An error bound for high–dimensional Edgeworth expansion of Wilks' Lambda distribution

Hirofumi Wakaki
Department of Mathematics, Hiroshima University

Abstract

An error bound for the Edgeworth expansion of Wilks' Lambda distribution is derived when some of the parameters are large. Some tables of the error bounds are given, which shows that the derived bounds are very sharp.

1 Introduction

Let B and W be independent random matrices of size $p \times p$ distributed as $W_p(q, \Sigma)$ and $W_p(n, \Sigma)$, respectively. We assume that $n > p$. The distribution of $\Lambda = \det\{B(W+B)^{-1}\}$ is called Wilks' lambda distribution and denoted as $\Lambda_p(q, n)$. We assume that $p \geq q$ without any loss of generality, because it holds that

$$\Lambda_p(q, n) = \Lambda_q(p, n-p+q). \tag{1.1}$$

When n is large relative to p and q, an asymptotic expansion based on the formula of Box (1949) can be used. Let

$$M = n - \frac{1}{2}(p+q+1), \gamma = \frac{1}{48}pq(p^2+q^2-5) \tag{1.2}$$

and $f = pq$. Then

$$\Pr\{-M \log \Lambda \leq x\} = G_f(x) + \frac{\gamma}{M^2}\{G_{f+4}(x) - G_f(x)\} + O(M^{-4}). \tag{1.3}$$

Distributional results for Λ and other test statistics can be found in Muirhead (1982, chapter 10), Anderson (1984, chapter 8) and Siotani, Hayakawa and Fujikoshi (1985, chapter 6). A computable error bound for the above expansion formula can be found in Fujikoshi and Ulyanov (2006).

Tonda and Fujikoshi (2004) derived an asymptotic expansion of the distribution of $\log \Lambda$ when q was fixed and

$$p \to \infty,\ n \to \infty,\ \frac{p}{n} \to c \in (0,1), \tag{1.4}$$

by using the fact that $W_q(p, n-p+q)$ is the same distribution of $\prod_{j=1}^{q} X_j$ where $X_1, \cdots, X_q$ are independent and X_j is distributed as beta distribution with degree of freedom $\frac{p}{2}$ and $\frac{n-p+j}{2}$. They used an approximated cumulants derived by so called delta method, while wakaki (2006) used the exact cumulants of which the formulas are given as follows.

It is known that the hth moment of Λ is given by

$$\mathrm{E}[\Lambda^h] = \prod_{j=1}^{q} \frac{\Gamma[\frac{n-p+j}{2}+h]\Gamma[\frac{n+j}{2}]}{\Gamma[\frac{n-p+j}{2}]\Gamma[\frac{n+j}{2}+h]}. \tag{1.5}$$

By using this formula, the cumulant generating function of $T = -\log\Lambda$ can be expanded as

$$\begin{aligned} \log \mathrm{E}[\exp(itT)] &= \log \mathrm{E}[\Lambda^{-it}] \\ &= \sum_{j=1}^{q} \Big\{ \log\Gamma\Big[\frac{n-p+j}{2} - it\Big] - \log\Gamma\Big[\frac{n-p+j}{2}\Big] \\ &\qquad - \Big(\log\Gamma\Big[\frac{n+j}{2} - it\Big] - \log\Gamma\Big[\frac{n+j}{2}\Big]\Big)\Big\} \\ &= \sum_{s=1}^{\infty} \frac{\kappa^{(s)}}{s!}(it)^s, \end{aligned} \tag{1.6}$$

where

$$\kappa^{(s)} = (-1)^s \sum_{j=1}^{q} \Big\{ \psi^{(s-1)}\Big(\frac{n-p+j}{2}\Big) - \psi^{(s-1)}\Big(\frac{n+j}{2}\Big)\Big\}. \tag{1.7}$$

Here $\psi^{(s)}$ is the polygamma function defined by

$$\psi^{(s)}(a) = \Big(\frac{d}{da}\Big)^{s+1} \log\Gamma[a] = \begin{cases} -C + \sum_{k=0}^{\infty}\Big(\frac{1}{1+k} - \frac{1}{k+a}\Big) & (s=0) \\ \sum_{k=0}^{\infty} \frac{(-1)^{s+1}s!}{(k+a)^{s+1}} & (s=1,2,\cdots)\ . \end{cases} \tag{1.8}$$

Wakaki (2006) examined the order of the standardized cumulants $\kappa^{(s)}/(\kappa^{(2)})^{-s/2}$ $(s = 3,4,\cdots)$ in several cases that some of n,p,q get large and clarified whether the Edgeworth expansion is available or not. Numerical experiments shows that using the exact cumulants gives very accurate approximation formula.

Ulyanov, Wakaki and Fujikoshi (2006) derived a Berry–Esseen bound for a normal approximation under the set up (1.4). First they derived a Berry–Esseen bound for normal approximation of F–distribution when the degrees of freedom were large. Then they applied the result to the approximation for $\log\Lambda$. The obtained error bound is not good. One of the reason is that they only used the fact that the F–statistc is the ratio of two averages of i.i.d. random variables. If we use the exact formula of the characteristic function of Wilks' lambda dsitribution we can obtain very sharp bounds for normal approximation as well as the Edgeworth expansion formula of the arbitrary order as in the following sections.

2 Edgeworth expansion

Bounds for the standardized cumulants are given in the following lemma.

Lemma 2.1 *The following inequalities for the cumulants hold.*

$$2\log\frac{n(n-p+q)}{(n-p)(n+q)} < \kappa^{(2)} < 2\log\frac{(n-\frac{1}{2})(n-p+q-\frac{1}{2})}{(n-p-\frac{1}{2})(n+q-\frac{1}{2})}, \tag{2.1}$$

$$\begin{aligned} 0 < \kappa^{(s)} < \frac{2^{s-1}(s-3)!}{(n-p-\frac{1}{2})^{s-2}}\Big\{1 - \Big(\frac{n-p-\frac{1}{2}}{n-p+q-\frac{1}{2}}\Big)^{s-2} \\ - \Big(\frac{n-p-\frac{1}{2}}{n-\frac{1}{2}}\Big)^{s-2} + \Big(\frac{n-p-\frac{1}{2}}{n+q-\frac{1}{2}}\Big)^{s-2}\Big\} \quad (s \geq 3). \end{aligned} \tag{2.2}$$

Proof If $1 \leq A < B$, it holds that

$$\frac{1}{A^2} - \frac{1}{B^2} > \int_{-1}^{0}\Big\{\int_0^1 \frac{1}{(A+x+\frac{y}{2})^2} - \frac{1}{(B+x+\frac{y}{2})^2}dx\Big\}dy.$$

Hence

$$\begin{aligned} \kappa^{(2)} &= \sum_{j=1}^{q}\sum_{k=0}^{\infty}\Big\{\frac{1}{(\frac{n-p+j}{2}+k)^2} - \frac{1}{(\frac{n+j}{2}+k)^2}\Big\} \\ &> \int_0^q\Big\{\int_0^\infty \frac{1}{(\frac{n-p+y}{2}+x)^2} - \frac{1}{(\frac{n+y}{2}+x)^2}dx\Big\}dy. \end{aligned}$$

The last integral gives the lower bound for $\kappa^{(2)}$.

Since

$$\frac{1}{(\frac{n-p+y}{2}+x)^s} - \frac{1}{(\frac{n+y}{2}+x)^s}$$

is convex as a function of x and y when $\frac{n-p+y}{2}+x > 0$ and $s > 0$,

$$\frac{1}{(\frac{n-p+j}{2}+k)^s} - \frac{1}{(\frac{n+j}{2}+k)^s} < \int_{j-1/2}^{j+1/2}\Big\{\int_{k-1/2}^{k+1/2} \frac{1}{(\frac{n-p+y}{2}+x)^s} - \frac{1}{(\frac{n+y}{2}+x)^s}dx\Big\}dy.$$

Hence

$$\begin{aligned} 0 < \kappa^{(s)} &= \sum_{j=1}^{q}\sum_{k=0}^{\infty}\Big\{\frac{(s-1)!}{(\frac{n-p+j}{2}+k)^s} - \frac{(s-1)!}{(\frac{n+j}{2}+k)^s}\Big\} \\ &< \int_{1/2}^{q+1/2}\Big\{\int_{-1/2}^{\infty} \frac{(s-1)!}{(\frac{n-p+y}{2}+x)^s} - \frac{(s-1)!}{(\frac{n+y}{2}+x)^s}dx\Big\}dy. \end{aligned}$$

The last integration in the case of $s = 2$ and $s > 2$ gives the upper bound in (2.1) and (2.2), respectively. ∎

Let m and b_s $(s = 0, 1, 2, \cdots)$ be defined by

$$m = \frac{n-p-\frac{1}{2}}{2}(\kappa^{(2)})^{1/2},$$
$$b_s = \frac{2}{(s+3)(s+2)(s+1)}(\kappa^{(2)})^{-1} \cdot \left\{1 - \left(\frac{n-p-\frac{1}{2}}{n-p+q-\frac{1}{2}}\right)^{s+1} - \left(\frac{n-p-\frac{1}{2}}{n-\frac{1}{2}}\right)^{s+1} + \left(\frac{n-p-\frac{1}{2}}{n+q-\frac{1}{2}}\right)^{s+1}\right\}. \tag{2.3}$$

With using (2.1), we can check that b_s is bounded and that m becomes large if at least two of $n-p, p$ and q become large. The orders of m and b_s are summarized in table 1 in section 4.

Lemma 2.1 gives a bound for the standardized cumulant $\widetilde{\kappa}^{(s)} := \kappa^{(s)}(\kappa^{(2)})^{-s/2}$,

$$0 < \widetilde{\kappa}^{(s)} < s!\, m^{-(s-2)} b_{s-3} \ (s = 3, 4, \cdots). \tag{2.4}$$

Let $\widetilde{T}$ be the standardized statistic defined by

$$\widetilde{T} = \frac{T - \kappa^{(1)}}{(\kappa^{(2)})^{1/2}}. \tag{2.5}$$

The characteristic function of $\widetilde{T}$ can be expanded as

$$\begin{aligned}
\varphi(t) := \mathrm{E}[\exp(it\widetilde{T})] &= \exp\left\{-\frac{t^2}{2} + \sum_{s=3}^{\infty} \frac{\widetilde{\kappa}^{(s)}}{s!}(it)^s\right\} \\
&= \exp\left(-\frac{t^2}{2}\right)\left\{1 + \sum_{k=1}^{\infty} \frac{1}{k!}(it)^{3k}\left(\sum_{s=0}^{\infty} \frac{\widetilde{\kappa}^{(s+3)}}{(s+3)!}(it)^s\right)^k\right\} \\
&= \exp\left(-\frac{t^2}{2}\right)\left\{1 + \sum_{k=1}^{\infty} \frac{1}{k!}(it)^{3k}\sum_{j=0}^{\infty}\left(\sum_{s_1+\cdots+s_k=j} \frac{\widetilde{\kappa}^{(s_1+3)}\cdots\widetilde{\kappa}^{(s_k+3)}}{(s_1+3)!\cdots(s_k+3)!}\right)(it)^j\right\}.
\end{aligned} \tag{2.6}$$

From (2.3)

$$\gamma_{k,j} := \sum_{s_1+\cdots+s_k=j} \frac{\widetilde{\kappa}^{(s_1+3)}\cdots\widetilde{\kappa}^{(s_k+3)}}{(s_1+3)!\cdots(s_k+3)!} = O(m^{-(j+k)}) \tag{2.7}$$

as $m \to \infty$. Therefore, let

$$\varphi_s(t) = \exp\left(-\frac{t^2}{2}\right)\left\{1 + \sum_{k=1}^{s} \frac{1}{k!}(it)^{3k}\sum_{j=0}^{s-k}\gamma_{k,j}(it)^j\right\}, \tag{2.8}$$

then

$$\varphi(t) = \varphi_s(t) + O(m^{-(s+1)}) \quad (m \to \infty).$$

Inverting the characteristic function formally, we obtain the Edgeworth expansion of the distribution function of $\widetilde{T}$ up to the order $O(m^{-s})$ as

$$Q_s(x) = \Phi(x) - \phi(x)\left\{\sum_{k=1}^{s} \frac{1}{k!} \sum_{j=0}^{s-k} \gamma_{k,j} h_{3k+j-1}(x)\right\}, \tag{2.9}$$

where Φ and ϕ are the distribution function and the probability density function of the standard normal distribution, respectively, $\gamma_{k,j}$ is given by (2.7), and $h_r(x)$ is the r–th order Hermite polynomial defined by

$$\left(\frac{d}{dx}\right)^r \exp\left(-\frac{x^2}{2}\right) = (-1)^r h_r(x) \exp\left(-\frac{x^2}{2}\right).$$

3 Error bound

Using the inverse Fourier transformation we obtain a uniform bound for the error of the Edgeworth expansion as

$$\begin{aligned}\sup_x \left|\mathrm{P}(\widetilde{T} \le x) - Q_s(x)\right| &\le \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{1}{|t|} \left|\varphi(t) - \varphi_s(t)\right| dt \\ &= \frac{1}{2\pi}(I_1[v] + I_2[v] + I_3[v]),\end{aligned} \tag{3.1}$$

where

$$I_1[v] = \int_{-mv}^{mv} \frac{1}{|t|} \left|\varphi(t) - \varphi_s(t)\right| dt,$$
$$I_2[v] = \int_{|t|>mv} \frac{1}{|t|} \left|\varphi_s(t)\right| dt \quad \text{and} \quad I_3[v] = \int_{|t|>mv} \frac{1}{|t|} \left|\varphi(t)\right| dt$$

with some positive constant $v < 1$.

First we derive a bound for $I_1[v]$. Let

$$L[v] = \begin{cases} \dfrac{3v-2}{4v} - \dfrac{(1-v)^2}{2v^2} \log(1-v) & (0 < |v| < 1) \\ 0 & (v = 0). \end{cases} \tag{3.2}$$

Then it is easily checked that $L[v]$ can be expanded as

$$L[v] = \sum_{s=1}^{\infty} \frac{1}{s(s+1)(s+2)} v^s.$$

So $B[v] := \sum_{s=0}^{\infty} b_s v^s$ can be expressed as

$$B[v] = \frac{2}{v\kappa^{(2)}}\left\{L[v] - L\left[\frac{n-p-\frac{1}{2}}{n-p+q-\frac{1}{2}}v\right] - L\left[\frac{n-p-\frac{1}{2}}{n-\frac{1}{2}}v\right] + L\left[\frac{n-p-\frac{1}{2}}{n+q-\frac{1}{2}}v\right]\right\} \tag{3.3}$$

Let

$$R_{k,l}[v] = v^{-l}\Big\{(B[v])^k - \sum_{j=0}^{l-1}\Big(\sum_{s_1+\cdots+s_k=j} b_{s_1}\cdots b_{s_k}\Big)v^j\Big\}. \tag{3.4}$$

Note that the second term in the above braces is the Taylor expansion of $(B[v])^k$ up to the order v^{l-1}.

Then from (2.6), (2.7) and (2.8) with using (2.4), if $|t| \le mv$

$$\begin{aligned}
&\frac{1}{|t|}|\varphi(t)-\varphi_s(t)| \\
&\quad\le \exp\Big(-\frac{t^2}{2}\Big)\Big\{\sum_{k=1}^{s}\frac{1}{k!}|t|^{3k-1}\sum_{j=s-k+1}^{\infty}\Big(\sum_{s_1+\cdots+s_k=j} b_{s_1}\cdots b_{s_k}\Big)m^{-(j+k)}|t|^j \\
&\qquad\qquad + \sum_{k=s+1}^{\infty}\frac{1}{k!}|t|^{3k-1}\Big(\sum_{j=0}^{\infty} b_j m^{-(j+1)}|t|^j\Big)^k\Big\} \\
&\quad= \exp\Big(-\frac{t^2}{2}\Big)\Big\{\sum_{k=1}^{s}\frac{1}{k!}|t|^{3k-1}m^{-k}\Big(\frac{|t|}{m}\Big)^{s-k+1}R_{k,s-k+1}\Big[\frac{|t|}{m}\Big] \\
&\qquad\qquad + \sum_{k=s+1}^{\infty}\frac{1}{k!}|t|^{3k-1}m^{-k}\Big(B\Big[\frac{|t|}{m}\Big]\Big)^k\Big\} \\
&\quad\le m^{-(s+1)}\exp\Big(-\frac{t^2}{2}\Big)\Big\{\sum_{k=1}^{s}\frac{1}{k!}|t|^{s+2k}R_{k,s-k+1}[v] \\
&\qquad\qquad + \frac{1}{(s+1)!}|t|^{3s+2}(B[v])^{s+1}\exp\Big(t^2vB[v]\Big)\Big\}.
\end{aligned}$$

Integrating the last expression, we obtain a bound for $I_1[v]$:

$$\begin{aligned}
I_1[v] \le U_1[v] := \frac{2}{m^{s+1}}\Big\{&\sum_{k=1}^{s}\frac{1}{k!}R_{k,s-k+1}[v]\int_0^{mv} t^{s+2k}\exp\Big(-\frac{t^2}{2}\Big)dt \\
&+ \frac{1}{(s+1)!}(B[v])^{s+1}\int_0^{mv} t^{3s+2}\exp\Big(-\frac{t^2}{2}c_v\Big)dt\Big\},
\end{aligned} \tag{3.5}$$

where

$$c_v = 1 - 2vB[v]. \tag{3.6}$$

Note that $U_1[v] = O(m^{-(s+1)})$ if $c_v > 0$ since

$$\begin{aligned}
2\int_0^{mv} t^k\exp\Big(-\frac{t^2}{2}c_v\Big)dt &= \Big(\frac{c_v}{2}\Big)^{-(k+1)/2}\int_0^{mv}\Big(\frac{t^2}{2}c_v\Big)^{(k-1)/2}\exp\Big(-\frac{t^2}{2}c_v\Big)(tc_v)dt \\
&= \Big(\frac{c_v}{2}\Big)^{-(k+1)/2}\int_0^{m^2v^2c_v/2} s^{(k-1)/2}e^{-s}ds < \Big(\frac{c_v}{2}\Big)^{-(k+1)/2}\Gamma\Big[\frac{k+1}{2}\Big].
\end{aligned} \tag{3.7}$$

More simple, but a little loose upper bound for $I_1[v]$ is given by

$$\widetilde{U}_1[v] := \frac{2}{m^{s+1}}\Big\{\sum_{k=1}^{s}\frac{1}{k!}R_{k,s-k+1}[v]2^{-(s+1)/2-k}\Gamma\Big[\frac{s+1}{2}+k\Big] \\ +\frac{1}{(s+1)!}(B[v])^{s+1}\Big(\frac{c_v}{2}\Big)^{-(3s+3)/2}\Gamma\Big[\frac{3s+3}{2}\Big]\Big\}. \tag{3.8}$$

Next we consider $I_2[v]$ in (3.1), which is represented as

$$I_2[v] = 2\Big\{\int_{mv}^{\infty}\exp\Big(-\frac{t^2}{2}\Big)t^{-1}dt+\sum_{k=1}^{s}\frac{1}{k!}\sum_{j=0}^{s-k}\gamma_{k,j}\int_{mv}^{\infty}\exp\Big(-\frac{t^2}{2}\Big)|t|^{3k+j-1}dt\Big\}. \tag{3.9}$$

The integrals included in $I_2[v]$ are not difficult to compute. For positive constant a,

$$2\int_{mv}^{\infty}t^k\exp(-at^2)dt = a^{-(k+1)/2}\int_{m^2v^2a}^{\infty}s^{(k-1)/2}e^{-s}ds$$

is the incomplete gamma function. There are several algorithms to compute it. Using the above formula,

$$2\int_{mv}^{\infty}t^k\exp\Big(-\frac{t^2}{2}\Big)dt = 2\int_{mv}^{\infty}t^k\exp\Big(-\frac{t^2}{2}(1-c+c)\Big)dt \\ < \exp\Big(-\frac{m^2v^2}{2}(1-c)\Big)\Big(\frac{c}{2}\Big)^{-(k+1)/2}\Gamma\Big[\frac{k+1}{2}\Big],$$

where $0 < c < 1$. So we obtain a simple upper bound for $I_2[v]$ as

$$I_2[v] < \exp\Big(-\frac{m^2v^2}{2}(1-c)\Big)\Big\{1+\sum_{k=1}^{s}\frac{1}{k!}\sum_{j=0}^{s-k}\gamma_{k,j}\Big(\frac{c}{2}\Big)^{-(3k+j)/2}\Gamma\Big[\frac{3k+j}{2}\Big]\Big\}. \tag{3.10}$$

We note that $I_2[v] = O(\exp[-\frac{m^2v^2}{2}(1-c)])$ $(m \to \infty)$ for fixed v $(0 < v)$ and c $(0 < c < 1)$.

Finally we derive a bound for $I_3[v]$ in (3.1). From (1.5) the characteristic function of T is given by

$$\varphi(t) = \prod_{j=1}^{q}\frac{\Gamma[\frac{n-p+j}{2}-i\widetilde{t}\,]\Gamma[\frac{n+j}{2}]}{\Gamma[\frac{n-p+j}{2}]\Gamma[\frac{n+j}{2}-i\widetilde{t}\,]}, \tag{3.11}$$

where $\widetilde{t} = (\kappa^{(2)})^{-1/2}t$. It is known that

$$\left|\frac{\Gamma[x+yi]}{\Gamma[x]}\right|^2 = \prod_{k=0}^{\infty}\Big\{1+\frac{y^2}{(x+k)^2}\Big\}^{-1} \tag{3.12}$$

for any real number $x, y; (x > 0)$. Since if $A < B$

$$\log\Big\{1+\frac{t^2}{(A+x)^2}\Big\} - \log\Big\{1+\frac{t^2}{(B+x)^2}\Big\}$$

is a decreasing function of $x > 0$,

$$
\begin{aligned}
\log|\varphi(t)| &= -\frac{1}{2}\sum_{j=1}^{q}\sum_{k=0}^{\infty}\Big\{\log\Big(1+\frac{\tilde{t}^2}{(\frac{n-p+j}{2}+k)^2}\Big)-\log\Big(1+\frac{\tilde{t}^2}{(\frac{n+j}{2}+k)^2}\Big)\Big\} \\
&< -\frac{1}{2}\sum_{j=1}^{q}\int_0^\infty\Big\{\log\Big(1+\frac{\tilde{t}^2}{(\frac{n-p+j}{2}+x)^2}\Big)-\log\Big(1+\frac{\tilde{t}^2}{(\frac{n+j}{2}+x)^2}\Big)\Big\}dx \\
&= -\frac{1}{4}\sum_{j=1}^{q}\int_{n-p}^{n}\log\Big(1+\frac{4\tilde{t}^2}{(x+j)^2}\Big)dx \\
&< -\frac{1}{4}\int_1^{q+1}\Big\{\int_{n-p}^{n}\log\Big(1+\frac{4\tilde{t}^2}{(x+y)^2}\Big)dx\Big\}dy = -\tilde{t}^2\,G(\tilde{t},n,p,q),
\end{aligned}
\tag{3.13}
$$

where

$$
\begin{aligned}
G(t,n,p,q) &= F\Big(\frac{n+q+1}{2t}\Big)-F\Big(\frac{n+1}{2t}\Big)-F\Big(\frac{n-p+q+1}{2t}\Big)+F\Big(\frac{n-p+1}{2t}\Big), \\
F(x) &= \int_0^x\Big\{\int_0^y\log\Big(1+\frac{1}{z^2}\Big)dz\Big\}dy \\
&= \frac{x^2}{2}\log\Big(1+\frac{1}{x^2}\Big)+2x\arctan x-\frac{1}{2}\log(1+x^2).
\end{aligned}
$$

Therefore a bound for $I_3[v]$ is given by

$$
I_3[v] < \int_{m_0 v}^{\infty}\frac{2}{t}\exp\big\{-t^2\,G(t,n,p,q)\big\}dt = U_3[v] \quad \text{(say)}, \tag{3.14}
$$

where $m_0 = \frac{1}{2}(n-p-\frac{1}{2})$.

A simple bound for $I_3[v]$ is given as follows. Since $\log(1+\frac{1}{z^2})$ is convex

$$
\begin{aligned}
t^2 G(t,n,p,q) &= \frac{1}{4}\sum_{j=1}^{q}\int_{n-p}^{n}\log\Big(1+\frac{4t^2}{(x+j)^2}\Big)dx \\
&> \frac{p}{4}\sum_{j=1}^{q}\log\Big(1+\frac{4t^2}{n-\frac{p}{2}+j}\Big) > \frac{pq}{4}\log\Big\{1+\frac{4t^2}{(n-\frac{p-q-1}{2})^2}\Big\}.
\end{aligned}
\tag{3.15}
$$

Hence

$$
\begin{aligned}
I_3[v] &< \int_{m_0 v}^{\infty}\frac{2}{t}\Big\{1+\frac{4t^2}{(n-\frac{p-q-1}{2})^2}\Big\}^{-pq/4}dt \\
&= \int_{1+\alpha}^{\infty}(s-1)^{-1}s^{-pq/4}ds = \int_{1+\alpha}^{\infty}\sum_{k=0}^{\infty}s^{-pq/4-k-1}ds = \sum_{k=0}^{\infty}\frac{1}{\frac{pq}{4}+k}(1+\alpha)^{-pq/4-k} \\
&< \frac{4}{pq}(1+\alpha)^{-pq/4}\Big(1-\frac{1}{1+\alpha}\Big)^{-1} = \frac{4}{pq}(1+\alpha)^{-pq/4}\frac{1+\alpha}{\alpha} = \widetilde{U}_3[v] \text{ (say)},
\end{aligned}
\tag{3.16}
$$

where

$$\alpha = \frac{4m_0^2 v^2}{(n - \frac{p-q-1}{2})^2} = \frac{(n-p-\frac{1}{2})^2 v^2}{(n-p+\frac{p+q+1}{2})^2}. \tag{3.17}$$

The result obtained here is summarized in the following theorem.

Theorem 3.1 *Let $\widetilde{T}$ be the standardized statistic given by* (2.5) *and Q_s be the Edgeworth expansion of the distribution function of $\widetilde{T}$ up to the order $O(m^{-s})$ given by* (2.9)*. Then*

$$\sup_x \left| \mathrm{P}(\widetilde{T} \le x) - Q_s(x) \right| < \frac{1}{2\pi}(U_1[v] + I_2[v] + U_3[v]), \tag{3.18}$$

where U_1 and U_3 are given by (3.5) *and* (3.14)*. More simple bound is obtained by using $\widetilde{U}_1$ in* (3.8), (3.10) *with appropriate c and $\widetilde{U}_3$ in* (3.16)*.*

4 The order of the error

In this section we show the order of the error of the Edgeworth expansion when some of $n-p, p$ and q becomes large.

Table 1 gives the orders of m and b_s, where the notations $\gg, \sim$ and $\gtrsim$ are defined as

$$a \sim b \Leftrightarrow \frac{a}{b} \text{ and } \frac{b}{a} \text{ are bounded,}$$
$$a \gg b \Leftrightarrow \frac{b}{a} \to 0, \quad a \gtrsim b \Leftrightarrow \frac{b}{a} \text{ is bounded.}$$

Table 1: The orders of m and b_s

		$m \sim$	$b_s \sim$
(i)	$l \gtrsim p \gtrsim q \gg 1$	$(pq)^{1/2}$	1
(ii)	$l \gtrsim p \gg q \sim 1$	$p^{1/2}$	1
(iii)	$p \gg l \gtrsim q \gg 1$	$(lq)^{1/2}$	1
(iv)	$p \gg l \gg q \sim 1$	$l^{1/2}$	1
(v)	$p \gtrsim q \gg l \gg 1$	$l\{\log(q/l)\}^{1/2}$	$\{\log(q/l)\}^{-1}$
(vi)	$p \gtrsim q \gg l \sim 1$	$(\log q)^{1/2}$	$(\log q)^{-1}$

$(l = n - p)$

In section 3 we noted that (3.7) and (3.4) show that

$$I_1[v] + I_2[v] = O(m^{-(s+1)})$$

for any positive number s. So we look at the order of $I_3[v]$.

In the cases (i) and (ii) in table 1, some positive constant c exists such that α in (3.17) is larger than c for all n, p and q. Then (3.16) shows that

$$I_3[v] = O\Big(\frac{1}{pq}(1+c)^{-pq/4}\Big).$$

In the cases (iii) and (iv), α converges to 0. So we need another bound instead of $\widetilde{U}_3$. Narrowing the region of integration in (3.15),

$$\begin{aligned} t^2 G(t,n,p,q) &> \frac{1}{4}\int_1^{q+1}\Big\{\int_{n-p}^{n-p+p'} \log\Big(1+\frac{4t^2}{(x+y)^2}\Big)dx\Big\}dy \\ &> \frac{p'q}{4}\log\Big\{1+\frac{4t^2}{(n-p+1+\frac{p'+q}{2})^2}\Big\} \quad (0 < p' \le p). \end{aligned} \tag{4.1}$$

Similarly as (3.16) we obtain

$$I_3[v] < \frac{4}{p'q}(1+\alpha')^{-p'q/4}\frac{1+\alpha'}{\alpha'},$$

where

$$\alpha' = \frac{4m_0^2 v^2}{(n-p+\frac{p'+q+1}{2})^2} = \frac{(n-p-\frac{1}{2})^2 v^2}{(n-p+1+\frac{p'+q}{2})^2}.$$

Taking $p' = l = n-p$ shows that

$$I_3[v] = O\Big(\frac{1}{lq}(1+c')^{-lq/4}\Big) = o(m^{-(s+1)})$$

for any positive number s where c' is some positive constant.

In the cases (v) and (vi) in table 1, α' also converges to 0. However the order of $I_3[v]$ is still $O[(lq)^{-1}]$, which shows that

$$I_3[v] = O\Big((\log q)^{-(s+1)/2}\Big)$$

for all positive s. Hence only the case (v) is remained. In this case we can see that

$$\kappa^{(2)} = O\Big(\log\frac{q}{n-p}\Big).$$

Since $\log(1+x) - (x - \frac{1}{2}x^2)$ is a nonnegative increasing function,

$$\begin{aligned} \log|\varphi(t)| &= -\frac{1}{2}\sum_{j=1}^{q}\sum_{k=0}^{\infty}\Big\{\log\Big(1+\frac{t^2}{\kappa^{(2)}(\frac{n-p+j}{2}+k)^2}\Big) - \log\Big(1+\frac{t^2}{\kappa^{(2)}(\frac{n+j}{2}+k)^2}\Big)\Big\} \\ &\le -\frac{1}{2}\sum_{j=1}^{q}\sum_{k=0}^{\infty}\Big[\Big\{\frac{t^2}{\kappa^{(2)}(\frac{n-p+j}{2}+k)^2} - \frac{t^2}{\kappa^{(2)}(\frac{n+j}{2}+k)^2}\Big\} \\ &\qquad - \frac{1}{2}\Big\{\Big(\frac{t^2}{\kappa^{(2)}(\frac{n-p+j}{2}+k)^2}\Big)^2 - (\frac{t^2}{\kappa^{(2)}(\frac{n+j}{2}+k)^2}\Big)^2\Big] \\ &= -\frac{1}{2}t^2\Big\{1-\frac{1}{12}t^2\widetilde{\kappa}^{(4)}\Big\} \le -\frac{1}{2}t^2\Big\{1-\frac{1}{6}t^2 m^{-2}(\kappa^{(2)})^{-1}\Big\}, \end{aligned} \tag{4.2}$$

where we used (2.4) and (2.3) for the last inequality. We divide the integral $I_3[v]$ into two part :

$$I_3[v] = I_{31}[v] + I_{32}[v],$$
$$I_{31}[v] = \int_{mv}^{m(\kappa^{(2)})^{1/2}} \frac{2}{t}|\varphi(t)|dt, \quad I_{32}[v] = \int_{m(\kappa^{(2)})^{1/2}}^{\infty} \frac{2}{t}|\varphi(t)|dt.$$

From (4.2)

$$\begin{aligned} I_{31}[v] &\le \int_{mv}^{m(\kappa^{(2)})^{1/2}} \frac{2}{t}\exp(-ut^2) = \int_{m^2v^2u^2}^{m^2\kappa^{(2)}u^2} \frac{1}{s}e^{-s}ds \\ &\le \frac{1}{m^2v^2u^2}\exp\Big\{-m^2v^2u^2\Big\} = U_{31}[v] \text{ (say)}, \end{aligned} \tag{4.3}$$

where $u = \frac{5}{12}$.

Similarly to (4.1), we obtain

$$t^2G(t,n,p,q) > \frac{q'^2}{4}\log\Big\{1 + \frac{4t^2}{(n-p+1+q')^2}\Big\},$$

where $q' = \min\{q, m\}$. Hence

$$I_{32}[v] < \frac{4}{q'^2}(1+\alpha'')^{-q'^2/4}\frac{1+\alpha''}{\alpha''},$$

where

$$\alpha'' = \Big(\frac{2m}{n-p+1+q'}\Big)^2 \ge \Big(\frac{2m}{(n-p+1+m)}\Big)^2 \to 4. \tag{4.4}$$

Therefore $I_{31}[v] + I_{32}[v] = o(m^{-(s+1)})$ for any positive number s since $q'^2 \gg m$ in the cases of (v) and (vi).

5 Examples

In this section we show some tables of error bounds.

In the case of $s = 2$, $\varphi_s(t)$ in (2.8) becomes

$$\varphi_2(t) = \exp\Big(-\frac{t^2}{2}\Big)\Big\{1 + \frac{1}{6}\widetilde{\kappa}^{(3)}(it)^3 + \frac{1}{72}(\widetilde{\kappa}^{(3)})^2(it)^6 + \frac{1}{24}\widetilde{\kappa}^{(4)}(it)^4\Big\}.$$

Hence $I_2[v]$ in (3.13) is

$$I_2[v] = 2\int_{mv}^{\infty} \exp\Big(-\frac{t^2}{2}\Big)\Big\{\frac{1}{t} + \frac{1}{6}\widetilde{\kappa}^{(3)}t^2 + \frac{1}{72}(\widetilde{\kappa}^{(3)})^2t^5 + \frac{1}{24}\widetilde{\kappa}^{(4)}t^3\Big\}dt$$

$U_1[v]$ in (3.5) is given by

$$U_1[v] = \frac{2}{m^3}\int_0^{mv}\Big\{\Big(R_{1,2}[v]t^4 + \frac{1}{2}R_{2,1}[v]t^6\Big)\exp\Big(-\frac{t^2}{2}\Big) + \frac{1}{6}(B[v])^3t^8\exp\Big(-\frac{t^2}{2}c_v\Big)\Big\}dt,$$

where $B[v]$ is given by (3.3),

$$R_{1,2}[v] = \frac{1}{v^2}\{B[v] - b_0 - b_1 v\} \text{ and } R_{2,1}[v] = \frac{1}{v}\{(B[v])^2 - b_0^2\}.$$

$U_3[v]$ and $\widetilde{U}_3[v]$ given by (3.14) and (3.16) do not depend on the choice of s.

Although we can minimize the bound given in theorem 3.1 with respect to v numerically, it is sufficient to calculate the bounds at

$$v = 0.05, 0.10, \cdots, 0.95$$

and choose the minimum for actual use. Table 2 gives the bounds

$$\text{BOUND–1} = \min_{v=0.05,\cdots,0.95} \frac{1}{2\pi}\big(U_1[v] + I_2[v] + U_3[v]\big),$$
$$\text{BOUND–2} = \min_{v=0.05,\cdots,0.95} \frac{1}{2\pi}\big(U_1[v] + I_2[v] + \widetilde{U}_3[v]\big)$$

for

$$n = 50;\ p = 10, 20, 30, 40\ ; q = 5, 10.$$

The values in the parentheses are the values of v at the minimum.

We can see that (i) if q becomes large the bounds becomes small, (ii) if $n-p$ small the bounds becmes a little bit large. The simple version BOUND–2 suffices for actual use except the cases that $n-p$ is small relative to p.

Table 2: The error bounds for $s = 2$ and $n = 50$.

p	q	BOUND–1	BOUND–2
10	5	0.0272 (0.55)	0.0288 (0.55)
20	5	0.0110 (0.55)	0.0114 (0.55)
30	5	0.0077 (0.55)	0.0085 (0.60)
40	5	0.0100 (0.70)	0.0159 (0.75)
10	10	0.0099 (0.50)	0.0101 (0.50)
20	10	0.0035 (0.45)	0.0037 (0.45)
30	10	0.0023 (0.50)	0.0026 (0.50)
40	10	0.0029 (0.65)	0.0044 (0.70)

Table 3 gives the bounds for different orders of the expansion formulas and

$$n = 50;\ p = 10, 20;\ q = 5.$$

We can see that the bounds are similar for $s = 2, 3, 4$ if $p = 10$, while the bounds becomes small as s becomes large if $p = 20$.

The error bound for Box's formula (1.2) given by Fujikoshi and Ulyanov (2006) is not very sharp because it is an application of the formula of error bounds for wide class of statistics. It will be possible to derive a sharp bound for (1.2) using similar method in this paper. Then we will be able to choose better approximation formula between Box' formula and Edgeworth expansion formula.

Table 3: The error bounds for different orders with $n = 50, q = 5$.

s	p	BOUND–1	BOUND–2
0	10	0.0621	0.0676
1	10	0.0368	0.0398
2	10	0.0272	0.0288
3	10	0.0237	0.0244
4	10	0.0238	0.0246
0	20	0.0406	0.0429
1	20	0.0181	0.0192
2	20	0.0110	0.0114
3	20	0.0078	0.0083
4	20	0.0061	0.0063

References

ANDERSON, T. W. (1984). *An introduction to multivariate statistical analysis* (2nd ed.), John Wiley & Sons, New York.

BOX, G. E. P (1949). A general distribution theory for a class of likelihood criteria, *Biometrika*, **36**, 317–346.

FUJIKOSHI, Y. and ULYANOV, V. (2006). Error bounds for asymptotic expansions ofWilks' lambda distribution, *J. Multivariate Analysis*, **97**, 1941–1957.

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

SIOTANI, M., HAYAKAWA, T. and FUJIKOSHI, Y. (1985). *Modern multivariate statistical analysis: a graduate course and handbook*, American Sciences Press, Columbus, Ohio.

TONDA, T. and FUJIKOSHI, Y. (2004). Asymptotic expansion of the null distribution of LR statistic for multivariate linear hypothesis when the dimension is large, *Comm. Statist.-Theory Meth.*, **33**, 1205-1220.

ULIANOV, V., WAKAKI, H. and FUJIKOSHI, Y. (2006). Berry.Esseen bound for high dimensional asymptotic approximation of Wilks 'Lambda distribution, *Statistics & Probability Letters*, **76**, 1191–1200.

WAKAKI, H. (2006). Edgeworth expansion of Wilks' lambda statistic, *J. Multivariate Analysis*, **97**, 1958–1964.

WAKAKI, H. and FUJIKOSHI, Y. and ULYANOV, V. (2002). Asymptotic Expansions of the Distributions of MANOVA Test Statistics when the Dimension is Large, *Hiroshima Statistical Group Technical Report*, **10**, 1-10.