\Sigma_{12} \neq 0_{p_1,p_2}$) is determined by the quantities $n$, $p_1$, $p_2$ and $P = \text{diag}\{\rho_1, \ldots, \rho_{p_1}\}$ where $\rho_1, \ldots, \rho_{p_1}$ are the eigenvalues of $\Sigma_{11}^{-1} \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$. The noncentral moments of $\Lambda_{BI}$ are given in Section 4.1 (see formula (11)).
Table 4 displays sequential saddlepoint approximations for the noncentral CDF of $\log \Lambda_{B1}$ and has similar structure and interpretation as Table 1. The entries are the sequential saddlepoint approximation evaluated at empirical quantiles for the listed percentage levels. We see remarkable accuracy with the largest percentage relative error as 2.1%. The empirical percentiles were determined by simulating $10^6$ independent values of $\log \Lambda_{B1}$ and are displayed in Table 5.

<table>
<thead>
<tr>
<th>$(p_1,p_2,n)$</th>
<th>1</th>
<th>5</th>
<th>10</th>
<th>30</th>
<th>50</th>
<th>70</th>
<th>90</th>
<th>95</th>
<th>99</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(2,3,10)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lug. &amp; Rice</td>
<td>.9883</td>
<td>4.965</td>
<td>9.936</td>
<td>29.90</td>
<td>49.89</td>
<td>69.88</td>
<td>89.95</td>
<td>94.98</td>
<td>99.00</td>
</tr>
<tr>
<td>$O(n^{-3/2})$</td>
<td>5.295</td>
<td>24.85</td>
<td>41.63</td>
<td>77.05</td>
<td>92.08</td>
<td>97.99</td>
<td>98.22</td>
<td>97.56</td>
<td>97.01</td>
</tr>
<tr>
<td>Normal</td>
<td>.4320</td>
<td>4.292</td>
<td>10.05</td>
<td>32.68</td>
<td>52.52</td>
<td>70.15</td>
<td>86.92</td>
<td>91.51</td>
<td>96.23</td>
</tr>
<tr>
<td>$O(n^{-2})$</td>
<td>1.836</td>
<td>8.617</td>
<td>16.31</td>
<td>42.92</td>
<td>65.78</td>
<td>85.81</td>
<td>101.2</td>
<td>103.1</td>
<td>102.0</td>
</tr>
<tr>
<td>Noncent. $\chi^2$</td>
<td>2.335</td>
<td>8.494</td>
<td>13.74</td>
<td>30.63</td>
<td>46.15</td>
<td>62.62</td>
<td>82.97</td>
<td>89.79</td>
<td>97.10</td>
</tr>
</tbody>
</table>

| $(5,7,20)$    |      |      |      |      |      |      |      |      |      |
| Lug. & Rice   | .9915| 4.977| 9.968| 30.02| 50.03| 70.04| 90.02| 95.00| 99.00|
| $O(n^{-3/2})$ | 11.31| 56.46| 99.49| 182.6| 184.6| 140.2| 71.57| 57.33| 60.12|
| Normal        | .2389| 2.094| 5.104| 20.00| 37.26| 56.60| 79.74| 87.15| 95.26|
| Noncent. $\chi^2$ | 4.712| 13.41| 21.01| 42.97| 60.20| 75.66| 90.83| 94.91| 98.67|

| $(10,13,40)$  |      |      |      |      |      |      |      |      |      |
| Lug. & Rice   | .9760| 4.953| 9.957| 29.91| 49.95| 69.99| 90.03| 95.01| 99.00|

| $(25,28,70)$  |      |      |      |      |      |      |      |      |      |
| Lug. & Rice   | .9903| 5.004| 9.990| 29.96| 49.97| 69.99| 90.05| 95.01| 98.99|

Table 4. Sequential saddlepoint approximation to the CDF of the noncentral distribution for $\log \Lambda_{B1}$. The "Lug. & Rice" table entries are the CDF approximation evaluated at the simulated empirical percentiles listed in Table 6. Accuracy is judged by the closeness of the table entries to the true percentile levels in the top row.

Table 4 compares these saddlepoint approximations to four other approxi-
mations for case $p_1 = 2, 5$. Sugiura and Fujikoshi (1969) suggest a $O(n^{-3/2})$
expansion based on the central limit tendency of $\ln \Lambda_{B1}$ which is denoted as
$O(n^{-3/2})$. Values for the leading normal term in this expansion are listed as
"Normal" and are seen to be much more accurate than the expansion itself. Lee
(1971), Muirhead (1972), and Sugiura (1973) proposed a $O(n^{-3})$ expansion
under local alternatives in which $P \to 0$ which is denote as $O(n^{-3})$. Such local
approximations should not be expected to be accurate here as may be seen in
the table. The leading terms in these expansions are denoted as "noncentral $\lambda^2$"
and demonstrate greater accuracy. None of these alternative approximations ap-
pear capable of delivering the accurate probabilities seen with the saddlepoint
approximation.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>5</th>
<th>10</th>
<th>30</th>
<th>50</th>
<th>70</th>
<th>90</th>
<th>95</th>
<th>99</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(p_1, p_2, n) = (2, 3, 10)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\mu}$</td>
<td>-1.985</td>
<td>-1.985</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\sigma}$</td>
<td>.7525</td>
<td>.7521</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\mu}$</td>
<td>-3.328</td>
<td>-2.980</td>
<td>-2.322</td>
<td>-1.916</td>
<td>-1.547</td>
<td>-1.076</td>
<td>-0.8775</td>
<td>-0.5565</td>
<td></td>
</tr>
<tr>
<td>$\hat{\sigma}$</td>
<td>-4.043</td>
<td>-3.325</td>
<td>-2.976</td>
<td>-2.319</td>
<td>-1.913</td>
<td>-1.544</td>
<td>-1.074</td>
<td>-0.8764</td>
<td>-0.5570</td>
</tr>
<tr>
<td>$(p_1, p_2, n) = (5, 7, 20)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\mu}$</td>
<td>-5.284</td>
<td>-5.283</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\sigma}$</td>
<td>.9098</td>
<td>.9104</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\mu}$</td>
<td>-6.852</td>
<td>-6.475</td>
<td>-5.727</td>
<td>-5.240</td>
<td>-4.777</td>
<td>-4.149</td>
<td>-3.865</td>
<td>-3.359</td>
<td></td>
</tr>
<tr>
<td>$(p_1, p_2, n) = (10, 13, 40)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\mu}$</td>
<td>-11.52</td>
<td>-11.52</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\sigma}$</td>
<td>.8711</td>
<td>.8730</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\sigma}$</td>
<td>-13.63</td>
<td>-12.98</td>
<td>-11.96</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$(p_1, p_2, n) = (25, 28, 70)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\mu}$</td>
<td>-32.66</td>
<td>-32.66</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\sigma}$</td>
<td>-35.57</td>
<td>-34.69</td>
<td>-34.23</td>
<td>-33.29</td>
<td>-32.64</td>
<td>-32.01</td>
<td>-31.10</td>
<td>-30.67</td>
<td>-29.89</td>
</tr>
</tbody>
</table>

Table 5. Empirical percentiles (top) and saddlepoint percentiles (bottom; not
shown when identical to table accuracy) for the noncentral distribution of $\log \Lambda_{B1}$. The
empirical percentiles were determined from $10^6$ independent simulations of $\log \Lambda_{B1}$. The
case $p_1 = 25$ 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 with simulation.
3.2 Noncentral Distribution for Bartlett’s M Statistic

Let \( x_1, \ldots, x_{N_1} \) be a random sample from a \( N_p(\mu_1, \Sigma_1) \) population, and let \( y_1, \ldots, y_{N_2} \) be a random sample from a \( N_p(\mu_2, \Sigma_2) \) population. The null hypothesis of equal covariance matrices (ECM) may be written

\[
H_{ECM} : \Sigma_1 = \Sigma_2 = \Sigma, \text{ with } \Sigma, \mu_1 \text{ and } \mu_2 \text{ unrestricted.}
\]

The Bartlett M test (or the modified likelihood ratio test) rejects \( H_{ECM} \) for small values of

\[
\Lambda_{ECM} = \frac{|A_1|^{n_1/n}|A_2|^{n_2/n}}{|A_1 + A_2|}
\]

<table>
<thead>
<tr>
<th>( (p, n_1, n_2) )</th>
<th>1</th>
<th>5</th>
<th>10</th>
<th>30</th>
<th>50</th>
<th>70</th>
<th>90</th>
<th>95</th>
<th>99</th>
</tr>
</thead>
<tbody>
<tr>
<td>(3, 5, 8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lug. &amp; Rice</td>
<td>1.022</td>
<td>5.033</td>
<td>10.04</td>
<td>29.81</td>
<td>49.65</td>
<td>69.60</td>
<td>89.86</td>
<td>94.95</td>
<td>99.00</td>
</tr>
<tr>
<td>Norm. ( O\left(n^{-3/2}\right) )</td>
<td>0.4508</td>
<td>0.8000</td>
<td>0.9811</td>
<td>19.92</td>
<td>53.08</td>
<td>86.37</td>
<td>112.0</td>
<td>114.7</td>
<td>111.1</td>
</tr>
<tr>
<td>( \chi^2 ) to ( O\left(n^{-3/2}\right) )</td>
<td>4.049</td>
<td>3.113</td>
<td>6.921</td>
<td>23.47</td>
<td>41.75</td>
<td>61.90</td>
<td>85.41</td>
<td>92.26</td>
<td>98.31</td>
</tr>
<tr>
<td>(8, 15, 20)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lug. &amp; Rice</td>
<td>1.114</td>
<td>5.341</td>
<td>10.46</td>
<td>30.56</td>
<td>50.43</td>
<td>70.21</td>
<td>90.04</td>
<td>94.99</td>
<td>98.99</td>
</tr>
<tr>
<td>( \chi^2 ) to ( O\left(n^{-3/2}\right) )</td>
<td>1.022</td>
<td>0.6798</td>
<td>1.618</td>
<td>7.402</td>
<td>16.93</td>
<td>32.36</td>
<td>61.73</td>
<td>75.20</td>
<td>93.28</td>
</tr>
<tr>
<td>(16, 22, 27)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lug. &amp; Rice</td>
<td>1.257</td>
<td>5.866</td>
<td>11.43</td>
<td>32.39</td>
<td>52.43</td>
<td>71.93</td>
<td>90.75</td>
<td>95.40</td>
<td>99.09</td>
</tr>
<tr>
<td>(16, 50, 60)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lug. &amp; Rice</td>
<td>1.117</td>
<td>5.400</td>
<td>10.66</td>
<td>31.08</td>
<td>51.08</td>
<td>70.82</td>
<td>90.34</td>
<td>95.18</td>
<td>99.06</td>
</tr>
<tr>
<td>(32, 120, 125)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lug. &amp; Rice</td>
<td>1.170</td>
<td>5.654</td>
<td>11.00</td>
<td>31.81</td>
<td>51.89</td>
<td>71.47</td>
<td>90.69</td>
<td>95.41</td>
<td>99.08</td>
</tr>
</tbody>
</table>

Table 6. Sequential saddlepoint approximation to the CDF of the noncentral distribution for \( \log \Lambda_{ECM} \). Table entries are the CDF approximation evaluated at the simulated empirical percentiles listed in Table 7. Accuracy is judged by the closeness of the table entries to the true percentile levels in the top row.
where \( n_i = N_i - 1 \), \( n = n_1 + n_2 \),

\[
A_1 = \sum_{i=1}^{N_1} (x_i - \bar{x}) (x_i - \bar{x})^T, \quad A_2 = \sum_{j=1}^{N_2} (y_j - \bar{y}) (y_j - \bar{y})^T,
\]

and \( \bar{x} \) and \( \bar{y} \) are the respective sample means.

The non-central distribution of \( \LambdaECM \) (i.e. the distribution of \( \LambdaECM \) when \( \Sigma_1 \neq \Sigma_2 \)) is determined by the quantities \( p, n_1, n_2 \) and \( \Delta = \text{diag} \{ \delta_1, \ldots, \delta_p \} \) where \( \delta_1, \ldots, \delta_p \) are the eigenvalues of \( \Sigma_1 \Sigma_2^{-1} \). The non-central moments of \( \LambdaECM \) are given in Section 4.1 (see formula (12)).

\[
\begin{array}{cccccccccc}
1 & 5 & 10 & 30 & 50 & 70 & 90 & 95 & 99 \\
\hline
(p, n_1, n_2) = (3, 5, 8) & \hat{\mu} = -2.905 & (-2.906) & \hat{\sigma} = .4355 & (.4325) \\
(p, n_1, n_2) = (8, 15, 20) & \hat{\mu} = -7.834 & (-7.828) & \hat{\sigma} = .4210 & (.4162) \\
(p, n_1, n_2) = (16, 22, 27) & \hat{\mu} = -17.77 & (-17.73) & \hat{\sigma} = .6425 & (.6325) \\
(p, n_1, n_2) = (16, 50, 60) & \hat{\mu} = -15.10 & (-15.09) & \hat{\sigma} = .2982 & (.2962) \\
(p, n_1, n_2) = (32, 120, 125) & \hat{\mu} = -30.17 & (-30.16) & \hat{\sigma} = .2846 & (.2827) \\
-30.83 & -30.63 & -30.52 & -30.30 & -30.16 & -30.01 & -29.80 & -29.70 & -29.51 \\
-30.84 & -30.64 & -30.54 & -30.32 & -30.17 & -30.02 & -29.81 & -29.71 & -29.52 \\
\end{array}
\]

Table 7. Empirical percentiles (top) and saddlepoint percentiles (bottom; not shown when identical to table accuracy) for the noncentral distribution of \( \log \LambdaECM \). The empirical percentiles were determined from \( 10^6 \) independent simulations of \( \log \LambdaECM \). The approximate means and standard deviations \( \hat{\mu} \) and \( \hat{\sigma} \) were determined from the approximated CGF while those in parentheses were determined with simulation.

The results in Table 6 show less accuracy in the center of the distribution than in the previous examples, but still have extremely high accuracy in the tails, and this should be acceptable for most practical purposes.
Table 6 also shows two expansions of order $O(n^{-3/2})$ for comparison in the cases $p = 3, 8$. Approximation "Norm. $O(n^{-3/2})$" is an expansion for fixed $\Delta$ that was suggested by Sugiura (1974, eqn. (3.9)) and has a normal approximation as its leading term. The expansion denoted as "$\chi^2$ to $O(n^{-3/2})$" was also suggested by Sugiura (1974 eqn. (3.5)) as a local expansion about the null hypothesis in which $\Delta - I = O(n^{-1/2})$ and which has a leading term that is noncentral $\chi^2$. None of these expansions deliver the accuracy of the saddlepoint approximation.

4 Computational Details of the Approximations

We now explain how to calculate our approximations to the non-null distribution of the three test statistics discussed in Sections 2 and 3.

4.1 Moment Generating Functions

In what follows,

$$\Gamma_p(a) = \pi^{p(p-1)/4} \prod_{i=1}^{p} \Gamma \left\{ a - \frac{1}{2} (i - 1) \right\}$$

is the multivariate gamma function, and $\Gamma_1(a) \equiv \Gamma(a) = \int_0^{\infty} x^{a-1} e^{-x} dx$ is the usual (scalar) gamma function; see e.g. Muirhead (1982, §2.1.2) for further details.

(i) The non-null MGF of $\log A_{GLM}$ is specified in Theorem 10.5.1 of Muirhead (1982) as

$$M_{GLM}(s) = E \left( e^{s \log A_{GLM}} \right)$$

$$= \frac{\Gamma_p \left( \frac{n}{2} + s \right) \Gamma_p \left\{ \frac{1}{2} (n + m) \right\} \Gamma_p \left( \frac{1}{2} (n + m) + s \right)}{\Gamma_p \left( \frac{n}{2} \right) \Gamma_p \left( \frac{1}{2} (n + m) + s \right)} \left\{ \frac{1}{2} \right\}^{1 \choose 1} \left\{ \frac{1}{2} (n + m) + s; -\frac{1}{2} \Omega \right\}. \quad (10)$$

(ii) The log likelihood ratio statistic for testing block independence has non-null
\( M_{B1}(s) = E\left(e^{s \log A_{B1}}\right) = \frac{\Gamma_{p_1} \left(\frac{s}{2}\right) \Gamma_{p_1} \left\{ \frac{1}{2} (n - p_2) + s \right\}}{\Gamma_{p_1} \left(\frac{s}{2} + \frac{n}{2}\right) \Gamma_{p_1} \left\{ \frac{1}{2} (n - p_2) \right\}} \times \left| I_{p_1} - P^2 \right|^{n/2} F_1\left(\frac{n}{2}, \frac{n}{2}; \frac{n}{2} + s; P^2\right). \) \( (11) \)

(iii) The log Bartlett \( M \)-statistic, \( \log A_{ECM} \), for testing equality of covariance matrices has non-null MGF

\( M_{ECM}(s) = E\left(e^{s \log A_{ECM}}\right) = \frac{\Gamma_p \left(\frac{s}{2}\right) \Gamma_p \left\{ \frac{n}{2} \left(1 + \frac{2s}{n}\right) \right\} \Gamma_p \left\{ \frac{n}{2} \left(1 + \frac{2s}{n}\right) \right\}}{\Gamma_p \left(\frac{n}{2}\right) \Gamma_p \left\{ \frac{n}{2} \left(1 + \frac{2s}{n}\right) \right\}} \times |\Delta|^{n/2} F_1\left\{ s, \frac{n}{2}, 1 + \frac{2s}{n}; \frac{n}{2}, 1 + \frac{2s}{n}; I_p - \Delta \right\}. \) \( (12) \)

The quantities \( _1F_1(a; b; X) \) and \( _2F_1(a, b; c; X) \) in the above formulae are hypergeometric functions of matrix argument.

### 4.2 Laplace Approximations to \( _1F_1 \) and \( _2F_1 \)

The following Laplace approximations to \( _1F_1 \) and \( _2F_1 \) were given in Butler and Wood (2002a), to which we refer the reader for further details.

The calibrated approximations \( \hat{F}_1(a; b; X) \) and \( \hat{F}_1(a, b; c; X) \) both have explicit form. The approximation \( \hat{F}_1(a; b; X) \) is given by

\[ \hat{F}_1(a; b; X) = x^{a-p}(p+1)/R_{1,1}^{1/2} \prod_{i=1}^{p} \left\{ \frac{\hat{y}_i}{a} \left( \frac{1 - \hat{y}_i}{b - a} \right)^{b-a} e^{x_i \hat{y}_i} \right\} \] \( (13) \)

where \( X = \text{diag}(x_1, \ldots, x_p) \),

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

and, for \( i = 1, \ldots, p \),

\[ \hat{y}_i = \frac{2a}{b - x_i + \sqrt{(x_i - b)^2 + 4ax_i}}. \]
The approximation $2 \hat{F}_1(a, b; c; X)$ is given by

$$
2 \hat{F}_1(a, b; c; X) = c^{p-\rho(p+1)/4} R_{2,1}^{-1/2} \prod_{i=1}^{p} \left\{ \left( \frac{\hat{y}_i}{a} \right)^{a} \frac{(1-\hat{y}_i)}{(c-a)} (1-x_i \hat{y}_i)^{-b} \right\}
$$

(14)

where $X = \text{diag}(x_1, \ldots, x_p)$,

$$
R_{2,1} = \prod_{i=1}^{p} \prod_{j=1}^{p} \left\{ \frac{\hat{y}_i \hat{y}_j}{a} + \frac{(1-\hat{y}_i)(1-\hat{y}_j)}{c-a} - \frac{b}{a(c-a)} S_i S_j \right\},
$$

$$
S_i = x_i \hat{y}_i (1-\hat{y}_i)/(1-x_i \hat{y}_i);
$$

and, putting $\tau_i = x_i (b-a) - c$,

$$
\hat{y}_i = \frac{2a}{\sqrt{\tau_i^2 - 4ax_i (c-b) - \tau_i}}.
$$

These approximations are calibrated so that they achieve equality at $X = 0_p$, i.e.

$$
1 \hat{F}_1(a; b; 0_p) = 1 F_1(a; b; 0_p) = 1 \quad \text{and} \quad 2 \hat{F}_1(a, b; c; 0_p) = 2 F_1(a, b; c; 0_p) = 1
$$

and

$$
2 \hat{F}_1(a, b; c; 0_p) = 2 F_1(a, b; c; 0_p) = 1;
$$

for this reason we often refer to $1 \hat{F}_1$ and $2 \hat{F}_1$ as calibrated Laplace approximations.

### 4.3 The Lugannanni and Rice Approximation

Given a random variable $X$ with MGF $M(s)$ and cumulant generating function $K(s) = \log M(s)$, the Lugannanni and Rice (1980) approximation to the tail probability $\Pr(X > y)$ is given by

$$
\Pr(X > y) \sim 1 - \Phi(r) + \phi(r) \left( \frac{1}{u} - \frac{1}{r} \right)
$$

where $\Phi$ and $\phi$ are the cumulative distribution function and density, respectively, of a standard normal random variable;

$$
r = \text{sgn}(\hat{s}) \sqrt{2 \{ \hat{s} \hat{y} - K(\hat{s}) \}} \quad \text{and} \quad u = \hat{s} \sqrt{K''(\hat{s})},
$$

25
where \( \text{sgn}(\hat{s}) = \pm 1 \) or 0 depending on whether \( \hat{s} \) is positive, negative or zero; and \( \hat{s} \) is the solution to the saddlepoint equation

\[
K''(\hat{s}) = y.
\]

In the above, \( K' \) and \( K'' \) are the first and second derivative, respectively, of the cumulant generating function \( K \).

The Lugananni and Rice approximation is known to be extremely accurate in a wide variety of settings, and this is certainly the case for the applications presented in Sections 2 and 3. However, the principal difficulty in applying Lugananni and Rice in the applications considered here is the computation of the matrix argument hypergeometric functions \( _1F_1 \) and \( _2F_1 \).

4.4 Summary of the Approximations

Our approximation to the non-null distribution of \( \Lambda_{\text{CLM}} \) consists of the following two-stage approach.

**Stage I:** Use the calibrated Laplace approximation \( \tilde{F}_1 \) to approximate \( _1F_1 \).

**Stage II:** Use the Lugananni and Rice tail probability approximation, but replace \( _1F_1 \) by \( \tilde{F}_1 \) in the MGF of \( \log \Lambda_{\text{CLM}} \).

All that changes for the two statistics, \( \Lambda_{\text{BI}} \) and \( \Lambda_{\text{ECM}} \), discussed in Section 3 is that we approximate \( _2F_1 \) with the calibrated Laplace approximation \( \tilde{F}_1 \).

In each of the three cases, nearly all of the required calculations are explicit. The hardest part of the computation is solving the (approximate) saddlepoint equation

\[
\hat{K}''(\hat{s}) = y
\]

where \( \hat{K} \) is the relevant cumulant generating function with \( _1F_1 \) (or \( _2F_1 \)) replaced by \( \tilde{F}_1 \) (or \( \tilde{F}_1 \)). It is straightforward to solve this equation numerically using analytical differentiation to calculate \( \hat{K}' \). The analytical expression for
\( \hat{K}' \) is however quite complicated and is therefore provided in Appendix B. It is also convenient to use one numerical derivative to calculate \( \hat{K}'' \) from \( \hat{K}' \) thus simplifying the programming considerably but also maintaining sufficient computational accuracy.

5 A Central Limit Theorem in High Dimensions

The behavior of the noncentral distributions when the dimension \( p \to \infty \) as the sample size \( n \) goes to infinity may be examined in terms of the saddlepoints involved in Table 1. Saddlepoints for the standardized value of \( \log \Lambda_{GLM} \), given by

\[
Z = \frac{1}{\sigma} (\log \Lambda_{GLM} - \mu)
\]

are even more revealing. It is easily shown that if \( \hat{s} \) is the saddlepoint for \( \log \Lambda_{GLM} \) at value \( y \), then \( \sigma \hat{s} \) is the saddlepoint for \( Z \) at \( z = (y - \mu) / \sigma \). However,

<table>
<thead>
<tr>
<th>p</th>
<th>1</th>
<th>5</th>
<th>10</th>
<th>30</th>
<th>50</th>
<th>70</th>
<th>90</th>
<th>95</th>
<th>99</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>-1.830</td>
<td>-1.366</td>
<td>-1.092</td>
<td>-0.4533</td>
<td>.0498</td>
<td>.6109</td>
<td>1.546</td>
<td>2.056</td>
<td>3.132</td>
</tr>
<tr>
<td>7*</td>
<td>-1.962</td>
<td>-1.445</td>
<td>-1.151</td>
<td>-0.4810</td>
<td>.0285</td>
<td>.5786</td>
<td>1.450</td>
<td>1.909</td>
<td>2.846</td>
</tr>
<tr>
<td>7**</td>
<td>-2.097</td>
<td>-1.539</td>
<td>-1.227</td>
<td>-0.5334</td>
<td>-.0209</td>
<td>.5197</td>
<td>1.345</td>
<td>1.767</td>
<td>2.599</td>
</tr>
<tr>
<td>12</td>
<td>-2.062</td>
<td>-1.500</td>
<td>-1.187</td>
<td>-0.4912</td>
<td>.0209</td>
<td>.5612</td>
<td>1.392</td>
<td>1.814</td>
<td>2.651</td>
</tr>
<tr>
<td>25*</td>
<td>-2.062</td>
<td>-1.501</td>
<td>-1.184</td>
<td>-0.4904</td>
<td>.0226</td>
<td>.5619</td>
<td>1.392</td>
<td>1.813</td>
<td>2.638</td>
</tr>
<tr>
<td>25**</td>
<td>-2.133</td>
<td>-1.544</td>
<td>-1.217</td>
<td>-.5084</td>
<td>.0^2684</td>
<td>.5408</td>
<td>1.349</td>
<td>1.749</td>
<td>2.532</td>
</tr>
<tr>
<td>50</td>
<td>-2.108</td>
<td>-1.523</td>
<td>-1.199</td>
<td>-.4965</td>
<td>.0179</td>
<td>.5545</td>
<td>1.369</td>
<td>1.781</td>
<td>2.582</td>
</tr>
<tr>
<td>z</td>
<td>-2.326</td>
<td>-1.645</td>
<td>-1.282</td>
<td>-.5244</td>
<td>0</td>
<td>.5244</td>
<td>1.282</td>
<td>1.645</td>
<td>2.326</td>
</tr>
</tbody>
</table>

Table 8. Saddlepoints associated with the standardized value of \( \log \Lambda_{GLM} \) and used in finding the saddlepoint approximations of Table 1. Row \( z \) contains the various standard normal percentiles. Superscripts * and ** denote respectively the settings with the smaller and larger entries in \( \Omega \).
since $\mu$ and $\sigma$ are unknown, we replace them with estimates $\hat{\mu}$ and $\hat{\sigma}$ obtained from simulation, and assume these estimates are exact. With this assumption, then the saddlepoints for $Z$ as determined for Table 1 are given in Table 8. Row $z$ contains the standard normal percentiles which are well-known to also be saddlepoints for these percentiles under a standard normal distribution. The saddlepoints appear to be converging to the standard normal saddlepoints as $p \to \infty$. If so then this would be indicative of a central limit theorem. To see this, suppose that $K_p(s)$ and $F_p(z)$ are the CGF and CDF of $Z$ in $p$ dimensions and consider the mapping from the saddlepoint to the 100th percentile. If $\delta_{p,\alpha}$ is the saddlepoint for $F_p^{-1}(\alpha)$, so that $K_p'(\delta_{p,\alpha}) = F_p^{-1}(\alpha)$, then

$$
\delta_{p,\alpha} = \left\{ \frac{1}{K_p'} \circ F_p^{-1} \right\}(\alpha).
$$

If the function in curly braces converges pointwise to $\Phi^{-1}$ as $p \to \infty$, so the right side approaches $\Phi^{-1}(\alpha) = z_\alpha$, then a central limit theorem must hold.

There is also clear evidence for a central limit tendency in high dimensions with the other two test statistics as may be seen in Tables 9 and 10.

<table>
<thead>
<tr>
<th>$(p_1,p_2,n)$</th>
<th>1</th>
<th>5</th>
<th>10</th>
<th>30</th>
<th>50</th>
<th>70</th>
<th>90</th>
<th>95</th>
<th>99</th>
</tr>
</thead>
<tbody>
<tr>
<td>(2,3,10)</td>
<td>-1.465</td>
<td>-1.154</td>
<td>-0.9457</td>
<td>-0.3958</td>
<td>0.0948</td>
<td>0.6990</td>
<td>1.863</td>
<td>2.609</td>
<td>4.616</td>
</tr>
<tr>
<td>(5,7,20)</td>
<td>-1.846</td>
<td>-1.377</td>
<td>-1.100</td>
<td>-0.4547</td>
<td>0.04881</td>
<td>0.6063</td>
<td>1.519</td>
<td>2.007</td>
<td>3.035</td>
</tr>
<tr>
<td>(10,13,40)</td>
<td>-2.098</td>
<td>-1.520</td>
<td>-1.197</td>
<td>-0.4954</td>
<td>0.01963</td>
<td>0.5571</td>
<td>1.376</td>
<td>1.784</td>
<td>2.584</td>
</tr>
<tr>
<td>(25,28,70)</td>
<td>-2.185</td>
<td>-1.566</td>
<td>-1.230</td>
<td>-0.5063</td>
<td>0.01170</td>
<td>0.5436</td>
<td>1.339</td>
<td>1.727</td>
<td>2.471</td>
</tr>
</tbody>
</table>

| $z$         | -2.326 | -1.645 | -1.282 | -0.5244 | 0 | 0.5244 | 1.282 | 1.645 | 2.326 |

Table 9. Saddlepoints associated with the standardized value of $\log \Lambda_{n1}$ and used in finding the saddlepoint approximations of Table 4.
\begin{table}
\centering
\begin{tabular}{cccccccccc}
\hline
$p$ & 1   & 0.5 & 10  & 30  & 50  & 70  & 90  & 95  & 99  \\
3   & -1.194 & -0.9704 & -0.8072 & -0.3327 & 0.1519 & 0.8292 & 2.365 & 3.485 & 6.874 \\
8   & -1.772 & -1.325 & -0.5777 & -0.4331 & 0.06465 & 0.6217 & 1.549 & 2.050 & 3.114 \\
16  & -1.908 & -1.386 & -1.084 & -0.4122 & 0.09409 & 0.6379 & 1.483 & 1.921 & 2.799 \\
16  & -2.087 & -1.499 & -1.173 & -0.4670 & 0.04565 & 0.5787 & 1.386 & 1.791 & 2.589 \\
32  & -2.175 & -1.535 & -1.194 & -0.4609 & 0.05665 & 0.5819 & 1.360 & 1.743 & 2.464 \\
\hline
$z$  & -2.326 & -1.645 & -1.282 & -0.5244 & 0.5244 & 1.282 & 1.645 & 2.326 & \\
\hline
\end{tabular}
\caption{Saddlepoints associated with the standardized value of $\log \Lambda_{ECM}$ and used in finding the saddlepoint approximations of Table 6.}
\end{table}

We formalize the observations above with a central limit theorem.

**Theorem 1** Suppose that $m = m_n$ and $p = p_n$ satisfy

\begin{equation}
1 \leq m_n \leq \theta n \quad \text{and} \quad 1 \leq p_n \leq \theta n,
\end{equation}

where $\theta \in (0, 1)$ does not depend on $n$. Let $\log \Lambda_{GLM}^{(n)}$, $n = 1, 2, \ldots$ be a sequence of random variables with moment generating functions of the form $M_{GLM}^{(n)}(s)$ given in (10), with noncentrality matrix $\Omega = \Omega^{(n)}$ assumed to satisfy

\begin{equation}
p_n^{-1} \text{tr}(\Omega^{(n)}) = o(m_n).
\end{equation}

Write

\[ \sigma_n^2 = \frac{n^2}{m_n p_n} \text{Var}(\log \Lambda_{GLM}) \quad \text{and} \quad \tau_n = \frac{1}{\sqrt{m_n p_n}} + \left( \frac{\text{tr}(\Omega^{(n)})}{m_n p_n} \right)^{3/2}. \]

Then as $n \to \infty$, $\max\{\sigma_n^2, 1/\sigma_n^2\} = O(1)$ and

\[ \Pr \left[ \frac{n}{\sqrt{m_n p_n}} \left( \frac{\log \Lambda_{GLM} - E(\log \Lambda_{GLM})}{\sigma_n} \right) \leq x \right] = \Phi(x) + O(\tau_n), \]

where $\Phi$ is the standard normal CDF.

Weak convergence to normality does not in itself imply convergence of the saddlepoint approximation to the true distribution, though it turns out that
the latter type of convergence does hold here. Below, we give a result which

describes the convergence of the saddlepoint approximation under the same

conditions as Theorem 1.

**Theorem 2** Assume the asymptotic setup in Theorem 1, i.e. that (15) and

(16) both hold. Suppose, in addition, that \( m_n p_n \to \infty \), and let \( \tau_n \) be defined as

in Theorem 1. Then for any \( a \) and \( b \) such that \( 0 < a < b < 1 \),

\[
\sup_{\alpha \in (a,b)} |\delta_{p,\alpha} - z_\alpha| = O(\tau_n);
\]

and, uniformly for any bounded interval \( I \),

\[
\sup_{t \in I} |K_p(t) - \frac{1}{2} t^2|, \quad \sup_{t \in I} |K'_p(t) - t| \quad \text{and} \quad \sup_{t \in I} |K''_p(t) - 1|
\]

are all of size \( O(\tau_n) \) as \( n \to \infty \). Consequently, the saddlepoint approximation
to the standardized CDF converges uniformly to the standard normal CDF \( \Phi \) on

compact intervals.

Proofs of these results are given in Butler and Wood (2002b).

We follow with some comments.

1. Note that when \( p_n, m_n \) stays bounded as \( n \to \infty \) in Theorem 1, the central

limit theorem does not hold. In such cases we have the usual chi-squared limit.

2. Results broadly similar to Theorems 1 and 2 can be obtained for the other

two statistics \( \Lambda_{BI} \) and \( \Lambda_{ECM} \).

3. Butler and Wood (2002a, Theorem 2) give a result complementary to Theorems 1 and 2 which states that, for fixed \( n \) (and fixed \( m_n, p_n \), and \( \Omega^{(n)} \)), the

relative error of the saddlepoint approximation is bounded in both tails and

close to zero in the limit.

4. The condition (16) in Theorems 1 and 2 ensures that the contribution from

the "noncentral" component of (10) is asymptotically negligible compared with

the contribution from the "null" component (after taking account of the mean

of the noncentral component).
5. An interesting and challenging open problem is to extend Theorems 1 and 2 to settings in which (16) does not hold. Our expectation is that a limit distribution will exist under mild conditions but this seems difficult to prove.

A different problem involving “high dimension” asymptotics in multivariate analysis has been considered by Johnstone (2001); he studies the limit distribution of the largest eigenvalue of Wishart matrix as the dimension of the matrix goes to infinity.

Appendix A

Consider the factorizations of $X$ and $C$ given by

$$X = Q_1 D \quad X (N \times q), \quad Q_1 (N \times q), \quad D (q \times q)$$

and

$$C = FP_1 \quad C (m \times q), \quad F (m \times m), \quad P_1 (m \times q)$$

where $Q_1^T Q_1 = I_q, P_1 P_1^T = I_m$ and $D$ and $F$ have full rank. Let $Q_2$ denote a $N \times (N - q)$ matrix such that the block matrix $(Q_1 : Q_2) \in \mathcal{O}(N)$, the space of orthogonal matrices; and let $P_2$ be a $(m - q) \times q$ matrix such that the block matrix $(P_1^T : P_2^T)^T \in \mathcal{O}(q)$. Then canonical variables $Y_1^* (m \times p)$ and $Y_3^* ((q - m) \times p)$ are given by

$$Y^* = \begin{pmatrix} Y_1^* \\ Y_2^* \\ Y_3^* \end{pmatrix} = \begin{pmatrix} P_1 Q_1^T Y \\ P_2 Q_2^T Y \\ Q_3^T Y \end{pmatrix}$$

Under the model (3), $E(Y_3^*) = 0_{N-q,p}$ and therefore $SS_{err} = (Y_3^*)^T Y_3^*$ has a (central) Wishart$_p (N - q, \Sigma)$ distribution; and

$$M_1 = E(Y_1^*) = P_1 Q_1^T E(Y) = P_1 Q_1^T X B = F^{-1} C Q_1^T Q_1 D B = F^{-1} C D B.$$
Therefore $SS_{\text{hypo}} = (Y_1^*)^T Y_1^*$ has a non-central Wishart distribution $(m, \Sigma, \Omega)$ with non-centrality matrix

$$\Omega = \Sigma^{-1} M_1^T M_1 = \Sigma^{-1} B^T D^T C^T (F^T)^{-1} F^{-1} C D B.$$

Also, $Y_1^*$ and $Y_2^*$ are independent; this follows from Muirhead (1982), Theorem 10.1.2 and the discussion on pages 436–37.

Appendix B

The expression needed to solve the saddlepoint equation in the case of Wilks' statistic is

$$\hat{K}'(s) = \sum_{i=1}^{p} \left\{ \psi \left[ \frac{n}{2} + s - \frac{1}{2} (i - 1) \right] - \psi \left[ \frac{n + m}{2} + s - \frac{1}{2} (i - 1) \right] \right\} + \nabla(s)$$

where $\psi(z) = d\ln \Gamma(z) / dz$ and

$$\nabla(s) = \frac{\partial}{\partial s} \ln \Gamma\left(\frac{a}{b}; -\frac{1}{2} \Omega \right).$$

Here and also below we use $a = s$ and $b = (n + m) / 2 + s$ to simplify the expressions. Derivative $\nabla$ is computed using implicit differentiation and, after substantial simplification, yields

$$\nabla(s) = \left\{ b - \frac{1}{4} (p + 1) \right\} \frac{p}{b} + p \ln b$$

$$- \frac{1}{2} \sum_{i=1}^{p} \sum_{j=1}^{p} \left\{ \frac{\hat{y}_i \hat{y}_j}{a} + \frac{(1 - \hat{y}_i)(1 - \hat{y}_j)}{b - a} \right\}^{-1} \left\{ \left( \frac{\hat{y}_i}{a} - \frac{1 - \hat{y}_i}{b - a} \right) \frac{\partial \hat{y}_j}{\partial s} \right. \right.$$  

$$+ \left( \frac{\hat{y}_j}{a} - \frac{1 - \hat{y}_j}{b - a} \right) \frac{\partial \hat{y}_i}{\partial s} - \hat{y}_i \hat{y}_j \right\}$$

$$+ \sum_{i=1}^{p} \left\{ \ln \left( \frac{\hat{y}_i}{a} \right) - 1 + \left( \frac{a}{\hat{y}_i} - \frac{b - a}{1 - \hat{y}_i} + x_i \right) \frac{\partial \hat{y}_i}{\partial s} \right\},$$

with

$$\frac{\partial \hat{y}_i}{\partial s} = 2 \frac{b - x_i + \sqrt{q_i} - a - a(x_i + b) / \sqrt{q_i}}{(b - x_i + \sqrt{q_i})^2}$$

and

$$q_i = (x_i - b)^2 + 4ax_i.$$
The computation of $\hat{K}''(s)$ is best performed using a numerical derivative of $\hat{K}'(s)$.

References


distribution of the sum of independent random variables. Adv. Appl. Probab
12, 475-490.


Wiley, New York.

calculations for general linear multivariate models including repeated mea-

power in testing the multivariate general linear hypothesis. Computational
Statistics and Data Analysis, 2, 143-158.


Math. 25 153-163.

criteria for covariance matrix under local alternatives. Ann. Statist. 1, 718-
728.

Math. 26, 117-125.