High-dimensional and large-sample approximations in statistics

Vladimir V. Ulyanov

Moscow State University
Faculty of Computational Mathematics and Cybernetics

Russian-Chinese Seminar on Asymptotic Methods in Probability Theory and Mathematical Statistics
June 10–14, 2013, Saint Petersburg

Structure

1. Quadratic forms
2. Almost quadratic forms
3. Non-quadratic forms

Parts 1 and 2 are based on
Yu.V. Prokhorov and V.V. Ulyanov, “Some approximation problems in statistics and probability”, *Limit Theorems in Probability, Statistics and Number Theory. Berlin–Heidelberg: Springer-Verlag*, 42, 235–252 (2013)

Part 3 is based on
Y. Fujikoshi, V.V. Ulyanov and R. Shimizu, *Multivariate statistics: High-dimensional and large-sample approximations*, John Wiley and Sons, Hoboken, NJ, (2010)

Notation

$X, X_1, X_2, \ldots$ – i.i.d. random elements in a Hilbert space H.
$dim(H)$ – dimension of H, may be infinite or finite

Let (x, y) for $x, y \in H$ denote the inner product in H.
Put $|x| = (x, x)^{1/2}$.
Assume $\mathbf{E}|X_1|^2 < \infty$. For simplicity $\mathbf{E}X_1 = 0$ and $\mathbf{E}|X_1|^2 = 1$
Denote by V a covariance operator of X_1

$$(Vx, y) = \mathbf{E}(X_1, x)(X_1, y).$$

Let $\sigma_1^2 \geqslant \sigma_2^2 \geqslant \ldots$ be the eigenvalues of V Put

$$S_n = n^{-1/2} \sum_{i=1}^{n} X_i,$$

Let Y be a H-valued Gaussian $(0, V)$ random element.

Many asymptotic problems in probability theory and statistics can be described in terms of closeness of $f(S_n)$ and $f(Y)$.

If $f(x)$ is linear we get two gems of probability theory – law of large numbers and CLT.

It would be natural to extend the results to quadratic forms. Moreover, in mathematical statistics there are numerous asymptotic problems which can be formulated in terms of quadratic or almost quadratic forms.

Notation

Let Y be H-valued Gaussian $(0, V)$ random element. Put for any $a \in H$

$$F(x) = P\{|S_n - a|^2 \leqslant x\}, \quad F_0(x) = P\{|Y - a|^2 \leqslant x\},$$

$$\delta_n(a) = \sup_x |F(x) - F_0(x)|.$$

It is known (see e.g. Sazonov (1968) and Bentkus (1986)) that in the case $H = \mathbf{R}^d$, $d < \infty$, i.e. in the finite dimensional case, we have

$$\delta_n(a) \leqslant c \, \mathbf{E}|X_1|^3 \, \sigma_d^{-3} \, n^{-1/2}$$

and the bound is optimal with respect to the dependence on moments, eigenvalues as well as n.

However in the infinite dimensional case the situation essentially changes. Here we have (see e.g. Sazonov (1981), ch.2)

$$\sup_a \delta_n(a) \geqslant 1/2.$$

History, 3 phases

The improvements of bounds for $\delta_n(a)$ in the infinite dimensional case can be divided roughly into three phases: proving bounds with optimal

- dependence on n;
- moment conditions;
- dependence on the eigenvalues of V.

History, 3 phases

The improvements of bounds for $\delta_n(a)$ in the infinite dimensional case can be divided roughly into three phases: proving bounds with optimal

- dependence on n;
- moment conditions;
- dependence on the eigenvalues of V.

History, 3 phases

The improvements of bounds for $\delta_n(a)$ in the infinite dimensional case can be divided roughly into three phases: proving bounds with optimal

- dependence on n;
- moment conditions;
- dependence on the eigenvalues of V.

History. Approximation. Third phase

The first phase started in the middle of 60-s with bounds of logarithmic order for $\delta_n(a)$ (see Kandelaki (1965))

At the end of the third phase it was proved (see Sazonov, Ulyanov and Zalesskii (1988b), Nagaev (1989), Senatov (1989))

$$\delta_n(a) \leqslant \frac{c\ c_6(V)}{\sqrt{n}}\,(1+|a|^3)\,\mathbf{E}|X_1|^3, \tag{1}$$

where

$$c_k(V) = \prod_1^k \sigma_i^{-1}$$

It is known (see Senatov (1985)) that for any $c_0 > 0$ and for any given eigenvalues $\sigma_1^2, \dots, \sigma_6^2 > 0$ of a covariance operator V there exist a vector $a \in H$, $|a| > c_0$, and a sequence $X_1, X_2, \dots$ of i.i.d. random elements in H with zero mean and covariance operator V such that

$$\liminf_{n\to\infty} \sqrt{n}\, \delta_n(a) \geqslant c\, c_6(V)\, (1 + |a|^3)\, \mathbf{E}|X_1|^3. \qquad (2)$$

Due to (2) the bound (1) is the best possible in case of finite third moment of $|X_1|$.

Notation, first term of expansion

Better approximations for $F(x)$ are available when we use an additional term, say $F_1(x)$, of its asymptotic expansion. This term $F_1(x)$ is defined as the unique function satisfying $F_1(-\infty) = 0$ with Fourier-Stieltjes transform equal to

$$\hat{F}_1(t) = -\frac{2t^2}{3\sqrt{n}} \mathbf{E} e\{t|Y-a|^2\} \left(3(X, Y-a)|X|^2 + 2it(X, Y-a)^3\right). \tag{3}$$

Here and in the following we write $e\{x\} = \exp\{ix\}$.

Introduce the error

$$\Delta_n(a) = \sup_x |F(x) - F_0(x) - F_1(x)|.$$

Note that $\hat{F}_1(t) = 0$ and hence $F_1(x) = 0$ when $a = 0$ or X has a symmetric distribution, i.e. when X and $-X$ are identically distributed. Therefore, we get

$$\Delta_n(0) = \delta_n(0).$$

History. Asymptotic expansions. First phase

Similar to the developments of bounds for $\delta_n(a)$ the first task consisted in deriving the bounds for $\Delta_n(a)$ with optimal dependence on n. Starting with a seminal paper by Esseen (1945) for finite dimensional spaces $H = \mathbf{R}^d,\ d < \infty$ who proved

$$\Delta_n(0) = \mathcal{O}(n^{-d/(d+1)}), \tag{4}$$

a comparable bound

$$\Delta_n(0) = \mathcal{O}(n^{-\gamma})$$

with $\gamma = 1 - \varepsilon$ for any $\varepsilon > 0$ was finally proved in Götze(1979, 1984), based on Weyl type inequalities

History, First and second phases

Fifty years after Essen's result the optimal bounds (in n and moments)

$$\Delta_n(0) \leqslant \frac{c(9,V)}{n}\, \mathbf{E}|X_1|^4, \tag{5}$$

$$\Delta_n(a) \leqslant \frac{c(13,V)}{n}\, (1+|a|^6)\, \mathbf{E}|X_1|^4, \tag{6}$$

where $c(i,V) \leqslant \exp\{c\sigma_i^{-2}\}, \quad i = 9, 13$, were finally established in Bentkus and Götze (1997), using new techniques which allowed to prove optimal bounds in classical lattice point problems as well.

The bounds (5) and (6) are optimal with respect to the dependence on n and on the moments.
The bound (5) improves as well Esseen's result (4) for Euclidean spaces $\mathbf{R}^d$ with $d > 8$.

However the dependence on covariance operator V in (5), (6) can be improved.

Main result

Theorem. Götze and Ulyanov (2011) *There exists an absolute constant c such that for any $a \in H$*

$$\Delta_n(a) \leqslant \frac{c}{n} \cdot c_{12}(V) \cdot \left(\mathbf{E}|X_1|^4 + \mathbf{E}(X_1, a)^4\right) \times \left(1 + (Va, a)\right), \tag{7}$$

where

$$c_{12}(V) = \prod_1^{12} \sigma_i^{-1}$$

Optimality

It follows from Götze and Ulyanov (2000) that for any given eigenvalues $\sigma_1^2, \ldots, \sigma_{12}^2 > 0$ of a covariance operator V there exist $a \in H$, $|a| > 1$, and a sequence $X_1, X_2, \ldots$ of i.i.d. random elements in H with zero mean and covariance operator V such that

$$\liminf_{n\to\infty} n\,\Delta_n(a) \geqslant c\; c_{12}(V)\,(1+|a|^6)\,\mathbf{E}|X_1|^4.$$

Optimality

Thus, (7) is the best bound in the following sense:

- it is impossible that $\Delta_n(a)$ is of order $\mathcal{O}(n^{-1})$ *uniformly* for distributions of X_1 with arbitrary eigenvalues $\sigma_1^2, \sigma_2^2, \ldots$;
- the form of dependence on the eigenvalues of V, on n and on $\mathbf{E}|X_1|^4$ in (7) coincides with one given in lower bound.

Case $a = 0$

In special cases the number of eigenvalues which are necessary for optimal bounds may well decrease below 12.
Lower bounds for $n\Delta_n(a)$ in the case $a = 0$ are not available. A conjecture in Götze (1998) said that in that case the five first eigenvalues of V suffice. That conjecture was confirmed in Götze and Zaitsev (2008) with result $\Delta_n(0) = \mathcal{O}(n^{-1})$ provided that $\sigma_5 > 0$ only.
Note that for some centered ellipsoids in $\mathbf{R}^d$ with $d \geq 5$ the bounds of order $\mathcal{O}(n^{-1})$ were obtained in Götze and Ulyanov (2003).

Case $a \neq 0$ and $dim(H) < \infty$

It was proved recently (see Götze and Zaitsev (2010)) that even for $a \neq 0$ we have $\Delta_n(a) = \mathcal{O}(n^{-1})$ when $H = \mathbf{R}^d$, $5 \leq d < \infty$, and the upper bound for $\Delta_n(a)$ depends on the smallest eigenvalue σ_d.

It is necessary to emphasize that (7) implies $\Delta_n(a) = \mathcal{O}(n^{-1})$ for infinite dimensional space H with dependence on first twelve eigenvalues only.

Proof

The proofs of recent results due to Götze, Ulyanov and Zaitsev are based on the reduction of the original problem to lattice valued random vectors and on symmetrization techniques developed in number of papers, see e.g. Götze (1979), Yurinskii (1982), Sazonov, Ulyanov and Zalesskii (1988a, 1988b, 1991), Götze and Ulyanov (2000), Bogatyrev, Götze and Ulyanov (2006). In the proofs we use also the new inequalities obtained in Lemma 6.5 in Götze and Zaitsev (2011) and in Götze and Margulis (2010) (see Lemma 8.2 in Götze and Zaitsev (2011)). In fact, the bounds in Götze and Zaitsev (2011) are constructed for more general quadratic forms of the type $(\mathbb{Q}x, x)$ with non-degenerate linear symmetric bounded operator in $\mathbf{R}^d$.

Notation

- $(Y_1, \ldots, Y_k)^T$ – vector with multinomial distribution $M_k(n, \pi)$:

$$\Pr(Y_1 = n_1, \ldots, Y_k = n_k) = \begin{cases} n! \prod_{j=1}^{k} \pi_j^{n_j} / n_j!, & n_1 + \cdots + n_k = n \\ 0, & \text{otherwise} \end{cases}$$

- Hypothesis

$$H_0 : \pi = \boldsymbol{p}$$

Notation

- $(Y_1, \ldots, Y_k)^T$ – vector with multinomial distribution $M_k(n, \pi)$:

$$\Pr(Y_1 = n_1, \ldots, Y_k = n_k) = \begin{cases} n! \prod_{j=1}^{k} \pi_j^{n_j} / n_j!, & n_1 + \cdots + n_k = n \\ 0, & \text{otherwise} \end{cases}$$

- Hypothesis

$$H_0 \colon \pi = \boldsymbol{p}$$

Goodness-of-fit tests

- Chi-square statistics

$$t_1(\boldsymbol{Y}) = \sum_{j=1}^{k} \frac{(Y_j - np_j)^2}{np_j}$$

- Log-likelihood ratio statistics

$$t_0(\boldsymbol{Y}) = 2 \sum_{j=1}^{k} Y_j \log(Y_j/(np_j))$$

- Power-divergence family of statistics (Cressie and Read, 1984)

$$t_\lambda(\boldsymbol{Y}) = \frac{2}{\lambda(\lambda+1)} \sum_{j=1}^{k} Y_j \left[\left(\frac{Y_j}{np_j} \right)^\lambda - 1 \right], \ \lambda \in \mathbb{R}$$

Goodness-of-fit tests

- Chi-square statistics

$$t_1(\boldsymbol{Y}) = \sum_{j=1}^{k} \frac{(Y_j - np_j)^2}{np_j}$$

- Log-likelihood ratio statistics

$$t_0(\boldsymbol{Y}) = 2 \sum_{j=1}^{k} Y_j \log(Y_j/(np_j))$$

- Power-divergence family of statistics (Cressie and Read, 1984)

$$t_\lambda(\boldsymbol{Y}) = \frac{2}{\lambda(\lambda+1)} \sum_{j=1}^{k} Y_j \left[\left(\frac{Y_j}{np_j} \right)^\lambda - 1 \right], \ \lambda \in \mathbb{R}$$

Goodness-of-fit tests

- Chi-square statistics

$$t_1(\boldsymbol{Y}) = \sum_{j=1}^{k} \frac{(Y_j - np_j)^2}{np_j}$$

- Log-likelihood ratio statistics

$$t_0(\boldsymbol{Y}) = 2\sum_{j=1}^{k} Y_j \log(Y_j/(np_j))$$

- Power-divergence family of statistics (Cressie and Read, 1984)

$$t_\lambda(\boldsymbol{Y}) = \frac{2}{\lambda(\lambda+1)} \sum_{j=1}^{k} Y_j \left[\left(\frac{Y_j}{np_j} \right)^{\lambda} - 1 \right], \ \lambda \in \mathbb{R}$$

Formulation of the Problem

- It is well known that

$$t_\lambda(\boldsymbol{Y}) \Rightarrow \chi^2_{k-1} \quad \text{as} \quad n \to \infty$$

- We study the rate of convergence

$$\Pr(t_\lambda(\boldsymbol{Y}) < c) = \Pr(\chi^2_{k-1} < c) + O(n^?)$$

Formulation of the Problem

- It is well known that

$$t_\lambda(\boldsymbol{Y}) \Rightarrow \chi^2_{k-1} \quad \text{as} \quad n \to \infty$$

- We study the rate of convergence

$$\Pr(t_\lambda(\boldsymbol{Y}) < c) = \Pr(\chi^2_{k-1} < c) + O(n^{?})$$

- Yarnold (1972) proved for chi-square statistics

$$\Pr(t_1(\boldsymbol{Y}) < c) = \Pr(\chi^2_{k-1} < c) + O(n^{-\boldsymbol{(k-1)/k}})$$

- Götze and Ulyanov (2003) for $k > 5$

$$\Pr(t_1(\boldsymbol{Y}) < c) = \Pr(\chi^2_{k-1} < c) + O(n^{-\boldsymbol{1}})$$

History, Pearson statistics

- Yarnold (1972) proved for chi-square statistics

$$\Pr(t_1(\boldsymbol{Y}) < c) = \Pr(\chi^2_{k-1} < c) + O(n^{-\boldsymbol{(k-1)/k}})$$

- Götze and Ulyanov (2003) for $k > 5$

$$\Pr(t_1(\boldsymbol{Y}) < c) = \Pr(\chi^2_{k-1} < c) + O(n^{-\boldsymbol{1}})$$

$J(n)$ term

We take Yarnold's expansion

$$\Pr(t_\lambda(\boldsymbol{Y}) < c) = \Pr(\chi^2_{k-1} < c) + J(n) + O(n^{-\mathbf{1}})$$

where

$$J(n) = -\frac{1}{\sqrt{n}} \sum\nolimits_{l=1}^{r} n^{-(r-l)/2} \sum\nolimits_{x_{l+1} \in L_{l+1}} \cdots \sum\nolimits_{x_r \in L_r} \left[\int \cdots \int_{B_l^\lambda} [S_1(\sqrt{n} x_l + np_l)\varphi(\boldsymbol{x})]_{\lambda_l(x^*)}^{\theta_l(x^*)} \, dx_1, \cdots, dx_{l-1} \right]$$

$$L_j = \left\{ \boldsymbol{x} \colon \; x_j = \frac{n_j - np_j}{\sqrt{n}}, \text{ with } p_j \text{ defined as before} \right\}$$

$$S_1(x) = x - \lfloor x \rfloor - 1/2, \; \lfloor x \rfloor \text{ is the integer part of } x$$

Transformation of statistics

We consider transformation

$$X_j = (Y_j - np_j)/\sqrt{n},\ j = 1, \dots, k,\ r = k-1,\ \boldsymbol{X} = (X_1, \dots, X_r)^T.$$

The statistic $t_\lambda(\boldsymbol{Y})$ can be expressed as a function of $\boldsymbol{X}$ in the form

$$T_\lambda(\boldsymbol{X}) = \frac{2n}{\lambda(\lambda+1)} \left[\sum_{j=1}^{k} p_j \left(\left(1 + \frac{X_j}{\sqrt{n p_j}} \right)^{\lambda+1} - 1 \right) \right],$$

and then, via the Taylor's expansion, transformed to the form

$$T_\lambda(\boldsymbol{X}) = \sum_{i=1}^{k} \left(\frac{X_i^2}{p_i} + \frac{(\lambda-1)X_i^3}{3\sqrt{n}p_i^2} + \frac{(\lambda-1)(\lambda-2)X_i^4}{12 p_i^3 n} + O\left(n^{-3/2}\right) \right).$$

$J(n)$ term

Write

$$\Pr(t_\lambda(\boldsymbol{Y}) < c) = \Pr(T_\lambda(\boldsymbol{X}) < c) = \Pr(\boldsymbol{X} \in B^\lambda)$$

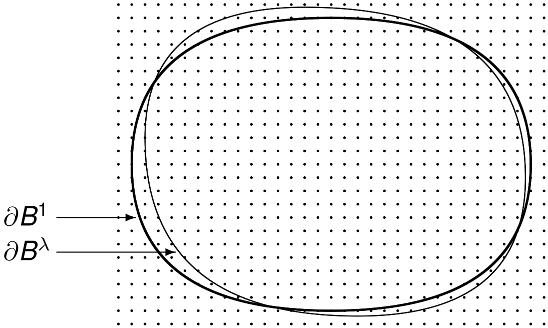
Shiotani, Fujikoshi and Read (1984) showed

$$J(n) = (N^\lambda - n^{r/2} V^\lambda)\, e^{-c/2} \Big/ \left((2\pi n)^r \prod\nolimits_{j=1}^{k} p_j \right)^{1/2} + o(1)$$

Ulyanov and Zubov (2009):

$$J(n) = (N^\lambda - n^{r/2} V^\lambda)\, e^{-c/2} \Big/ \left((2\pi n)^r \prod\nolimits_{j=1}^{k} p_j \right)^{1/2} + O(n^{-1})$$

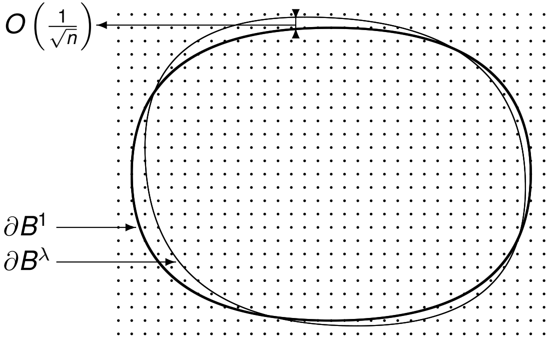
Studied objects



$$B^\lambda = \{\boldsymbol{x} : T_\lambda(\boldsymbol{x}) < c\},$$

L – lattice with lattice spacing $n^{-1/2}$.

Yarnold's argument

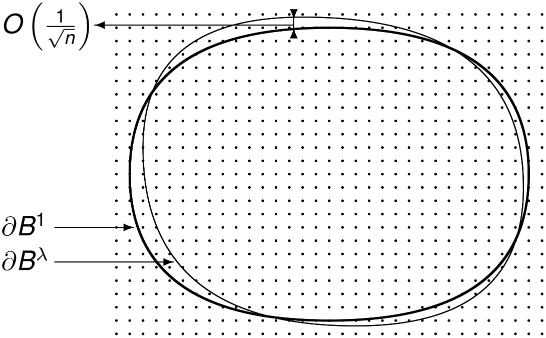


$B^\lambda = \{T_\lambda < c\}$

L – lattice with lat. sp. $\frac{1}{\sqrt{n}}$

$$J(n) = dn^{-\frac{k-1}{2}}\left(N^\lambda - n^{\frac{k-1}{2}}V^\lambda\right) + O(n^{-1}).$$

Yarnold's argument



$B^\lambda = \{T_\lambda < c\}$

L – lattice with lat. sp. $\frac{1}{\sqrt{n}}$

$$J(n) = dn^{-\frac{k-1}{2}}\left(N^\lambda - n^{\frac{k-1}{2}}V^\lambda\right) + O(n^{-1}).$$

Hlawka result

E. Hlawka (1950): Let B be a compact convex set in $\mathbb{R}^r$ with the origin as its inner point. We denote the volume of this set by V. Assume that the boundary of this set is an $(r-1)$-dimensional surface of class $\mathbb{C}^\infty$, the Gaussian curvature being non-zero and finite everywhere on the surface. Also assume that a specially defined canonical map from the unit sphere to B is one-to-one and belongs to the class $\mathbb{C}^\infty$. Then in the set that is obtained from the initial one by translation along an arbitrary vector and by linear expansion with the factor $\sqrt{n}$ the number of integer points is

$$N = n^{r/2} V + O\left(D\, n^{r/2 - r/(r+1)}\right),$$

where the constant D is a number dependent only on the properties of the curve C, but not on the parameters n or V.

Results

- Ulyanov and Zubov (2009) for $k \geq 4$

$$Pr(t_\lambda(\mathbf{Y}) < c) = Pr(\chi_{k-1} < c) + O(n^{-(k-1)/k})$$

- Assylbekov, Ulyanov and Zubov (2011) for $k = 3$

$$Pr(t_\lambda(\mathbf{Y}) < c) = Pr(\chi_2 < c) + O(n^{-3/4+\beta})$$

with $\beta = 0.065$

Multivariate Linear Model

Model

$$Y = X\Theta + (\varepsilon_1, \ldots, \varepsilon_n)'$$

$$\begin{aligned}&\varepsilon_1, \ldots, \varepsilon_n \sim i.i.d.\\ &\mathrm{E}[\varepsilon_j] = 0, \quad Cov(\varepsilon_j) = \Sigma\end{aligned}$$

Hypothesis Testing

$$H_0 : C\Theta = 0$$

SS & SP Matrices

$$\begin{aligned} S_h &= \hat{\Theta}' C' \{C(X'X)^{-1}C'\}^{-1} C\hat{\Theta} \\ S_e &= Y'(I_n - X(X'X)^{-1}X')Y \end{aligned}$$

$$\hat{\Theta} = (X'X)^{-1}X'Y$$

Test Statistics

(i) $T_{LR} = -n \log\{|S_e|/|S_e + S_h|\}$

(ii) $T_{LH} = T_0^2 = n\,\mathrm{tr} S_h S_e^{-1}$

$$S_e \sim W_p(n, \Sigma) \quad S_h \sim W_p(q, \Sigma)$$

$$S_e \perp S_h \quad \text{independent; } \Sigma = I$$

The case $p = 1$

(i) $T_{LR} = n \log\left(1 + \frac{1}{n} T_0^2\right)$

(ii) $T_{LH} = T_0^2 = \left(\frac{1}{n}\chi_n^2\right)^{-1} \chi_q^2$

Error Bound for Asymptotic Expansions of $\Pr(T_0^2 \le x)$ UFS (2005)

$$|\Pr(T_0^2 \le x) - G_r(x) - \frac{r}{4n}\{(q-p-1)G_r(x)$$

$$-2qG_{r+2}(x) + (q+p+1)G_{r+4}(x)\}|$$

$$\le \frac{r}{48n^2}(|h_1| + |h_2| + 48q)$$

$$+\frac{1}{2n^2}p(2p^2+5p+5)\min\{\eta_{-1,4,p}, \nu_{-1,4,p}\}$$

Function of Multivariate scale mixture

An Expression for $T_0^2 (= T_{LH})$

$$T_{LH} = T_0^2 = n\,\mathrm{tr} S_h S_e^{-1} = X_1 + \ldots + X_p = f(S \cdot Z),$$

where

(i) $X_i = S_i Z_i, \quad i = 1, \ldots, p$
(ii) $Z_1, \ldots, Z_p \sim$ i.i.d. $\quad \chi_q^2$
(iii) $S_i = Y_i^{-1} (i = 1, \ldots, p)$ and

$$Y_1 > \ldots > Y_p > 0$$

are the characteristic roots of W and $nW \sim W_p(n, I_p)$

$$
\begin{aligned}
T_{LR} &= n \log \Lambda = n \log \prod_{j=1}^{p} V_j \\
& \qquad V_j \sim Be\left(\frac{1}{2}(n-j+1), \frac{1}{2}q\right) \\
&= \sum_{i=1}^{p} n \log(1 + X_i/n)
\end{aligned}
$$

$X = (X_1, \ldots, X_p)'$
(i) $X_i = S_i Z_i, \quad i = 1, \ldots, p$
(ii)$Z_1, \ldots, Z_p \sim i.i.d. \chi_q^2$
(iii) $S_i = Y_i^{-1} (i = 1, \ldots, p)$ and $nY_i \sim \chi_{m_i}^2,\ m_i = n - (i-1)$

High Dimensional Case, U-Fujikoshi-Wakaki (2006)

The distribution of $\Lambda = |S_e|/|S_e + S_h|$ can be regarded as

$$\begin{aligned} \Lambda_{q,p,n+q-p} &= \prod_{j=1}^{q} Be(\frac{1}{2}m_j, \frac{1}{2}p_j) \\ &= \prod_{j=1}^{q} \left(1 + \chi^2_{p_j}/\chi^2_{m_j}\right)^{-1}, \end{aligned}$$

where $p_j = p,\ m_j = n + q - p + 1 - j,\ j = 1, \dots, q$ and all the χ^2-variables are independent.

Further, the limiting distribution of

$$T_{LR} = \frac{\sqrt{p}}{a\sqrt{q}} \{-\log \Lambda - q \log(1+q)\}$$

provided that

$$p/n \to c \in (0,1),$$

is the standard normal distribution (see, e.g., Tonda and Fujikoshi (2004)), where

$$r = \frac{p}{m}, \quad m = n - p + q, \quad a = \frac{\sqrt{2}r}{\sqrt{1+r}}.$$

How to get

$$\sup_x | \Pr(T_{LR} \le x) - \Phi(x)| \le D?$$

Theorem. UFW (2006) Let $X_1, \ldots, X_p$ be i.i.d. positive random variables and $Y_1, \ldots, Y_n$ be i.i.d. positive random variables.
Let $EX_1 = EY_1$, $Var(X_1) = Var(Y_1) = \sigma^2$ and
$E|X_1|^3 = E|Y_1|^3 = \beta$.
Put $S_p(X) = (X_1 + \ldots + X_p)/p$ and
$S_n(Y) = (Y_1 + \ldots + Y_n)/n$.
Then for

$$A = \left(\frac{np}{\sigma^2(n+p)} \right)^{1/2}$$

we have

$$\sup_x \left| \Pr\left(A \left(\frac{S_p(X)}{S_n(Y)} - 1 \right) \leq x \right) - \Phi(x) \right| \leq c_0 \frac{\beta}{\sigma^3} \left(\frac{1}{n} + \frac{1}{p} \right)^{1/2}$$