Filomat 35:15 (2021), 5253–5261
https://doi.org/10.2298/FIL2115253L

Published by Faculty of Sciences and Mathematics,
University of Niš, Serbia
Available at: http://www.pmf.ni.ac.rs/filomat

The Uniform Asymptotic Normality of a Matrix-T Distribution

Kai Can Li[a], Jian Lin Zhang[b], Di Shou Mao[a]

[a]*School of Mathematics and Statistics, Hubei Normal University, Huangshi, 435002, China*
[b]*Huangzhou Middle School, Huanggang, 438000, China*

Abstract. Using the Kullback-Leibler distance between two density functions about a matrix T distribution and a matrix normal distribution,we obtain a Berry-Esseen boundary for the T distribution. Further, we give the condition under which a matrix T is uniformly asymptotically matrix normal distribution, and point out the convergence rate.

1. Introduction

As an extension of the one-dimensional Student's t distribution, the multivariate t distribution is closer to the real data with respect to the normal distribution due to the tail characteristics, therefore, it has been widely used in cluster analysis, discriminant analysis and regression analysis. To see Kotz (2004), Roth (2013)[1, 3]. In recent years, Jiang et al. (2020)[8] have applied multivariate t distributions to causal analysis. In the study of Bayes analysis, Dickey (1967)[5]introduced matrix T distributions and discussed their properties, and since then matrix T distributions have received much attention from many researchers, a variety of t distributions, see Chapters 4 and 5 in Kotz (2004)[1], were proposed and extend to related fields. For example, studying spherically symmetric distributions, Fang (1990)[4] found that ellipsoidal distributions are also closely related to matrix T distributions, see §3.5.3 in Fang (1990)[4]. Whether it is a one-dimensional t distribution, a multivariate t distribution, or a matrix T distribution, the density function form is much complicated than those of the normal distributions, and it is inconvenient to use. It has been found that the density function of the one-dimensional t distribution approaches the one-dimensional normal distribution with unlimited increasing degrees of freedom, see Kotz (2004)[1] page 2, and this convergence is essentially an in distribution convergence. For a multivariate t distribution, a matrix T distribution, one naturally asks whether it is converg to a normal distribution. Is it uniformly convergent? What is its rate of convergence? In order to calculate probability integrals of multivariate t distributions, a series of studies have been conducted by Fujikoshi (1989, 1997)[6, 7] and others. From the perspective of the Kullback-Leibler distance. Akimoto (1994)[2] gave a results for the consistent convergence of the Wishart distribution to a normal distribution, these are the motivation for this paper to study the consistent asymptotic normality of the matrix T distribution, although so far we have found no results of this study.

2020 *Mathematics Subject Classification*. Primary 62E17,62E20
Keywords. matrix T distribution, matrix Γ distribution, Berry-Esseen boundary, uniformly asymptotic normality
Received: 10 January 2021; Revised: 23 April 2021; Accepted: 08 November 2021
Communicated by Dragana Cvetković-Ilić
Corresponding author: Di Shou Mao
The research is partially supported National Science Foundation of China (No. 10771175)
Email address: likc@hbnu.edu.cn (Di Shou Mao)

In this paper, after describing the relationship between the matrix Γ distribution and the matrix T distribution in the section 2, we give the definition that a matrix T distribution is uniformly asymptotic normality in the section 3. In the section 4, we intend to discuss the Berry-Esseen inequality of the matrix T distribution, and thus give conclusions about consistent asymptotic normality of the matrix T distribution and its convergence rate.

2. Overview of the matrix T distribution

In this paper, let $A = (a_{ij})$ be a matrix of $m \times m$ order. $A > 0$ means that A is a positive definite matrix, the determinant of A is represented by$|A|$. when $A > 0$, $A^{\frac{1}{2}}$ was defined by Muirhead (1982), page 588[9]. If $A = (a_{ij})_{n\times m}$ is a random matrix, the joint distribution of the $nm \times 1$ random vector

$$\overrightarrow{A} = (a_{11}, a_{21}, ..., a_{n1}, a_{12}, a_{22}, ..., a_{n2}, a_{13}, a_{23}, ..., a_{n3}, ..., a_{1m}, ..., a_{nm})^{\top}$$

is called the distribution of the random matrix $A = (a_{ij})$. If $A = (a_{ij})$ is a symmetric matrix with $m \times m$ order, the joint distribution of the $\frac{m(m+1)}{2} \times 1$ dimension random vector

$$\widetilde{A} = (a_{11}, a_{12}, ..., a_{1m}, a_{22}, a_{23}, ..., a_{2m}, a_{33}, ..., a_{mm})^{\top}$$

is regarded as the distribution of the random matrix $A = (a_{ij})$.Where, $\overrightarrow{A}$ represents the straightening of the whole matrix, and $\widetilde{A}$ is the straightening of the main diagonal of the symmetric matrix and the triangular elements below the main diagonal.

In mathematical statistics, if x and y are mutually independent random variables, and $x \sim N(0, 1)$, $y \sim \chi^2_{(n)}$, then

$$t = \frac{\sqrt{n}x}{\sqrt{y}} \sim t_{(n)}.$$

That is, t obeys the central t distribution whose degree of freedom is n. According to this way, a matrix T distribution can be defined similarly. Therefore, we need the concepts of matrix normal distribution and matrix χ^2 distribution. It is well known that X has matrix normal distribution, denoted by $X \sim N_{n\times m}(M, V, \Sigma)$, if the density function of $X = (x_{ij})_{n\times m}$ is

$$f(X) = (2\pi)^{-\frac{nm}{2}}|V|^{-\frac{n}{2}}|\Sigma|^{-\frac{m}{2}}\exp\{-\frac{1}{2}tr(X - M)^{\top}\Sigma^{-1}(X - M)V^{-1}\}. \tag{0}$$

On the other hands, following the Definition 2.1.10 of Muirhead (1982)[9], a matrix Γ distribution is a natural generalization for χ^2 distribution.

Definition 2.1. *If a $m \times m$ symmetric random matrix X has a density function:*

$$g(X|\alpha, \Sigma) = \begin{cases} \frac{|X|^{\alpha-\frac{m+1}{2}}}{\Gamma_m(\alpha)|\Sigma|^{\alpha}} exp\{-tr(\Sigma^{-1}X)\}, & X > 0; \\ 0, & other, \end{cases} \tag{1}$$

where $\Sigma = (\sigma_{ij}) > 0$, $\alpha > \frac{m-1}{2}$, $\Gamma_m(\alpha)$ is a multivariate Γ function, i.e. $\Gamma_m(\alpha) = \pi^{m(m-1)/4}\Pi_{i=1}^{m}\Gamma(\alpha - (i - 1)/2)$, Then it is said that X obeys the matrix Γ distribution, denoted as $X \sim Ga_m(\alpha, \Sigma)$.

In Definition 2.1, let $m = 1$, $\alpha = \frac{n}{2}$, $\Sigma = 2$, equation (1) is the density function of the central χ^2 distribution with degree of freedom n . On the other hand, the central Wishart distribution is also a special case of the matrix Γ distribution, see the Theorem 3.2.1 in Muirhead (1982) page 85[9].

Definition 2.2. *Suppose the random matrix $X \sim N_{n\times m}(O, I_m, I_n)$, where O is the $n \times m$ order 0 matrix , I_m and I_n are unit matrices. $W \sim Ga_m(\frac{\gamma}{2}, 2I_m)$, $\gamma > 0$. X and W are independent of each other, the distribution of the random matrix $T = \sqrt{\gamma}XW^{-\frac{1}{2}}$ is said to be a central matrix T distribution with degrees of freedom γ, denoted as $T \sim MT(m, n, \gamma)$.*

In Definition 2.2, if $n, m = 1$, the matrix T distribution simplifies to a one-dimensional center t distribution with the degree of freedom γ. Since the degrees of freedom γ of t distribution derives from the capacity of a sampling, when discussing the limiting distribution of the matrix T distribution, In general, it is reasonable to treat the limit process as $\gamma \to \infty$.

Lemma 2.3. *Let $T = \sqrt{\gamma}XW^{-\frac{1}{2}}$ be the random matrix given by definition 2.2, then the density function of the center matrix T distribution $MT(m, n, \gamma)$ is*

$$g(T) = (\pi\gamma)^{-\frac{nm}{2}} \frac{\Gamma_m(\frac{\gamma+n}{2})}{\Gamma_m(\frac{\gamma}{2})} |I_m + \frac{1}{\gamma}T^\top T|^{-\frac{n+\gamma}{2}}. \tag{2}$$

Proof. Notice that X, W are independent, following the typical method to derive the density function of a random variate, for each nonnegative Borel measurable function $g(\cdot)$,

$$Eg(T) = \int_{W>0,X} g(\sqrt{\gamma}XW^{-\frac{1}{2}})(2\pi)^{-\frac{nm}{2}} \frac{1}{\Gamma_m(\frac{\gamma}{2})2^{\frac{m\gamma}{2}}} |W|^{\frac{\gamma}{2}-\frac{m+1}{2}} \exp\{-\frac{1}{2}tr(W + X^\top X)\}dXdW.$$

Set $T = \sqrt{\gamma}XW^{-\frac{1}{2}}, S = W$, then $X = \frac{1}{\sqrt{\gamma}}TS^{\frac{1}{2}}, W = S$, Using Theorem 2.1.5 in Muirhead (1982)[9], the transformed Jacobian determinant is

$$J((X, W) \to (T, S)) = \gamma^{-\frac{nm}{2}} |S|^{\frac{n}{2}},$$

hence

$$Eg(T) = \int_T [g(T)(2\pi)^{-\frac{nm}{2}} \gamma^{-\frac{nm}{2}} \frac{1}{\Gamma_m(\frac{\gamma}{2})2^{\frac{m\gamma}{2}}} \int_{S>0} |S|^{\frac{\gamma+n}{2}-\frac{m+1}{2}} \exp\{-\frac{1}{2}trS^{\frac{1}{2}}(I + \frac{1}{\gamma}T^\top T)S^{\frac{1}{2}}\}dS]dT.$$

Using equation (1) to the inner integration, we get

$$Eg(T) = \int_T g(T)\{(\pi\gamma)^{-\frac{nm}{2}} \frac{\Gamma_m(\frac{\gamma+n}{2})}{\Gamma_m(\frac{\gamma}{2})} |I_m + \frac{1}{\gamma}T^\top T|^{-\frac{n+\gamma}{2}}\}dT.$$

the density function of the central matrix T distribution is exactly equation (2). □

Lemma 2.4. *Let $X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim SS(\varphi)$, i.e., X obeys a spherically symmetric distribution whose eigenfunctions are $\varphi(\cdot)$, if X_i is a random array of $n_i \times m$, $i = 1, 2$. when $n_i \geq m$, the random array $T = \sqrt{n_2}X_1(X_2^\top X_2)^{-\frac{1}{2}} \sim MT(m, n_1, n_2)$.*

A detailed proof of lemma 2.4 is given in Theorem 3.5.4 of Fang (1990)[4] page 113. This conclusion shows that it is useful to study the matrix t distribution for applications of spherically symmetric distributions. this is not the purpose of this paper, it will not be deeply studied . In the definition of one-dimensional t distribution, when the normal distribution is not the standard normal, the concept of a non-central t distribution also arises, so it is necessary to extend the central matrix T to the general matrix T distribution.

Definition 2.5. *the random matrix $X = (x_{ij})_{n\times m}$ is called to obey the matrix T distribution, noted as $X \sim T_{n\times m}(M, B^{-1}, A, \gamma)$, If the density function of X is*

$$p(X) = (\pi)^{-\frac{nm}{2}} \frac{\Gamma_m(\frac{\gamma+n+m-1}{2})}{\Gamma_m(\frac{\gamma+m-1}{2})} |A|^{-\frac{n}{2}} |B|^{\frac{m}{2}} |I_m \quad + A^{-1}(X - M)^\top B(X - M)|^{-\frac{\gamma+n+m-1}{2}}. \tag{3}$$

where, the constant $\gamma > 0$, M is a constant matrix, and A, B are two positive definite matrices.

Remark 2.6. 1^o: *central matrix $MT(m,n,\gamma)$ is a special case of the matrix $T_{n\times m}(M,B^{-1},A,\gamma)$.*
2^o: *If $T\sim MT(m,n,\gamma)$, then $Y=M+\gamma^{-\frac{1}{2}}B^{-\frac{1}{2}}TA^{\frac{1}{2}}\sim T_{n\times m}(M,B^{-1},A,\gamma-m+1)$. where, M is a constant matrix , A,B are positive constant matrices.*

Indeed, 1^o is obvious from the density function that $T_{n\times m}(O,I_n,\gamma I_m,\gamma-m+1)$ is $MT(m,n,\gamma)$. The proof of the conclusion 2^o can be derived by the following Lemma 2.7. Remark 2.6 specifies the relationship between the matrix T distribution and the central matrix T, and it is clear that the properties of the central matrix T distribution can be obtained as a special case deduction as long as the properties of the matrix T distribution are studied.

Lemma 2.7. *Set $X\sim T_{n\times m}(M,B^{-1},A,\gamma)$, and $P_{n\times n}$, $Q_{m\times m}$ are non-singular constant matrices, then*

$$Z=PXQ\sim T_{n\times m}(PMQ,PB^{-1}P^\top,Q^\top AQ,\gamma).$$

Proof. The Jacobian determinant of that the transformation $Z\to X$ is $|J(Z\to X)|=|P|^m|Q|^n$, therefor, the density function of Z

$$g(Z)=(\pi)^{-\frac{nm}{2}}\frac{\Gamma_m(\frac{\gamma+n+m-1}{2})}{\Gamma_m(\frac{\gamma+m-1}{2})}|B|^{\frac{m}{2}}|A^{-1}|^{\frac{n}{2}}$$
$$\times|I_m+A^{-1}(P^{-1}ZQ^{-1}-M)^\top B(P^{-1}ZQ^{-1}-M)|^{-\frac{\gamma+n-m-1}{2}}|P^{-1}|^m|Q^{-1}|^n$$
$$=(\pi)^{-\frac{nm}{2}}\frac{\Gamma_m(\frac{\gamma+n+m-1}{2})}{\Gamma_m(\frac{\gamma+m-1}{2})}|(P^{-1})^\top BP^{-1}|^{\frac{m}{2}}|(Q^\top AQ)^{-1}|^{\frac{n}{2}}$$
$$\times|I_m+A^{-1}(Q^\top)^{-1}(Z-PMQ)^\top(P^{-1})^\top BP^{-1}(Z-PMQ)Q^{-1}|^{-\frac{\gamma+n-m-1}{2}}$$
$$=(\pi)^{-\frac{nm}{2}}\frac{\Gamma_m(\frac{\gamma+n+m-1}{2})}{\Gamma_m(\frac{\gamma+m-1}{2})}|(PB^{-1}P^\top)^{-1}|^{\frac{m}{2}}|(Q^\top AQ)^{-1}|^{\frac{n}{2}}$$
$$\times|I_m+(Q^\top AQ)^{-1}(Z-PMQ)^\top(PB^{-1}P^\top)^{-1}(Z-PMQ)|^{-\frac{\gamma+n-m-1}{2}}.$$

By definition, $Z=PXQ\sim T_{n\times m}(PMQ,PB^{-1}P^\top,Q^\top AQ,\gamma)$. □

3. A conception for uniformly asymptotic normality and related results

A statement for uniformly asymptotic normality and some related lemmas are given in this section. In the n-dimensional Euclidean space (R^n,B^n), let X,Y be a random vector of $n\times 1$ dimensions. P^X, P^Y is the probability distribution of X,Y, $f(x)$, $g(y)$ is the density of the distribution of X,Y, respectively, noting that $P^X(E)=\int_E dP^X,\quad E\in B^n$.Let

$$D(X,Y)=\sup_{E\in B^n}|P^X(E)-P^Y(E)|,\quad I(X,Y)=E_X[\ \ln\frac{f(X)}{g(X)}].$$

Obviously, $D(X,Y)$ and $I(X,Y)$ is the full variational distance and Kullback-Leibler distance between X and Y respectively. The full variational distance and Kullback-Leibler distance are widely used in mathematical statistics, information theory, and many other fields. There is an important conclusion between them.

Lemma 3.1. *If X,Y are random vectors with the densities $f(x)$ and $g(y)$, respectively, then the relational formula between the full-variance distance and Kullback-Leibler distance of X,Y is*

$$D(X,Y)\le\sqrt{\frac{I(X,Y)}{2}}.$$

For proof, see the Proposition 4.3.7 in Whittaker (1990)[11].

Definition 3.2. *Let $X \sim T_{n\times m}(M, B^{-1}, A, \gamma)$, $Y \sim N_{n\times m}(O, I_m, I_n)$, if there are nonsingular constant matrices $P_{n\times n}$, $Q_{m\times m}$, when $Z = P(X - M)Q$ so that the following inequality holds*

$$D(Z, Y) = \sup_{E\in B^{nm}} |P^Z(E) - P^Y(E)| \le HL_\gamma,$$

where, constant H is unrelated to γ but L_γ is related to γ, we call that the HL_γ is a Berry-Esseen boundary of the matrix T distribution. Further,when $\gamma \to \infty$, $D(Z, Y) \to 0$, we say that the matrix T distribution $T_{n\times m}(M, B^{-1}, A, \gamma)$ is uniformly asymptotic to a matrix normal distribution.

If $n = m = 1$, then the inequality in definition 3.2 is actually a Berry-Esseen inequality, which based on the full variation the distance between the standardized T distribution and the standardized normal distribution. Obviously, the Berry-Essen boundary of T distribution is not unique. The following results are also needed for future purposes.

Lemma 3.3. 1^o : *$ln\Gamma(x) = \frac{1}{2}\, ln 2\pi + (x - \frac{1}{2})\, ln x - x - R(x)$,*
where : $\quad 0 < R(x) < \frac{1}{64x^2(x+1)}, \quad x > 0.$

2^o:

$$ln\Gamma(\frac{\gamma}{2}) + \frac{1}{2}\, ln\frac{\gamma-2}{2} - ln\Gamma(\frac{\gamma+1}{2}) < \frac{1}{2\gamma}, \quad \gamma > 2.$$

3^o : *Set*

$$c(m, n, \gamma) = \frac{\Gamma_m(\frac{\gamma+m-1}{2})(\frac{\gamma-2}{2})^{\frac{nm}{2}}}{\Gamma_m(\frac{\gamma+n+m-1}{2})}, \quad \textit{then } lnc(m, n, \gamma) < \frac{nm}{2\gamma}.$$

Proof. A proof of 1^o to see Matsunawa (1976)[10].
The following to proof 2^o, when $\gamma > 2$, using 1^o, there is

$$\ln\Gamma(\frac{\gamma}{2}) + \frac{1}{2}\ln\frac{\gamma-2}{2} - \ln\Gamma(\frac{\gamma+1}{2}) =$$

$$\frac{\gamma-1}{2}\ln\frac{\gamma}{2} - \frac{\gamma}{2} - R(\frac{\gamma}{2}) + \frac{1}{2}\ln\frac{\gamma-2}{2} - \frac{\gamma}{2}\ln\frac{\gamma+1}{2} + \frac{\gamma+1}{2} + R(\frac{\gamma+1}{2}) =$$

$$\frac{\gamma}{2}(\ln\frac{\gamma}{2} - \ln\frac{\gamma+1}{2}) + \frac{1}{2}(\ln\frac{\gamma-2}{2} - \ln\frac{\gamma}{2}) + \frac{1}{2} + R(\frac{\gamma+1}{2}) - R(\frac{\gamma}{2}) =$$

$$\frac{\gamma}{2}\ln(1 - \frac{1}{\gamma+1}) + \frac{1}{2}\ln(\frac{\gamma-2}{\gamma}) + \frac{1}{2} + R(\frac{\gamma+1}{2}) - R(\frac{\gamma}{2}),$$

since $x > -1$, $\ln(1 + x) \le x$, and $R(x)$ are bounded,

$$\ln\Gamma(\frac{\gamma}{2}) + \frac{1}{2}\ln\frac{\gamma-2}{2} - \ln\Gamma(\frac{\gamma+1}{2}) < \frac{\gamma}{2}\ln(1 - \frac{1}{\gamma+1}) + \frac{1}{2} + R(\frac{\gamma+1}{2})$$

$$< \frac{\gamma}{2}(-\frac{1}{\gamma+1}) + \frac{1}{2} + \frac{1}{64(\frac{\gamma+1}{2})^2(\frac{\gamma+3}{2})} < \frac{1}{2(\gamma+1)} + \frac{1}{8(\gamma+1)^2(\gamma+3)} < \frac{1}{2\gamma}.$$

3^o By the expression $\Gamma_m(\cdot)$ in definition 2.1,

$$\ln c(m, n, \gamma) = \ln\Gamma_m(\frac{\gamma+m-1}{2}) + \frac{nm}{2}\ln(\frac{\gamma-2}{2}) - \ln\Gamma_m(\frac{\gamma+n+m-1}{2})$$

$$= \ln[\Pi_{i=1}^m \Gamma(\frac{\gamma+i-1}{2})] - \ln[\Pi_{i=1}^m \Gamma(\frac{\gamma+n+i-1}{2})] + \frac{nm}{2}\ln(\frac{\gamma-2}{2})$$

$$= \sum_{i=1}^{m}[\ \ln\Gamma(\frac{\gamma+i-1}{2}) - \ \ln\Gamma(\frac{\gamma+n+i-1}{2})] + \frac{nm}{2}\ln(\frac{\gamma-2}{2})$$

$$= \sum_{i=1}^{m}[\ \ln\Gamma(\frac{\gamma+i-1}{2}) + \frac{n}{2}\ln(\frac{\gamma-2}{2}) - \ \ln\Gamma(\frac{\gamma+n+i-1}{2})]$$

$$= \sum_{i=1}^{m}\{\sum_{j=1}^{n}[\ \ln\Gamma(\frac{\gamma+i+j-2}{2}) + \frac{1}{2}\ln(\frac{\gamma-2}{2}) - \ \ln\Gamma(\frac{\gamma+j+i-1}{2})]\},$$

according to the result of 2^o,

$$\ln c(m,n,\gamma) < \sum_{i=1}^{m}\{\sum_{j=1}^{n}[\frac{1}{2(\gamma+i+j-2)}] < \frac{nm}{2\gamma}.$$

□

Lemma 3.4. *Sppose the $f(x)$ to be a non-negative Borel measurable function on R^1, then*

$$\int_{R^m} f(x_1^2 + x_2^2 + \cdots + x_m^2)dx_1dx_2\cdots dx_m = \frac{\pi^{\frac{m}{2}}}{\Gamma(\frac{m}{2})}\int_0^{+\infty} y^{\frac{m}{2}-1}f(y)dy.$$

Detailed proof to see example 2.4.1 in Fang (1990)[4] page 52.

4. A Berry-Esseen boundary for matrix T distributions and its applications

In this section we shall give a Berry-Esseen bounds for the matrix T distribution. As an application, we give the proof that a matrix T distribution is uniformly asymptotic a normal distribution , also give a convergence rate.

Theorem 4.1. *If $T \sim T_{n\times m}(M, B^{-1}, A, \gamma)$, $S \sim N_{n\times m}(M, A, [(\gamma-2)B]^{-1})$, then when $\gamma > 2$.*

$$I(S,T) < \frac{mn(m+n+2)}{2(\gamma-2)}.$$

Proof. Assume $f(X)$ and $g(X)$ are the densities of the random matrices T, S respectively, then

$$f(X) = (\pi)^{-\frac{nm}{2}}\frac{\Gamma_m(\frac{\gamma+n+m-1}{2})}{\Gamma_m(\frac{\gamma+m-1}{2})}|A|^{-\frac{n}{2}}|B|^{\frac{m}{2}}|I_m + A^{-1}(X-M)^\top B(X-M)|^{-\frac{\gamma+n+m-1}{2}},$$

$$g(X) = (2\pi)^{-\frac{nm}{2}}|A|^{-\frac{n}{2}}|(\gamma-2)B|^{\frac{m}{2}}\exp\{-\frac{\gamma-2}{2}\ \mathrm{tr}[(X-M)^\top B(X-M)A^{-1}]\}$$
$$= (\frac{2}{\gamma-2}\pi)^{-\frac{nm}{2}}|A|^{-\frac{n}{2}}|B|^{\frac{m}{2}}\exp\{-\frac{\gamma-2}{2}\ \mathrm{tr}[(X-M)^\top B(X-M)A^{-1}]\}.$$

Note that

$$c(m,n,\gamma) = \frac{\Gamma_m(\frac{\gamma+m-1}{2})(\frac{\gamma-2}{2})^{\frac{nm}{2}}}{\Gamma_m(\frac{\gamma+n+m-1}{2})},$$

according to the definition of Kullback-Leibler distance, we have

$$
\begin{aligned}
I(S,T) = \int_{R^{nm}} \ln \frac{g(X)}{f(X)} g(X)dX = \ln c(m,n,\gamma) + \\
\int_{R^{nm}} \Big(\ln|I_m + A^{-1}(X-M)^\top B(X-M)|^{\frac{\gamma+n+m-1}{2}} \\
-\frac{\gamma-2}{2} \operatorname{tr}[(X-M)^\top B(X-M)A^{-1}]\Big) g(X)dX \\
= \ln c(m,n,\gamma) + \int_{R^{nm}} \Big(\ln|I_m + A^{-1}(X-M)^\top B(X-M)|^{\frac{\gamma+n+m-1}{2}} \\
-\frac{\gamma-2}{2} \operatorname{tr}[(X-M)^\top B(X-M)A^{-1}]\Big) \\
\times (\frac{2}{\gamma-2}\pi)^{-\frac{nm}{2}} |A|^{-\frac{n}{2}} |B|^{\frac{m}{2}} \\
\times \exp\{-\frac{\gamma-2}{2} \operatorname{tr}[(X-M)^\top B(X-M)A^{-1}]\} dX. \qquad (4)
\end{aligned}
$$

Taking an integral transformation in formula (4), $Y = B^{\frac{1}{2}}(X-M)A^{-\frac{1}{2}}$, using theorem 2.1.5 in Muirhead (1982)[9], the Jacobian determinant of the transformation is

$$J(X \to Y) = |A|^{\frac{n}{2}} |B|^{-\frac{m}{2}},$$

also note that in formula (4)

$$|I_m + A^{-1}(X-M)^\top B(X-M)|^{\frac{\gamma+n+m-1}{2}} = |(I_m + Y^\top Y)|^{\frac{\gamma+n+m-1}{2}}.$$

Using the matrix singular value decomposition theorem, there exist n order and m order orthogonal matrices P, Q such that

$$Y = P \begin{bmatrix} \Lambda & 0 \\ 0 & 0 \end{bmatrix} Q,$$

where $\Lambda = diag(\lambda_1, \lambda_2, \cdots, \lambda_r)$,$\lambda_i > 0, i = 1, 2, ...r$. $\lambda_1^2, \lambda_2^2, \cdots, \lambda_r^2$ are non-zero eigenvalues of $Y^\top Y$. Thus

$$
\begin{aligned}
|(I_m + Y^\top Y)|^{\frac{\gamma+n+m-1}{2}} &= |Q^\top (I_m + \begin{bmatrix} \Lambda^2 & 0 \\ 0 & 0 \end{bmatrix}) Q|^{\frac{\gamma+n+m-1}{2}} \\
&= \Big[(1+\lambda_1^2)(1+\lambda_2^2)\cdots(1+\lambda_r^2)\Big]^{\frac{\gamma+n+m-1}{2}}.
\end{aligned}
$$

and because $\ln(1+\lambda_i^2) \le \lambda_i^2, i = 1, 2, ...r.$ so that

$$\ln|(I_m + Y^\top Y)|^{\frac{\gamma+n+m-1}{2}} = \frac{\gamma+n+m-1}{2}[\sum_{i=1}^{r} \ln(1+\lambda_i^2)]$$

$$\le \frac{\gamma+n+m-1}{2}[\lambda_1^2 + \lambda_2^2 + \cdots + \lambda_r^2] = \frac{\gamma+n+m-1}{2} \operatorname{tr}(Y^\top Y).$$

On the other hand, in equation (4),

$$\operatorname{tr}[(X-M)^\top B(X-M)A^{-1}] = \operatorname{tr}[A^{-\frac{1}{2}}(X-M)^\top B(X-M)A^{-\frac{1}{2}}] = \operatorname{tr}(Y^\top Y).$$

Substituting both the above equation and inequality into the equation (4) yields

$$I(S,T) = \ln c(m,n,\gamma) + \int_{R^{nm}} \Big(\frac{\gamma+n+m-1}{2} \ln|I_m + Y^\top Y| - \frac{\gamma-2}{2} \operatorname{tr}(Y^\top Y)\Big)$$

$$\times(\frac{2}{\gamma-2}\pi)^{-\frac{nm}{2}}|A|^{-\frac{n}{2}}|B|^{\frac{m}{2}}\exp\{-\frac{\gamma-2}{2}\operatorname{tr}(Y^\top Y)\}|A|^{\frac{n}{2}}|B|^{-\frac{m}{2}}dY$$
$$\leq \ln c(m,n,\gamma)+\int_{R^{nm}}\left(\frac{\gamma+n+m-1}{2}\operatorname{tr}(Y^\top Y)-\frac{\gamma-2}{2}\operatorname{tr}(Y^\top Y)\right)$$
$$\times(\frac{2}{\gamma-2}\pi)^{-\frac{nm}{2}}\exp\{-\frac{\gamma-2}{2}\operatorname{tr}(Y^\top Y)\}dY.$$

So,

$$I(S,T)=\ln c(m,n,\gamma)+\frac{n+m+1}{2}(\frac{\gamma-2}{2\pi})^{\frac{nm}{2}}\int_{R^{nm}}\operatorname{tr}(Y^\top Y)\exp\{-\frac{\gamma-2}{2}\operatorname{tr}(Y^\top Y)\}dY. \tag{5}$$

Remember $Y=(y_{ij})_{n\times m}$ in equation (5), then $Y^\top Y=(\sum_{k=1}^n y_{ki}y_{kj})_{m\times m}$, therefore

$$\operatorname{tr}(Y^\top Y)=\sum_{i=1}^m\sum_{k=1}^n y_{ki}y_{ki}=\sum_{i=1}^m\sum_{k=1}^n y_{ki}^2.$$

By lemma 3.4

$$\int_{R^{nm}}\operatorname{tr}(Y^\top Y)\exp\{-\frac{\gamma-2}{2}\operatorname{tr}(Y^\top Y)\}dY$$
$$=\frac{\pi^{\frac{mn}{2}}}{\Gamma(\frac{mn}{2})}\int_0^{+\infty}y^{\frac{mn}{2}-1}\exp\{-\frac{\gamma-2}{2}y\}dy,$$

putting $t=\frac{\gamma-2}{2}y$, then

$$\frac{\pi^{\frac{mn}{2}}}{\Gamma(\frac{mn}{2})}\int_0^{+\infty}y^{\frac{mn}{2}-1}\exp\{-\frac{\gamma-2}{2}y\}dy$$
$$=\frac{\pi^{\frac{mn}{2}}}{\Gamma(\frac{mn}{2})}(\frac{2}{\gamma-2})^{\frac{mn}{2}+1}\int_0^{+\infty}t^{\frac{mn}{2}}\exp\{-t\}dt$$
$$=\frac{\pi^{\frac{mn}{2}}}{\Gamma(\frac{mn}{2})}(\frac{2}{\gamma-2})^{\frac{mn}{2}+1}\Gamma(\frac{mn}{2}+1)$$
$$=\frac{mn}{2}(\frac{2}{\gamma-2})^{\frac{mn}{2}+1}\pi^{\frac{mn}{2}},$$

therefor

$$\int_{R^{nm}}\operatorname{tr}(Y^\top Y)\exp\{-\frac{\gamma-2}{2}\operatorname{tr}(Y^\top Y)\}dY=\frac{mn}{2}(\frac{2}{\gamma-2})^{\frac{mn}{2}+1}\pi^{\frac{mn}{2}}. \tag{6}$$

Taking the conclusion 3^o from lemma 3.3 and equation (6) into equation (5) yields

$$I(S,T)\leq\frac{nm}{2\gamma}+\frac{n+m+1}{2}(\frac{\gamma-2}{2\pi})^{\frac{nm}{2}}\frac{mn}{2}(\frac{2}{\gamma-2})^{\frac{mn}{2}+1}\pi^{\frac{mn}{2}}$$
$$=\frac{nm}{2\gamma}+\frac{(n+m+1)nm}{2(\gamma-2)}$$
$$<\frac{(n+m+2)nm}{2(\gamma-2)}.$$

□

Theorem 4.2. *Set $T\sim T_{n\times m}(M,B^{-1},A,\gamma)$, then when $\gamma>2$.*

$$L=\frac{\sqrt{mn(m+n+2)}}{2}\sqrt{\frac{1}{\gamma-2}}$$

is a Berry-Esseen boundary of the matrix distribution T.

Proof. Without loss of generality, it can be assumed that $T \sim T_{n\times m}(O, B^{-1}, A, \gamma)$, since when $M \neq O$, one make $T_1 = T - M$, and take $n \times n$ order non-singular matrix $P = \sqrt{\gamma - 2}B^{\frac{1}{2}}$ and $m \times m$ order nonsingular matrix $Q = A^{-\frac{1}{2}}$, then by Lemma 2.7 $Z = PTQ \sim T_{n\times m}(O, (\gamma - 2)I_n, I_m, \gamma)$, pair matrix normal distribution $S \sim N_{n\times m}(O, I_m, I_n)$, by Theorem 4.1,

$$I(S, Z) < \frac{mn(m + n + 2)}{2(\gamma - 2)},$$

$$D(Z, S) = D(S, Z) \leq \sqrt{\frac{I(S, Z)}{2}} \leq \frac{\sqrt{mn(m + n + 2)}}{2}\sqrt{\frac{1}{\gamma - 2}} = L.$$

By Definition 3.2, L must be a Berry-Esseen boundary of the distributed $T_{n\times m}(M, B^{-1}, A, \gamma)$. □

Corollary 4.3. *When $\gamma \to \infty$, the matrix T distributes $T_{n\times m}(M, B^{-1}, A, \gamma)$, is uniformly asymptotic to the matrix normal distribution $N_{n\times m}(M, A, [(\gamma - 2)B]^{-1})$, whose rate of convergence is $O(\gamma^{-\frac{1}{2}})$.*

The proof is the direct result of Theorem 4.2.

Corollary 4.4. *The central matrix T distribution $MT_{n\times m}(m, n, \gamma)$ is uniformly asymptotic to the normal distribution $N_{n\times m}(O, \gamma I_m, \frac{1}{\gamma+m-1}I_n)$ when $\gamma \to \infty$, its convergence rate is equal to $O(\gamma^{-\frac{1}{2}})$.*

The proof of corollary 4.4 holds using corollary 4.3 and the preceding conclusion 1^0 in Remark 2.6.

From the conclusion that a matrix T distribution is uniformly asymptotically the matrix normal distribution, we can be sure that a one-dimensional t distribution or a multivariate t distribution is also uniformly asymptotically normal distribution.

In fact, it makes sense to study uniformly asymptotic normal distribution for the various forms of t distributions, which are described in Chapter 4 and Chapter 5 of Kotz (2004)[1], we will do not to discuss this topic here.

References

[1] S. Kotz, S. Nadarajah, Multivariate T-Distributions and Their Applications. Cambridge, Cambridge University Press,2004.
[2] Y.Akimoto, The Uniform Asymptotic Normality of the Wishart Distribution based on the K-L Information.Journal Janpan statist,24(1994)73-81.
[3] M.Roth , E. Ozkan ,F. Gustafsson , A Student's T Filter for Heavy Tailed process and Measurement Noise, IEEE International Conference on Acoustics, Speech and Signal Processing, (2013)5770–5774.
[4] K.T.Fang, and Y.T.Zhang, Generalized Multivariate Analysis. New York: Springer-Verlag,1990.
[5] J. M. Dickey, Matric-variate generalizations of the multivariate distribution, Annals of Mathematical Statistics, 38 (1967)511–518.
[6] Y. Fujikoshi, Error bounds for asymptotic expansions of the maximums of the multivariate t- and F-variables with common denominator,Hiroshima Mathematics Journal,19 (1989) 319–327.
[7] Y.Fujikoshi, An asymptotic expansion for the distribution of Hotelling's T2-statistic under nonnormality,Journal of Multivariate Analysis,61(1997) 187-193.
[8] Z.Jiang,P.Ding, Identification of Causal Effects Within Principal Strata Using Auxiliary Variables, Statistical Science(to appear) 2020.
[9] R. J. Muirhead, Aspects of multivariate statistical theory,New York: Wiley,1982.
[10] T. Matsunawa , Some inequalities based on factorial series, Ann. Inst.Statist. Math, 28(1976)291-305.
[11] J.Whittaker , Graphical Models in Applied Multivariate Statistics, Wiley, Chichester,1990.