OUKA
The University of Osaka
Institutional Knowledge Archive

Title	On Uniform Asymptotic Normality of Probability Distributions
Author(s)	秋本, 義久
Citation	大阪大学, 1996, 博士論文
Version Type	VoR
URL	https://doi.org/10.11501/3119657
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

The University of Osaka

On Uniform Asymptotic Normality of Probability Distributions

Yoshihisa Akimoto
Faculty of Econoinformatics,
Himeji Dokkyo University, Himeji 670, Japan

1996

Abstract

Theory of asymptotic equivalence of probability distributions has been developed by Ikeda. Among others uniform asymptotic equivalence is of interest. The usual asymptotic distribution theory is based on the notion of the so called in law convergence, while the notion of uniform asymptotic equivalence is defined by taking the absolute error evaluation into account. Regarding normal approximation, uniform approximation is the most important. On the other hand, Matsunawa has given a modified Stirling formula which sharpens the well-known Stirling asymptotic formula for natural numbers by presenting a double inequality. This result is useful in the uniform asymptotic approximation theory. However, so far as the author knows, applying the uniform normal approximation theory to concrete multivariate distributions has not been done, where the uniform asymptotic normality is a strictly stronger notion than the usual one which is based on the convergence in law.

In the present thesis, we consider two problems. One is the implication relation between two types of asymptotic equivalence. The other is the uniform normal approximation to multivariate distributions by using the modified Stirling formula.

In Chapter 2, we shall give a sufficient condition under which two types of asymptotic equivalence are mutually equivalent in one-dimensional real case and also some useful formulas for the numerical evaluation of the related quantities.

In Chapter 3, we shall prove the uniform asymptotic normality of the Wishart distribution $W_p(n, \Lambda)$ under the condition $p^3/n \to 0$, by giving an upper bound of the uniform error based on the Kullback-Leibler information. The condition $p^3/n \to 0$ is the best possible for which the information converges to zero.

In the last Chapter, we shall also prove the uniform asymptotic normality of the Dirichlet distribution under certain limiting process of related parameters. This result can be obtained by giving an upper bound of the uniform error based on the Kullback-Leibler information. These results in Chapters 3 and 4 are effective for exact sample theory because the upper bounds of the uniform errors are evaluated by certain inequality, respectively.

Contents

Chapter 1

Introduction

Let $\{\boldsymbol{X}_n; n \geq 1\}$ and $\{\boldsymbol{Y}_n; n \geq 1\}$ be two sequences of k-dimensional real random variables. It is assumed that $\boldsymbol{X}_n$ and $\boldsymbol{Y}_n$ are all absolutely continuous with respect to the Lebesgue measure μ over the measurable space $(\boldsymbol{R}_{(k)}, \boldsymbol{\mathcal{B}}_{(k)})$, $\boldsymbol{R}_{(k)}$ being the k-dimensional Euclidean space and $\boldsymbol{\mathcal{B}}_{(k)}$ the usual Borel field of subsets of $\boldsymbol{R}_{(k)}$. Let us denote the probability density functions of $\boldsymbol{X}_n$ and $\boldsymbol{Y}_n$ by $f_n(\boldsymbol{x})$ and $g_n(\boldsymbol{x})$, respectively. Further let $D(\boldsymbol{X}_n)$ and $D(\boldsymbol{Y}_n)$ denote the carriers of $f_n(\boldsymbol{x})$ and $g_n(\boldsymbol{x})$.

As is well-know, the Kullback-Leibler information for discrimination is given by

$$I(\boldsymbol{X}_n : \boldsymbol{Y}_n) = \int_{\boldsymbol{R}_{(k)}} f_n(\boldsymbol{x}) \log \frac{f_n(\boldsymbol{x})}{g_n(\boldsymbol{x})} d\mu(\boldsymbol{x}),$$

and we have $I(\boldsymbol{X}_n : \boldsymbol{Y}_n) \geq 0$, where equality holds if and only if $f_n(\boldsymbol{x}) = g_n(\boldsymbol{x})$ (a.e. μ) on $\boldsymbol{R}_{(k)}$. If $\mu(D(\boldsymbol{X}_n) - D(\boldsymbol{Y}_n)) \geq 0$, then $I(\boldsymbol{X}_n : \boldsymbol{Y}_n) = \infty$ ([9],[10],[19]).

Let $\boldsymbol{\mathcal{C}}$ be any given class of subsets of $\boldsymbol{R}_{(k)}$ belonging to $\boldsymbol{\mathcal{B}}_{(k)}$. Two sequences $\{\boldsymbol{X}_n; n \geq 1\}$ and $\{\boldsymbol{Y}_n; n \geq 1\}$ are then said to be asymptotically equivalent in the sense of type $(\boldsymbol{\mathcal{C}})_d$ and denoted briefly by

$$\boldsymbol{X}_n \sim \boldsymbol{Y}_n \ (\boldsymbol{\mathcal{C}})_d, \ (n \to \infty),$$

if it holds that

$$\delta_d(\boldsymbol{X}_n, \boldsymbol{Y}_n : \boldsymbol{\mathcal{C}}) = \sup_{E \in \mathcal{C}} \mid P^{X_n}(E) - P^{Y_n}(E) \mid \to 0, \text{ as } n \to \infty,$$

where P^{X_n} and P^{Y_n} designate the probability measures corresponding to the random variables $\boldsymbol{X}_n$ and $\boldsymbol{Y}_n$, respectively. First, it is noted that the quantity $\delta_d(\boldsymbol{X}_n, \boldsymbol{Y}_n : \boldsymbol{\mathcal{C}})$ defines a distance over the family of all probability distributions or of all random variables if we identify those random variables which have the same probability measure over $\boldsymbol{\mathcal{C}}$, $\boldsymbol{X}_n$ and $\boldsymbol{Y}_n$ being said to have the same probability measure over $\boldsymbol{\mathcal{C}}$ if it holds that $P^{X_n}(E) = P^{Y_n}(E)$ for every E belonging to the class. Next, in special case $\boldsymbol{\mathcal{C}} = \boldsymbol{\mathcal{B}}_{(k)}$, it holds that

$$\delta_d(\boldsymbol{X}_n, \boldsymbol{Y}_n : \boldsymbol{\mathcal{B}}_{(k)}) = \frac{1}{2}\int_{\boldsymbol{R}_{(k)}} |\, f_n(\boldsymbol{x}) - g_n(\boldsymbol{x}) \,|\, d\boldsymbol{x}.$$

Moreover, let φ is a measurable transformation from $(\boldsymbol{R}_{(k)}, \boldsymbol{\mathcal{B}}_{(k)})$ to another measurable space, then asymptotic equivalence is preserved by φ ([12], [14]). Type $(\boldsymbol{\mathcal{B}})_d$ asymptotic normality is sometimes called "uniform asymptotic normality" as in the title of the present thesis.

Now, we shall consider some of the familiar subclasses of $\boldsymbol{\mathcal{B}}_{(k)}$, as the subclass $\boldsymbol{\mathcal{C}}$ in the definition above:

$$\boldsymbol{\mathcal{M}}_{(k)} = \{(-\infty, a_1] \times \cdots \times (-\infty, a_k] \mid -\infty \le a_i \le \infty,\ i = 1, \cdots, k\},$$

$$\boldsymbol{\mathcal{S}}_{(k)} = \{(b_1, a_1] \times \cdots \times (b_k, a_k] \mid -\infty \le b_i \le a_i \le \infty,\ i = 1, \cdots, k\}.$$

$\boldsymbol{\mathcal{M}}_{(k)}$ is then a multiplicative class, and it immediately holds that $\boldsymbol{\mathcal{M}}_{(k)} \subset \boldsymbol{\mathcal{S}}_{(k)} \subset \boldsymbol{\mathcal{B}}_{(k)}$.

It should be noted that the convergence in the central limit theorem is always of type $(\boldsymbol{\mathcal{M}})_d$ ([12], [15]).

It is clear that $(\boldsymbol{\mathcal{B}})_d \Rightarrow (\boldsymbol{\mathcal{M}})_d$ in strongness between the asymptotic equivalence of type $(\boldsymbol{\mathcal{B}})_d$ and that of type $(\boldsymbol{\mathcal{M}})_d$, but the converse is not necessarily true. Now the usual asymptotic distribution theory has widely developed based on type $(\boldsymbol{\mathcal{M}})_d$ asymptotic equivalence. However, the argument based on type $(\boldsymbol{\mathcal{B}})_d$ asymptotic equivalence has been indispensability with the development of statistical theory and its applications. From this point of view, if the conditions under which two notions given above are mutually equivalent are clarified, then appling the results of type $(\boldsymbol{\mathcal{M}})_d$ to type $(\boldsymbol{\mathcal{B}})_d$ can be possible.

In Chapter 2, we shall give a sufficient condition under which two types of asymptotic equivalence, type $(\boldsymbol{\mathcal{B}})_d$ and type $(\boldsymbol{\mathcal{M}})_d$, are mutually equivalent in one-dimensional real case and also useful formulas for the numerical evaluation of related quantities, $\delta_d(\boldsymbol{X}_n, \boldsymbol{Y}_n : \mathcal{M})$, $\delta_d(\boldsymbol{X}_n, \boldsymbol{Y}_n : \mathcal{S})$ and $\delta_d(\boldsymbol{X}_n, \boldsymbol{Y}_n : \mathcal{B})$.

In Chapters 3 and 4, we shall consider the problem of the uniform normal approximation to the multivariate distributions. It is relatively well-known that both the Wishart distribution and the Dirichlet distribution converges in law to the normal distribution. However, it is unknown that its convergences are uniform. Now, these distributions are practically important distributions and its asymptotic normality has used by probability calculations. In Chapter 3, we shall prove that the Wishart distribution is asymptotically equivalent to the normal distribution with the same mean vector and variance-covariance matrix in the sense of type $(\mathcal{B})_d$ as the size n of the sample tends to infinity. The accuracy of the approximation is estimated by the upper bound of the uniform error based on the Kullback-Leibler information by using the general approximation theory which is stadied by Matsunawa([20], [21]), where the uniform error estimation is given by

$$\delta_d(\boldsymbol{X}_n, \boldsymbol{Y}_n : \mathcal{B}_{(k)}) \leq \sqrt{I(\boldsymbol{X}_n : \boldsymbol{Y}_n)/2},$$

([14]). In the same manner, the uniform normal approximation of the Dirichlet distribution is presented in Chapter 4.

Chapter 2

Implication Relation Between Two Types of Asymptotic Equivalence

2.1 Introduction

Let $\{X_n\,; n \geq 1\}$ and $\{Y_n\,; n \geq 1\}$ be two sequences of one-dimensional random variables, which are absolutely continuous with respect to the Lebesgue measure over the Borel σ-field $\boldsymbol{\mathcal{B}}$ on the real line $\boldsymbol{R}$. Let f_n and g_n be the probability density functions of X_n and Y_n, respectively.

Consider the subclasses of $\boldsymbol{\mathcal{B}}$:

$$\boldsymbol{\mathcal{M}} = \{(-\infty, a]\,; -\infty \leq a \leq \infty\},$$

$$\boldsymbol{\mathcal{S}} = \{(b, a]\,; -\infty \leq b \leq a \leq \infty\},$$

and define, for any given subclasses $\boldsymbol{\mathcal{C}}$ of $\boldsymbol{\mathcal{B}}$,

$$\delta_d(X_n, Y_n; \boldsymbol{\mathcal{C}}) = \sup_{E \in \mathcal{C}} |\ P^{X_n}(E) - P^{Y_n}(E)\ | \ .$$

Two sequences of random variables, $\{X_n\,; n \geq 1\}$ and $\{Y_n\,; n \geq 1\}$ are then said to be asymptotically equivalent in the sense of type $(\boldsymbol{\mathcal{C}})_d$ and denoted briefly by $X_n \sim Y_n\ (\boldsymbol{\mathcal{C}})_d$, $(n \to \infty)$, if it holds that

$$\delta_d(X_n, Y_n; \boldsymbol{\mathcal{C}}) \longrightarrow 0,\ (n \to \infty).$$

Then, since

$$(2.1) \qquad \delta_d(X_n, Y_n; \boldsymbol{\mathcal{M}}) \leq \delta_d(X_n, Y_n; \boldsymbol{\mathcal{S}}) \leq \delta_d(X_n, Y_n; \boldsymbol{\mathcal{B}})$$

and

$$\delta_d(X_n, Y_n; \boldsymbol{\mathcal{S}}) \leq 2\delta_d(X_n, Y_n; \boldsymbol{\mathcal{M}}),$$

we have the implication relations

$$(\boldsymbol{\mathcal{B}})_d \Rightarrow (\boldsymbol{\mathcal{S}})_d \Longleftrightarrow (\boldsymbol{\mathcal{M}})_d$$

among the three types of asymptotic equivalence.

Under the present setting of fixed basic space $\boldsymbol{R}$, if Y_n is identical to Y independently of n, Y being absolutely continuous, then the notion of $(\boldsymbol{\mathcal{M}})_d$-convergence, $X_n \to Y\ (\boldsymbol{\mathcal{M}})_d$, $(n \to \infty)$, is equivalent to the usual in law convergence.

In the following subsections we give a sufficient condition under which $(\boldsymbol{\mathcal{M}})_d$ and $(\boldsymbol{\mathcal{B}})_d$ are mutually equivalent. In section 2.3, some formulas for evaluating numerically the quantities given in (2.1) are presented.

As for the notions of asymptotic equivalence and its applications, the reader should refer to [11], [14], [15] and [16].

2.2 A sufficient condition

Let $\{A_n; n \geq 1\}$ be a sequence of main domains of $\{X_n; n \geq 1\}$ and $\{Y_n; n \geq 1\}$, *i.e.*, $A_n \in \boldsymbol{\mathcal{B}}$, $P^{X_n}(A_n) \to 1$, $P^{Y_n}(A_n) \to 1$ as $n \to \infty$.

Consider the Hahn-decomposition ([8]) of $\boldsymbol{R}$ with respect to the signed measure $P^{X_n} - P^{Y_n} (\equiv Q_n \text{ say})$:

$$\boldsymbol{R}_n^+ = \{x; f_n(x) - g_n(x) > 0\}, \qquad \boldsymbol{R}_n^- = \boldsymbol{R} - \boldsymbol{R}_n^+,\ (n \geq 1),$$

then, it is easy to see that

$$\delta_d(X_n, Y_n; \boldsymbol{\mathcal{B}}) = Q_n(\boldsymbol{R}_n^+), \quad (n \geq 1).$$

Let us put

$$W_n = A_n \bigcap \boldsymbol{R}_n^+, \quad (n \geq 1).$$

First we note the following lemma.

Lemma 2.1 *Suppose that the set W_n is the sum of a finite number of disjoint intervals:*

(2.2) $$W_n = \bigcup_{i=1}^{k_n} I_n^i, \quad I_n^i \bigcap I_n^j = \phi \ (i \neq j),$$

for all n. Then, it holds that

$$\delta_d(X_n, Y_n; \boldsymbol{\mathcal{B}}) \leq 2k_n \delta_d(X_n, Y_n; \boldsymbol{\mathcal{M}}) + \max\{P^{X_n}(A_n^c),\ P^{Y_n}(A_n^c)\},$$

where $A_n^c = \boldsymbol{R} - A_n$. (Each I_n^i may be considered as a member of $\boldsymbol{S}$.)

Thus, the following theorem gives a sufficient condition for $(\boldsymbol{\mathcal{M}})_d$ and $(\boldsymbol{\mathcal{B}})_d$ to be equivalent, which is straightforward from the lemma above.

Theorem 2.1 *For some sequence of asymptotic main domains $\{A_n ; n \geq 1\}$ of both $\{X_n ; n \geq 1\}$ and $\{Y_n ; n \geq 1\}$, suppose that (2.2) holds. Then, the condition*

$$k_n \delta_d(X_n, Y_n; \boldsymbol{\mathcal{M}}) \longrightarrow 0, \ (n \to \infty)$$

implies that

(2.3) $$\delta_d(X_n, Y_n; \boldsymbol{\mathcal{B}}) \longrightarrow 0, \ (n \to \infty)$$

and in particular, if there exists a positive constant k such that

$$k_n \leq k \quad (n \geq 1),$$

then

$$\delta_d(X_n, Y_n; \boldsymbol{\mathcal{M}}) \longrightarrow 0, \ (n \to \infty)$$

implies (2.3).

In what follows, some remarks are given on the evaluation of the number k_n.

Let $h(x)$ be a function defined over $\boldsymbol{R}$, which is left-continuous and has a finite number of discontinuities. Put

$$\boldsymbol{R}^+ = \{x; h(x) > 0\}.$$

Suppose that the set $\boldsymbol{R}^+$ is a disjoint sum of finite number k of the members of $\boldsymbol{S}$.

Let us designate by $[h]_0$, $[h]_{op}$ and $[h]_{dc}$, the number of zeros of h, that of optimal(extremum) points and of discontinuities, respectively. Then, we have the following lemma.

Lemma 2.2

(2.4) $$k \leq \frac{1}{2}\{[h]_0 + [h]_{dc}\} + 1,$$

(2.5) $$[h]_0 \leq [h]_{op} + [h]_{dc} + 1.$$

Proof. Let $\boldsymbol{R}^+$ be the sum of intervals $(b_i, a_i]$, $i = 1, 2, \cdots, k$, with $-\infty \leq b_1 < a_1 < b_2 < a_2 < \cdots < b_k < a_k \leq \infty$. Then, among the $2k$ points, a_i and b_i, at least $2k-2$ points must be zeros or discontinuities. Hence, $2(k-1) \leq [h]_0 + [h]_{dc}$, which implies (2.4).

Let $z_1 < z_2 < \cdots < z_s$, $s \geq 2$, be successive zeros of $h(x)$ such that z_1 and z_s are continuity points and the rest $s-2$ points are discontinuities. Then, since $h(x)$ is left-continuous, the interval $(z_1, z_2]$ contains at least one local optimum, and so does $(z_1, z_s]$.

Thus, if $[h]_0 = \infty$ then $[h]_{op} = \infty$, and therefore (2.5) is trivial. Supposing $[h]_0$ be finite, let the zeros of $h(x)$ be $-\infty < z_1 < z_2 < \cdots < z_m < \infty$, among which $z_{i_1} < z_{i_2} < \cdots < z_{i_n}$ are assumed to be continuity points and the other $m-n$ zeros and discontinuities. Then, each of the intervals $(z_{i_1}, z_{i_2}], (z_{i_2}, z_{i_3}], \cdots, (z_{i_{n-1}}, z_{i_n}]$ has at least one local optimum and hence at least $n-1$ local optimums in total. That is, $[h]_{op} \geq n-1$ and $[h]_{dc} \geq m-n$. Hence we have (2.5).

The following results are straightforward from the lemma above.

Lemma 2.3 *If h(x) continuous, then it holds that*

$$k \leq \frac{1}{2}[h]_0 + 1,$$

and

$$[h]_0 \leq [h]_{op} + 1.$$

Furthermore, if $h(x)$ is continuous and differentiable over $\boldsymbol{R}$, then

$$[h]_{op} = [h']_0.$$

2.3 Numerical evaluation of the related quantities

In the present section, we shall give useful formulas for evaluating numerically the quantities $\delta_d(X_n, Y_n; \boldsymbol{\mathcal{B}})$, $\delta_d(X_n, Y_n; \boldsymbol{\mathcal{S}})$, and $\delta_d(X_n, Y_n; \boldsymbol{\mathcal{M}})$, when the condition (2.2) holds. Without any loss of generality, we assume that $A_n = \boldsymbol{R}$. Throughout the present section, the suffix n is omitted.

As before, let $\boldsymbol{R}^+ = \{x; f(x) - g(x) \geq 0\}$ and $\boldsymbol{R}^- = \boldsymbol{R} - \boldsymbol{R}^+$ be a Hahn-decomposition of $\boldsymbol{R}$ with respect to the signed measure $Q = P^X - P^Y$. If $\boldsymbol{R}^+$ is the disjoint sum of k intervals in $\boldsymbol{\mathcal{S}}$, then $\boldsymbol{R}^-$ is also the disjoint sum of a finite number of intervals in $\boldsymbol{\mathcal{S}}$. Let the whole intervals be

$$I_1,\ I_2,\ I_3,\ \cdots,\ I_{2k},\ I_{2k+1}$$

in this order from the left, for which it is assumed without any loss of generality that $Q(I_{2i}) > 0$ and $Q(I_{2i-1}) < 0$, and therefore $\boldsymbol{R}^+ = \sum_{i=1}^{k} I_{2i}$ and $\boldsymbol{R}^- = \sum_{i=0}^{k} I_{2i+1}$. Put

$$I_i = (a_i, a_{i+1}],\ \ i = 1,\ 2,\ \cdots,\ 2k+1,$$

where $a_1 = -\infty$. If $a_{2k+1} = \infty$, we assume that $I_{2k+1} = \emptyset$.

Now, let $E_z = (-\infty, z]$ be a set in $\boldsymbol{\mathcal{M}}$ which attains the value of $\delta_d(X, Y; \boldsymbol{\mathcal{M}})$; such E_z exists since P^X and P^Y are assumed to be absolutely continuous with respect to the Lebesgus measure over $(\boldsymbol{R}, \boldsymbol{\mathcal{B}})$. Then, it is shown that

Lemma 2.4 *The set E_z must be identical with either one of the intervals*

$$\sum_{j=1}^{s} I_j,\ \ s = 1,\ 2,\ \cdots,\ 2k,$$

i.e.,

(2.6)
$$\delta_d(X, Y; \mathcal{M}) = \max_{1 \le s \le 2k} \{| Q(\sum_{j=1}^{s} I_j) |\}.$$

Proof. First, assume that $a_{2k+1} = \infty$, and suppose that z falls in I_{2i}. Then, if $Q(E_z) > 0$, it holds that $Q(E_z) \le Q(\sum_{j=1}^{2i} I_j)$, and if $Q(E_z) \le 0$, then $| Q(E_z) | \le | Q(\sum_{j=1}^{2i-1} I_j) |$. Hence, it is also clear that z can not fall in the last interval I_{2k}. When $z \in I_{2i+1}$, it holds that $Q(E_z) \le Q(\sum_{j=1}^{2i} I_j)$ or $| Q(E_z) | \le | Q(\sum_{j=1}^{2i+1} I_j) |$ according as $Q(E_z) > 0$ or < 0, respectively. Thus, we have

$$| Q(E_z) | = \max_{1 \le s \le 2k-1} \{| Q(\sum_{j=1}^{s} I_j) |\}.$$

Next, assume that $a_{2k+2} = \infty$. In this case, we also have

$$| Q(E_z) | = \max_{1 \le s \le 2k} \{| Q(\sum_{j=1}^{s} I_j) |\}.$$

Thus, we obtain (2.6).

In the next place, a formula for evaluating $\delta_d(X, Y; \mathcal{S})$ is given. First, we have the following lemma.

Lemma 2.5 *If a set* $E_{u,\nu} = (u, \nu]$ *attains* $\delta_d(X, Y; \mathcal{S})$, *then it holds that*

$$u = a_i \text{ and } \nu = a_{i'},$$

for some i *and* i', $1 < i < i' \le 2k + 1$, *i.e.*,

(2.7)
$$E_{u,\nu} = \sum_{j=1}^{i'-1} I_j.$$

Proof. Suppose that $u \in I_s$ and $\nu \in I_t$ for some $1 \le s < t \le 2k + 2$. Assume first that $Q(E_{u,\nu}) > 0$. If both s and t are even, then we have

$$Q(E_{u,\nu}) \le Q(E_{a_s, a_{t+1}}) = Q(\sum_{i=s}^{t} I_i).$$

If s is even and t is odd, then

$$Q(E_{u,\nu}) \le Q(E_{a_s, a_t}) = Q(\sum_{i=s}^{t-1} I_i),$$

if s is odd and t is even, then

$$Q(E_{u,\nu}) \leq Q(E_{a_{s+1},a_{t+1}}) = Q(\sum_{i=s+1}^{t} I_i),$$

and if both s and t are odd, then

$$Q(E_{u,\nu}) \leq Q(E_{a_{s+1},a_t}) = Q(\sum_{i=s+1}^{t-1} I_i).$$

Similar argument can be applied in case where $Q(E_{u,\nu}) < 0$, and therefore, we obtain (2.7) in all cases.

By this lemma, we get

$$\delta_d(X, Y; \boldsymbol{S}) = \max_{1 \leq s < t \leq 2k+1} \{| Q(\sum_{i=s}^{t} I_i) |\} \tag{2.8}$$

where we have assumed that $a_{2k+2} = \infty$. (In case $a_{2k+1} = \infty$, then the maximum in (2.8) should be taken over all integers such that $1 \leq s < t \leq 2k$.) Hereafter we assume that $a_{2k+2} = \infty$.

Now, some cases can be discarded from the right-hand side of (2.8). First, it is evident that one can discard the case $s = 1$ and $t = 2k + 1$, or $\sum_{i=1}^{2k+1} I_i = \boldsymbol{R}$. We now have the following lemma.

Lemma 2.6 *The sum of intervals $\sum_{i=s}^{t} I_i$ can not attain the maximum on the right-hand side of (2.8), if $Q(I_s)$ and $Q(I_t)$ have different signs.*

Proof. Suppose that $Q(I_s) > 0$ and $Q(I_t) < 0$. If $Q(\sum_{i=s}^{t} I_i) > 0$, then it holds that, for $s + 1 \leq t \leq 2k + 1$, $Q(\sum_{i=s}^{t} I_i) \leq Q(\sum_{i=s}^{t-1} I_i)$, because I_{t+1} is contained in $\boldsymbol{R}^{+}$. Also, if $Q(\sum_{i=s}^{t} I_i) < 0$, then, for $1 \leq s$, $| Q(\sum_{i=s}^{t} I_i) | \leq | Q(\sum_{i=s-1}^{t} I_i) |$. Next, suppose that $Q(I_s) < 0$ and $Q(I_t) > 0$. If $Q(\sum_{i=s}^{t} I_i) > 0$, then, for $1 \leq s$ and $s + 1 \leq t \leq 2k + 1$, $Q(\sum_{i=s}^{t} I_i) \leq Q(\sum_{i=s+1}^{t} I_i)$ and if $Q(\sum_{i=s}^{t} I_i) < 0$, then, for $1 \leq s$, $| Q(\sum_{i=s}^{t} I_i) | \leq | Q(\sum_{i=s}^{t-1} I_i) |$. This completes the proof of the lemma.

The lemma above means that one can discard the sums of any even number of intervals from (2.8), *i.e.*,

$$\delta_d(X, Y; \boldsymbol{S}) = \max_{1 \leq s,\, 0 \leq w,\, s+2w \leq 2k+1} \{| Q(\sum_{i=s}^{s+2w} I_i) |\}.$$

In case where $a_{2k+1} = \infty$, then

$$\delta_d(X,Y;\boldsymbol{S}) = \max_{1\leq s,\,0\leq w,\,s+2w\leq 2k}\{\mid Q(\sum_{i=s}^{s+2w} I_i) \mid\}.$$

For example, if $k = 2$, then

$$\begin{aligned}\delta_d(X,Y;\boldsymbol{S}) = \max\{ & \mid Q(I_1) \mid, \mid Q(I_2) \mid, \mid Q(I_3) \mid, \mid Q(I_4) \mid, \\ & \mid Q(I_1 + I_2 + I_3) \mid, \mid Q(I_2 + I_3 + I_4) \mid \}.\end{aligned}$$

Summarizing the results thus obtained, we state the following theorem.

Theorem 2.2 *Case (1): $a_{2k+2} = \infty$ and hence we have intervals $I_1, I_2, \cdots, I_{2k}, I_{2k+1}$ in all, for which $I_{2i} \subset \boldsymbol{R}^+$, $i = 1,2,\cdots,k$, and $I_{2i+1} \subset \boldsymbol{R}^-$, $i = 0,1,\cdots,k$. In this case, it holds that*

$$\begin{aligned}\delta_d(X,Y;\boldsymbol{B}) &= \sum_{i=1}^{k} Q(I_{2i}),\\ \delta_d(X,Y;\boldsymbol{S}) &= \max_{1\leq s,\,0\leq w,\,s+2w\leq 2k+1}\{\mid \sum_{i=s}^{s+2w} Q(I_i) \mid\},\end{aligned}$$

and

$$\delta_d(X,Y;\boldsymbol{\mathcal{M}}) = \max_{1\leq s\leq 2k}\{\mid \sum_{i=1}^{s} Q(I_i) \mid\}.$$

Case(2): $a_{2k+1} = \infty$, and hence $I_{2k+1} = \emptyset$. In this case,

$$\begin{aligned}\delta_d(X,Y;\boldsymbol{B}) &= \sum_{i=1}^{k} \mid Q(I_{2i}) \mid,\\ \delta_d(X,Y;\boldsymbol{S}) &= \max_{1\leq s,\,0\leq w,\,s+2w\leq 2k}\{\mid \sum_{i=s}^{s+2w} Q(I_i) \mid\},\end{aligned}$$

and

$$\delta_d(X,Y;\boldsymbol{\mathcal{M}}) = \max_{1\leq s\leq 2k-1}\{\mid \sum_{i=1}^{s} Q(I_i) \mid\}.$$

Example 2.1 Let Z_n be a normal random variable with mean n and variance $2n$. On the other hand, let X_n be a chi-square variable with n degrees of freedom. Then $E(X_n) = n$ and $Var(X_n) = 2n$. Let $\varphi_n(x)$ and $f_n(x)$ be their probability density functions respectively;

$$\varphi_n(x) = \frac{1}{2\sqrt{n\pi}} e^{-(x-n)^2/4n}, \ (-\infty < x < \infty),$$

$$f_n(x) = \begin{cases} \frac{1}{\sqrt{2^n}\Gamma(n/2)} x^{n/2-1} e^{-x/2} & \text{for } x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Now, the Kullback-Leibler information can be evaluated for sufficiently large n as follows;

$$I(X_n : Z_n) \sim \frac{1}{n},$$

which shows that the chi-square distribution is asymptotically normal and uniform over the Borel field. (The reader should refer to Capter 3.)

Numerical values of the error of uniform approximation, $\delta_d(X_n, Z_n; \boldsymbol{\mathcal{B}})$, together with the values of $\delta_d(X_n, Z_n; \boldsymbol{\mathcal{S}})$ and $\delta_d(X_n, Z_n; \boldsymbol{\mathcal{M}})$ are tabulated in the following table 2.2, and the related measures $Q(I_i)$ are tabulated in table 2.1.

Table 2.1.

Values of $Q(I_i)$.

$n \setminus Q$	$Q(I_1)$	$Q(I_2)$	$Q(I_3)$	$Q(I_4)$
10	−0.02784	0.10015	−0.07805	0.01842
20	−0.02431	0.06721	−0.05655	0.01444
30	−0.01932	0.05374	−0.04672	0.01235
40	−0.01623	0.04599	−0.04076	0.01100
50	−0.01420	0.04082	−0.03664	0.01003
60	−0.01276	0.03705	−0.03358	0.00929
70	−0.01166	0.03415	−0.03118	0.00870
80	−0.01080	0.03184	−0.02924	0.00821
90	−0.01010	0.02993	−0.02763	0.00780
100	−0.00952	0.02833	−0.02626	0.00744

Table 2.2.

Values of $\delta_d(X_n, Z; \mathcal{B})$, $\delta_d(X_n, Z; \mathcal{S})$ and $\delta_d(X_n, Z; \mathcal{M})$.

$n \setminus \delta_d$	$\delta_d(X_n, Z_n; \mathcal{B})$	$\delta_d(X_n, Z_n; \mathcal{S})$	$\delta_d(X_n, Z_n; \mathcal{M})$.
10	0.11857	0.10015	0.07231
20	0.08165	0.06721	0.04290
30	0.06609	0.05374	0.03442
40	0.05699	0.04599	0.02976
50	0.05085	0.04082	0.02662
60	0.04634	0.03705	0.02429
70	0.04285	0.03415	0.02249
80	0.04005	0.03184	0.02104
90	0.03773	0.02993	0.01983
100	0.03577	0.02833	0.01881

Example 2.2 Let Z be the standard normal random variable and let X_n be that of Student's t-distribution of n degrees of freedom. Let $\varphi(x)$ and $f_n(x)$ be their probability density functions respectively;

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \; (-\infty < x < \infty),$$

$$f_n(x) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)}(1 + \frac{x^2}{n})^{-(n+1)/2}, \; (-\infty < x < \infty).$$

It is well-known that the t-distribution tends to the standard normal distribution uniformly as $n \to \infty$. The related measures and quantities for some values of n are tabulated in the following table 2.3 and table 2.4 respectively.

Table 2.3.

Values of $Q(I_i)$.

$n \setminus Q$	$Q(I_1)$	$Q(I_2)$	$Q(I_3)$
10	−0.015542	0.031084	−0.015542
20	−0.007844	0.015687	−0.007844
30	−0.005245	0.010489	−0.005245
40	−0.003939	0.007878	−0.003939
50	−0.003154	0.006308	−0.003154
60	−0.002630	0.005260	−0.002630
70	−0.002255	0.004510	−0.002255
80	−0.001974	0.003948	−0.001974
90	−0.001755	0.003509	−0.001755
100	−0.001580	0.003160	−0.001580

Table 2.4.

Values of $\delta_d(X_n, Z; \mathcal{B})$, $\delta_d(X_n, Z; \mathcal{S})$ and $\delta_d(X_n, Z; \mathcal{M})$.

$n \setminus \delta_d$	$\delta_d(X_n, Z_n; \mathcal{B})$	$\delta_d(X_n, Z_n; \mathcal{S})$	$\delta_d(X_n, Z_n; \mathcal{M})$.
10	0.031084	0.031084	0.015542
20	0.015687	0.015687	0.007844
30	0.010489	0.010489	0.005245
40	0.007878	0.007878	0.003939
50	0.006308	0.006308	0.003154
60	0.005260	0.005260	0.002630
70	0.004510	0.004510	0.002255
80	0.003948	0.003948	0.001974
90	0.003509	0.003509	0.001755
100	0.003160	0.003160	0.001580

Chapter 3

Uniform Asymptotic Normality of the Wishart Distribution

3.1 Introduction

Nonaka [22] studied the uniform asymptotic normality of the Wishart distribution in the canonical case. We shall consider the same problem in the general case. The aim of this chapter is to give exact evaluation based on an inequality for the uniform error, and consequently we are to clarify the condition for the uniform convergence.

In [20] and [21], Matsunawa has given fairly sharp bounds(lower and upper) of the real gamma function and digamma function by evaluating the corresponding series of inverse factorials. This results are useful to our present purpose to evaluate uniform error based on the Kullback-Leibler information between the Wishart distribution and the normal distribution.

In Section 3.2, we shall show that the general case can be reduced to the canonical case. In Section 3.3, we shall give some lemmas, and in Section 3.4, we shall prove the uniform(*i.e.*, type $(\mathcal{B})_d$) asymptotic normality of the Wishart distribution $W_p(n, \Lambda)$ under the condition $p^3/n \to 0$, by giving the upper bound of the uniform error based on the Kullback-Leibler information. The condition $p^3/n \to 0$ is the best possible for which the information converges to zero.

Let $\boldsymbol{X}_\alpha = (X_{1\alpha}, X_{2\alpha}, \cdots, X_{p\alpha})'$, $\alpha = 1, 2, \cdots, N$, be a random sample of size N drawn from a p-dimensional non-degenerate normal distribution with mean vector $\boldsymbol{\mu} = (\mu_1, \mu_2,$

$\cdots, \mu_p)'$ and variance-covariance matrix $\Lambda = (\lambda_{ij})_{p\times p}$.

Let us put

$$A = \sum_{\alpha=1}^{N}(\boldsymbol{X}_\alpha - \overline{\boldsymbol{X}})(\boldsymbol{X}_\alpha - \overline{\boldsymbol{X}})'$$

with $\overline{\boldsymbol{X}} = \frac{1}{N}\sum_{\alpha=1}^{N}\boldsymbol{X}_\alpha$. Further, let us denote by $\boldsymbol{A}$ the $s = p(p+1)/2$ dimensional vector of distinct elements of A;

$$\boldsymbol{A} = (a_{11}, a_{12}, \cdots, a_{1p}, a_{22}, \cdots, a_{2p}, \cdots, a_{pp})'.$$

Then the density function of $\boldsymbol{A}$ is given by

$$f(\boldsymbol{A}) = \frac{\mid A \mid^{(n-p-1)/2}\exp(-\mathrm{tr}(\Lambda^{-1}A)/2)}{2^{np/2}\pi^{p(p-1)/4}\mid \Lambda \mid^{n/2}\prod_{i=1}^{p}\Gamma((n+1-i)/2)}, \tag{3.1}$$

where we have put $n = N - 1$, which is called the Wishart distribution $W_p(n, \Lambda)$. The range of variation of the components of $\boldsymbol{A}$ is over all values such that A is positive definite. We discuss the uniform asymptotic normality of this distribution under the limiting $N \to \infty$, where p may also vary with N.

It is known that

$$\begin{aligned} E(a_{ij}) &= n\lambda_{ij}, \quad i, j = 1, 2, \cdots, p, \\ Cov(a_{ij}, a_{kl}) &= n(\lambda_{ik}\lambda_{jl} + \lambda_{il}\lambda_{jk}), \\ Var(a_{ij}) &= n(\lambda_{ij}^2 + \lambda_{ii}\lambda_{jj}). \end{aligned} \tag{3.2}$$

Now, let $\boldsymbol{A}^*$ be a $s \times 1$ random vector having the s-dimensional normal distribution with mean $n\boldsymbol{\lambda}$ and variance matrix $n\Omega$, $N(n\boldsymbol{\lambda}, n\Omega)$, where we have put

$$\begin{aligned} E(\boldsymbol{A}) &= n\boldsymbol{\lambda}, \\ Var(\boldsymbol{A}) &= n\Omega. \end{aligned}$$

The probability density function of $\boldsymbol{A}^*$ is given by

$$g(\boldsymbol{A}) = (2\pi)^{-s/2}\mid n\Omega \mid^{-1/2}\exp\left\{-\frac{1}{2}(\boldsymbol{A} - n\boldsymbol{\lambda})'(n\Omega)^{-1}(\boldsymbol{A} - n\boldsymbol{\lambda})\right\}.$$

We shall show that under some conditions $\boldsymbol{A}$ and $\boldsymbol{A}^*$ are asymptotically equivalent $(\boldsymbol{\mathcal{B}})_d$ as $N \to \infty$, which is denoted by $\boldsymbol{A} \sim \boldsymbol{A}^*\ (\boldsymbol{\mathcal{B}})_d$ and means that

$$(3.3)\qquad \delta_d(\boldsymbol{A}, \boldsymbol{A}^* : \boldsymbol{\mathcal{B}}_{(s)}) = \sup_{E \in \mathcal{B}_{(s)}} \mid P^{\boldsymbol{A}}(E) - P^{\boldsymbol{A}^*}(E) \mid \longrightarrow 0,\ (N \to \infty),$$

where $\boldsymbol{\mathcal{B}}_{(s)}$ is the Borel field in the s-dimensional Euclidean space $\boldsymbol{R}_{(s)}$. A sufficient condition for this is given by

$$(3.4)\qquad I(\boldsymbol{A} : \boldsymbol{A}^*) = E_{\boldsymbol{A}}\left[\log \frac{f(\boldsymbol{A})}{g(\boldsymbol{A})}\right] \longrightarrow 0,\ (N \to \infty),$$

and an error estimation is given by

$$(3.5)\qquad \delta_d(\boldsymbol{A}, \boldsymbol{A}^* : \boldsymbol{\mathcal{B}}_{(s)}) \le \sqrt{I(\boldsymbol{A} : \boldsymbol{A}^*)/2}.$$

It should be noted that while under the present situation the inclusion relation of the carriers supp$(f) \subset$ supp (g) holds, $I(\boldsymbol{A}^* : \boldsymbol{A}) = E_{\boldsymbol{A}}[\log\{g(\boldsymbol{A})/f(\boldsymbol{A})\}]$ is undefined.

Remark. The error estimation (3.5) is proved as follows. Consider a decomposition of $\boldsymbol{R}_{(s)}$ such that $\boldsymbol{R}^+_{(s)} \equiv \{E; P^{\boldsymbol{A}}(E) \ge P^{\boldsymbol{A}^*}(E),\ E \in \boldsymbol{\mathcal{B}}_{(s)}\}$ and $\boldsymbol{R}^-_{(s)} \equiv \boldsymbol{R}_{(s)} - \boldsymbol{R}^+_{(s)}$. Then, we have $\delta_d(\boldsymbol{A}, \boldsymbol{A}^* : \boldsymbol{\mathcal{B}}_{(s)}) = P^{\boldsymbol{A}}(\boldsymbol{R}^+_{(s)}) - P^{\boldsymbol{A}^*}(\boldsymbol{R}^+_{(s)}) = P^{\boldsymbol{A}^*}(\boldsymbol{R}^-_{(s)}) - P^{\boldsymbol{A}}(\boldsymbol{R}^-_{(s)})$. Now, Kullback [19] gave the following inequality

$$I(\boldsymbol{A} : \boldsymbol{A}^*) \ge P^{\boldsymbol{A}}(\boldsymbol{R}^+_{(s)}) \log \frac{P^{\boldsymbol{A}}(\boldsymbol{R}^+_{(s)})}{P^{\boldsymbol{A}^*}(\boldsymbol{R}^+_{(s)})} + P^{\boldsymbol{A}}(\boldsymbol{R}^-_{(s)}) \log \frac{P^{\boldsymbol{A}}(\boldsymbol{R}^-_{(s)})}{P^{\boldsymbol{A}^*}(\boldsymbol{R}^-_{(s)})}.$$

On the other hand, it holds that

$$(3.6)\qquad p_1 \log \frac{p_1}{p_2} + q_1 \log \frac{q_1}{q_2} \ge 2(p_1 - p_2)^2$$

for $p_1 + q_1 = p_2 + q_2 = 1,\ 0 < p_1, p_2 < 1$. (*cf.* Kraft and Schmitz [18] gave a sharper inequality to (3.6).) Thus, combing the two results above, we get the error estimation (3.5).

3.2 Uniform asymptotic normality in canonical case

Let us consider the vectors

$$\boldsymbol{Z}_\alpha = \Lambda^{-1/2}(\boldsymbol{X}_\alpha - \boldsymbol{\mu}),\ \ \alpha = 1, 2, \cdots, N,$$

which are mutually independent and each has the standard normal, $N(0, I_p)$. For these variables

$$\overline{Z} = \Lambda^{-1/2}(\overline{X} - \mu),$$

and therefore

$$(3.7) \qquad \begin{aligned} B &= \sum_{\alpha=1}^{N}(Z_\alpha - \overline{Z})(Z_\alpha - \overline{Z})' \\ &= \Lambda^{-1/2}A\Lambda^{-1/2}. \end{aligned}$$

The transformation (3.7), $A \to B$, is non-singular, *i.e.*, it is measurable and one-to-one, and hence the inverse mapping is well-defined and measurable.

As before, let $\boldsymbol{B} = (b_{11}, \cdots, b_{1p}, b_{22}, \cdots, b_{2p}, \cdots, b_{pp})'$, where $b_{ij}(i \leq j,\ i, j = 1, 2, \cdots, p)$ are $s = p(p+1)/2$ distinct elements of the matrix B.

From (3.2), it then follows that $E(\boldsymbol{B}) = n\boldsymbol{I}_p = n(1, 0, \cdots, 0, 1, 0, \cdots, 0, \cdots, 1)' \equiv n\boldsymbol{\eta}$, or

$$E(b_{ij}) = n\delta_{ij}, \quad (i \leq j,\ i, j = 1, 2, \cdots, p).$$

Also,

$$(3.8) \qquad Cov(b_{ij}, b_{kl}) = 0 \ \text{ if } (i, j) \neq (k, l),$$

$$Var(b_{ij}) = \begin{cases} 2n & \text{if } i = j \\ n & \text{if } i \neq j \end{cases}$$

or

$$(3.9) \qquad Var(\boldsymbol{B}) = n \begin{pmatrix} 2 & & & & & & & & & & & \\ & 1 & & & & & & & & & & \\ & & \ddots & & & & & & & & 0 & \\ & & & 1 & & & & & & & & \\ & & & & 2 & & & & & & & \\ & & & & & 1 & & & & & & \\ & & & & & & \ddots & & & & & \\ & & & & & & & 1 & & & & \\ & & & & & & & & 2 & & & \\ & & 0 & & & & & & & 1 & & \\ & & & & & & & & & & \ddots & \\ & & & & & & & & & & & 2 \end{pmatrix}_{s \times s} \equiv n\Omega_o.$$

Let B^* be a $s \times 1$ random vector having the normal distribution $N(n\eta, n\Omega_o)$; whose density is given by

$$(3.10) \qquad g_o(B) = (2\pi)^{-s/2} \mid n\Omega_o \mid^{-1/2} \exp\left\{-\frac{1}{2}(B - n\eta)'(n\Omega_o)^{-1}(B - n\eta)\right\}.$$

On the other hand, B has the Wishart distribution

$$(3.11) \qquad f_o(B) = C_{np}^{-1} \mid B \mid^{(n-p-1)/2} \exp\left(-\frac{1}{2}\mathrm{tr}B\right)$$

where

$$C_{np} = 2^{np/2}\pi^{p(p-1)/4}\prod_{i=1}^{p}\Gamma\left(\frac{n+1-i}{2}\right).$$

We now quote a result on measurable transformation which transfer a type of asymptotic equivalence to the same or to another([12], [14]).

Lemma 3.1 *Let $\varphi(x)$ be any given $\mathcal{B}_{(s)}$-measurable mapping from $R_{(s)}$ to any Euclidean space, say $(R_{(t)}, \mathcal{B}_{(t)})$, and put $X' = \varphi(X)$, $Y' = \varphi(Y)$. Then $X \sim Y\,(\mathcal{B})_d$, $(n \to \infty)$, implies that $X' \sim Y'\,(\mathcal{B})_d$, in which case*

$$\delta_d(X, Y : \mathcal{B}_{(s)}) \geq \delta_d(X', Y' : \mathcal{B}_{(t)}).$$

Furthermore, if φ is non-singular, then $X \sim Y\,(\mathcal{B})_d$ and $X' \sim Y'\,(\mathcal{B})_d$ are mutually equivalent, and

$$\delta_d(X, Y : \mathcal{B}_{(s)}) = \delta_d(X', Y' : \mathcal{B}_{(t)}).$$

Thus, in order to prove (3.3), it is sufficient to show that

$$(3.12) \qquad I(B : B^*) = E_B\left[\log\frac{f_o(B)}{g_o(B)}\right] \longrightarrow 0, \quad (N \to \infty),$$

and an error estimation is given by

$$\delta_d(A, A^* : \mathcal{B}_{(s)}) = \delta_d(B, B^* : \mathcal{B}_{(s)}) \leq \sqrt{I(B : B^*)/2}.$$

Therefore, in order to show (3.4), it suffices to prove it in the canonical case where $\mu = 0$ and $\Lambda = I_p$.

3.3 Lemmas

In this section, we state some lemmas which play fundamental roles for the calculations of the Kullback-Leibler information. We shall begin with the following.

Lemma 3.2 *Let X be a random variable having the chi-square distribution of n degrees of freedom. Then*

$$E[\log X] = \log 2 + \frac{\Gamma'(n/2)}{\Gamma(n/2)}.$$

The proof of this lemma is simple and will be omitted.

Lemma 3.3 (Matsunawa)

$$\log\Gamma(x) = \frac{1}{2}\log 2\pi + (x - \frac{1}{2})\log x - x - R(x), \quad (x > 0),$$

where

$$0 < R(x) \equiv \sum_{i=1}^{\infty} \frac{a_{i+1}}{x(x+1)(x+2)\cdots(x+i)} < \frac{1}{64x^2(x+1)},$$

and a_{i+1} are defined by

$$a_{i+1} = \frac{1}{i+1}\int_0^1 t(1-t)(2-t)\cdots(i-t)(\frac{1}{2}-t)dt, \quad (i \geq 1).$$

Proof. See Formula 1 in [21].

Lemma 3.4 (Matsunawa)

$$\frac{\Gamma'(x)}{\Gamma(x)} = \log x - \frac{1}{2x} - \frac{1}{x}T(x), \quad (x > 1),$$

where

$$\frac{1}{72(x-1)} < T(x) \equiv \sum_{i=1}^{\infty} \frac{b_{i+1}}{(x+1)(x+2)\cdots(x+i)} < \frac{1}{12(x-1)}$$

and b_{i+1} are defined by

$$b_{i+1} = \frac{1}{i+1}\int_0^1 t(1-t)(2-t)\cdots(i-t)dt, \quad (i \geq 1).$$

Proof. See [20] (pp. 303-304).

Lemma 3.5 *Let* $Z_k = (z_{ij})$, $(1 \le i, j \le k)$, *be a symmetric matrix and* $z_{ii} = 1$, $(i = 1, 2, \cdots, k)$. *For* $2 \le k < n$,

$$
\begin{aligned}
(3.13)\qquad L_k &= \int_{Z_k > 0} | Z_k |^{(n-k-2)/2} \log | Z_k | \, dz_{12} \cdots dz_{k-1,k} \\
&= -\pi^{k(k-1)/4} \left\{ \prod_{i=1}^{k} \frac{\Gamma((n-i)/2)}{\Gamma((n-1)/2)} \right\} \left\{ \sum_{i=1}^{k-1} (k-i)\Delta_{ni} \right\},
\end{aligned}
$$

where $Z_k > 0$ *means that* Z_k *is positive definite, and*

$$
(3.14)\qquad \Delta_{ni} = \sum_{r=0}^{\infty} \frac{1}{r+1} \prod_{t=0}^{r} \frac{1+2t}{n-i+2t}.
$$

Remarks. (i) Under the same situation as in the lemma 3.5 above, Cramér [7] (pp. 392-393) gives the following result

$$
\begin{aligned}
(3.15)\qquad J_k &= \int_{Z_k > 0} | Z_k |^{(n-k-2)/2} dz_{12} \cdots dz_{k-1,k} \\
&= \pi^{k(k-1)/4} \prod_{i=1}^{k} \frac{\Gamma((n-i)/2)}{\Gamma((n-1)/2)}, \quad (2 \le k < n).
\end{aligned}
$$

The lemma 3.5 is proved by induction over k in the same way with the result.

(ii) Let

$$
a_r = \frac{1}{r+1} \prod_{t=0}^{r} \frac{1+2t}{n-i+2t},
$$

then we easily obtain

$$
\lim_{r \to \infty} r \left(\frac{a_r}{a_{r+1}} - 1 \right) = \frac{1}{2}(n-i+1) > 1, \quad (1 \le i \le k-1).
$$

Hence, the positive term series (3.14) converges by using the Laabe's criterion.

We shall indicate the general lines of the proof to the lemma 3.5. For $k = 2$, using the transform $z_{12}^2 = y$ and Maclaulin expansion $\log(1-y) = -\sum_{j=1}^{\infty} y^j/j$,

$$
\begin{aligned}
(3.16)\qquad L_2 &= \int_{-1}^{1} \left(1 - z_{12}^2\right)^{(n-4)/2} \log\left(1 - z_{12}^2\right) dz_{12} \\
&= \int_{0}^{1} y^{-1/2}(1-y)^{(n-4)/2} \log(1-y) dy \\
&= -B\left(\frac{1}{2}, \frac{n-2}{2}\right) \sum_{r=0}^{\infty} \frac{1}{r+1} \prod_{t=0}^{r} \frac{1+2t}{n-1+2t} \\
&= -\sqrt{\pi} \frac{\Gamma((n-2)/2)}{\Gamma((n-1)/2)} \Delta_{n1}
\end{aligned}
$$

which proves (3.13), where $B(p,q)$ designates the usual beta function. Suppose now that our relation has been proved for a certain value of k, and let us consider L_{k+1}. Using the expression

$$\begin{aligned} |\ Z_{k+1}\ | &= |\ Z_k\ | - \sum_{i,j=1}^{k} z_{ij}^* z_{i,k+1} z_{j,k+1} \\ &= |\ Z_k\ | - \boldsymbol{z}_{k+1}' Z_k^* \boldsymbol{z}_{k+1}, \end{aligned}$$

where z_{ij}^* designates the cofactor of z_{ij} in Z_k, Z_k^* is the adjugate matrix (z_{ij}^*) and $\boldsymbol{z}_{k+1} = (z_{1,k+1}, z_{2,k+1}, \cdots, z_{k,k+1})'$. Let $\lambda_1, \lambda_2, \cdots, \lambda_k$ be eigenvalues of Z_k^*, and let C be the matrix of the corresponding eigenvectors, $C = (\boldsymbol{c}_1, \boldsymbol{c}_2, \cdots, \boldsymbol{c}_k)$, and hence $C' Z_k^* C = \mathrm{Diag}(\lambda_1, \lambda_2, \cdots, \lambda_k)$. Using this matrix, make the transformation

$$C\boldsymbol{u} = \boldsymbol{z}_{k+1}, \qquad \boldsymbol{u} = (u_1, u_2, \cdots, u_k)'.$$

Then, we obtain for L_{k+1} the expression:

$$\begin{aligned} L_{k+1} &= \int_{Z_k>0} dz_{12} \cdots dz_{k-1,k} \int (|\ Z_k\ | - \boldsymbol{z}_{k+1}' Z_k^* \boldsymbol{z}_{k+1})^{(n-k-3)/2} \\ &\qquad \cdot \log(|\ Z_k\ | - \boldsymbol{z}_{k+1}' Z_k^* \boldsymbol{z}_{k+1}) dz_{1,k+1} \cdots dz_{k,k+1} \\ &\equiv \int_{Z_k>0} dz_{12} \cdots dz_{k-1,k} \cdot H_{k+1} \end{aligned}$$

where the integral with respect to the $z_{i,k+1}$ has to be extended over all values of the variables such that $\boldsymbol{z}_{k+1}' Z_k^* \boldsymbol{z}_{k+1} < |\ Z_k\ |$, and

(3.17)
$$\begin{aligned} H_{k+1} &= \int_{\sum_{i=1}^{k} \lambda_i u_i^2 < |Z_k|} \left(|\ Z_k\ | - \sum_{i=1}^{k} \lambda_i u_i^2 \right)^{(n-k-3)/2} \\ &\qquad \cdot \log \left(|\ Z_k\ | - \sum_{i=1}^{k} \lambda_i u_i^2 \right) du_1 \cdots du_k. \end{aligned}$$

Further, let us make the transformation

$$v_i = \sqrt{\frac{\lambda_i}{|\ Z_k\ |}}\, u_i, \qquad (i = 1, 2, \cdots, k).$$

Then, the Jacobian of the transformation is given by

$$\frac{| Z_k |^{k/2}}{(\lambda_1\lambda_2\cdots\lambda_k)^{1/2}} =| Z_k |^{1/2}.$$

Hence (3.17) becomes

$$(3.18)\quad H_{k+1} =| Z_k |^{(n-k-2)/2} \log | Z_k | \cdot \int_{\sum_{i=1}^k v_i^2<1} \left(1-\sum_{i=1}^k v_i^2\right)^{(n-k-3)/2} dv_1\cdots dv_k$$

$$+ | Z_k |^{(n-k-2)/2} \int_{\sum_{i=1}^k v_i^2<1} \left(1-\sum_{i=1}^k v_i^2\right)^{(n-k-3)/2} \log\left(1-\sum_{i=1}^k v_i^2\right) dv_1\cdots dv_k.$$

Now, the first term of (3.18) is equal to a well-known Dirichlet integral and the latter integral may be evaluated by the same methods as the Dirichlet integral and (3.16), and we obtain the following by the hypothesis of induction and (3.15).

$$\begin{aligned} L_{k+1} &= \pi^{k/2}\frac{\Gamma((n-k-1)/2)}{\Gamma((n-1)/2)}\left\{L_k - \left(\sum_{i=1}^k \Delta_{ni}\right) J_k\right\} \\ &= -\pi^{k(k+1)/4}\prod_{i=1}^{k+1}\frac{\Gamma((n-i)/2)}{\Gamma((n-1)/2)}\sum_{i=1}^k (k+1-i)\Delta_{ni}. \end{aligned}$$

Thus the relation holds for $k+1$, and the proof is completed.

3.4 Upper bound of the Kullback-Leibler information

Now, we shall check the condition (3.12), the validity of which implies uniform asymptotic normality of the Wishart distribution, *i.e.*, the condition (3.3) in general. By (3.10) and (3.11) we get

$$\begin{aligned} (3.19)\quad I(\boldsymbol{B}:\boldsymbol{B}^*) &= E_B\left[\log\frac{f_o(\boldsymbol{B})}{g_o(\boldsymbol{B})}\right] \\ &= E_B\left[\log\left\{C_{np}^{-1} | B |^{(n-p-1)/2} \exp\left(-\frac{1}{2}\mathrm{tr}B\right)\cdot(2\pi)^{s/2}n^{s/2}2^{p/2}\right.\right. \\ &\quad \left.\left.\cdot\exp\left(\frac{1}{2}(\boldsymbol{B}-n\boldsymbol{\eta})'(n\Omega_o)^{-1}(\boldsymbol{B}-n\boldsymbol{\eta})\right)\right\}\right] \\ &= \log C_{np}^{-1}(2\pi)^{p(p+1)/4}n^{p(p+1)/4}2^{p/2} + \frac{n-p-1}{2}E_B\left[\log | B |\right] \\ &\quad -\frac{1}{2}E_B\left[\mathrm{tr}B\right] + \frac{1}{2}E_B\left[(\boldsymbol{B}-n\boldsymbol{\eta})'(n\Omega_o)^{-1}(\boldsymbol{B}-n\boldsymbol{\eta})\right], \end{aligned}$$

where

$$C_{np} = 2^{np/2}\pi^{p(p-1)/4}\prod_{i=1}^{p}\Gamma\left(\frac{n+1-i}{2}\right).$$

We now calculate:

$$E_B\left[\log \mid B \mid\right] = C_{np}^{-1}\int_{B>0} \mid B \mid^{(n-p-1)/2} \log \mid B \mid \cdot \exp\left(-\frac{1}{2}\mathrm{tr}B\right) d\boldsymbol{B}.$$

We make the transformation $\boldsymbol{B} \to \boldsymbol{Z} = (z_{11}, \cdots, z_{1p}, z_{22}, \cdots, z_{2p}, \cdots, z_{pp})'$ such that

$$b_{ij} = z_{ij}\sqrt{b_{ii}b_{jj}} \quad (i \neq j), \qquad b_{ii} = z_{ii}.$$

Then

(3.20)
$$B = \begin{pmatrix} z_{11} & & z_{ij}\sqrt{z_{ii}z_{jj}} \\ & \ddots & \\ * & & z_{pp} \end{pmatrix}$$

$$= \begin{pmatrix} \sqrt{z_{11}} & & 0 \\ & \ddots & \\ 0 & & \sqrt{z_{pp}} \end{pmatrix} Z_p \begin{pmatrix} \sqrt{z_{11}} & & 0 \\ & \ddots & \\ 0 & & \sqrt{z_{pp}} \end{pmatrix},$$

where

$$Z_p = \begin{pmatrix} 1 & z_{12} & \cdots & z_{1p} \\ z_{21} & 1 & \cdots & z_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ z_{p1} & z_{p2} & \cdots & 1 \end{pmatrix}$$

and the Jacobian is given by $(z_{11}\cdots z_{pp})^{(p-1)/2}$. From (3.20) it follows that

$$\mid B \mid = z_{11}\cdots z_{pp} \mid Z_p \mid,$$

and consequently for the second member of the right-hand side of (3.19),

$$E_B[\log \mid B \mid] = \sum_{i=1}^{p} E_Z[\log z_{ii}] + E_Z[\log \mid Z_p \mid] \equiv I_1 + I_2, \quad \text{say.}$$

Since $z_{ii} = b_{ii}$ has a chi-square distribution of n degrees of freedom, lemma 3.2 gives us $E_Z[\log z_{ii}] = \log 2 + \Gamma'(n/2)/\Gamma(n/2)$. Hence

$$I_1 = p\left\{\log 2 + \frac{\Gamma'(n/2)}{\Gamma(n/2)}\right\}.$$

Also

$$I_2 = C_{np}^{-1} \int_{\mathcal{D}} ((z_{11} \cdots z_{pp}) \mid Z_p \mid)^{(n-p-1)/2} \log \mid Z_p \mid \cdot \exp\left(-\frac{1}{2}\sum_{i=1}^{p} z_{ii}\right) \cdot (z_{11} \cdots z_{pp})^{(p-1)/2} d\boldsymbol{Z}$$

$$= C_{np}^{-1} \int_{\mathcal{D}'} (z_{11} \cdots z_{pp})^{(n-2)/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{p} z_{ii}\right) dz_{11} \cdots dz_{pp} \times \int_{Z_p > 0} \mid Z_p \mid^{(n-p-1)/2} \log \mid Z_p \mid dz_{12} \cdots dz_{p-1,p}$$

where

$$\mathcal{D} = \{Z_p > 0,\ z_{ii} > 0 \ for\ i = 1, 2, \cdots, p\},$$

$$\mathcal{D}' = \{z_{ii} > 0 \ for\ i = 1, 2, \cdots, p\},$$

and

$$d\boldsymbol{Z} = dz_{11} dz_{12} \cdots dz_{1p} dz_{22} \cdots dz_{2p} \cdots dz_{p-1,p} dz_{pp}.$$

Here the first factor of the right-hand side is:

$$\int_{\mathcal{D}'} (z_{11} \cdots z_{pp})^{(n-2)/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{p} z_{ii}\right) dz_{11} \cdots dz_{pp}$$

$$= \prod_{i=1}^{p} \int_{z_{ii}>0} z_{ii}^{(n-2)/2} \exp\left(-\frac{1}{2} z_{ii}\right) dz_{ii}$$

$$= \prod_{i=1}^{p} 2^{n/2} \Gamma\left(\frac{n}{2}\right)$$

$$= 2^{np/2} \Gamma^{p}\left(\frac{n}{2}\right).$$

Also, by lemma 3.5,

$$\int_{Z_p > 0} \mid Z_p \mid^{(n-p-1)/2} \log \mid Z_p \mid dz_{12} \cdots dz_{p-1,p}$$

$$= -\pi^{p(p-1)/4} \left\{ \prod_{i=1}^{p} \frac{\Gamma((n+1-i)/2)}{\Gamma(n/2)} \right\} \left\{ \sum_{i=1}^{p-1} (p-i) \Delta_{n+1,i} \right\}.$$

Therefore

$$I_2 = -\sum_{i=1}^{p-1} (p-i) \Delta_{n+1,i}.$$

Summarizing these we get

$$E_B[\log \mid B \mid] = p\left\{\log 2 + \frac{\Gamma'(n/2)}{\Gamma(n/2)}\right\} - \sum_{i=1}^{p-1}(p-i)\Delta_{n+1,i}. \tag{3.21}$$

In the second place, we shall calculate for the third term of the right-hand side of (3.19):

$$E_B[\mathrm{tr}B] = E_B\left[\sum_{i=1}^{p} b_{ii}\right] = \sum_{i=1}^{p} E_B[b_{ii}] = np \tag{3.22}$$

because b_{ii} has a chi-square distribution of n degrees of freedom.

Finally, by (3.8) and (3.9)

$$\begin{aligned} & E_B\left[(\boldsymbol{B} - n\boldsymbol{\eta})'(n\Omega_o)^{-1}(\boldsymbol{B} - n\boldsymbol{\eta})\right] \\ = & \frac{1}{n}E_B\left[\sum_{i=1}^{p}\frac{(b_{ii} - n\eta_{ii})^2}{2} + \sum_{i<j}(b_{ij} - n\eta_{ij})^2\right] \\ = & \frac{1}{n}\left(np + \sum_{i<j} n\right) \\ = & \frac{p(p+1)}{2}. \end{aligned} \tag{3.23}$$

Substituting (3.21), (3.22) and (3.23) into (3.19), we obtain

$$\begin{aligned} I(\boldsymbol{B} : \boldsymbol{B}^*) = & \frac{p(p-2n+3)}{4}\log 2 + \frac{p}{2}\log \pi + \frac{p(p+1)}{4}\log n \\ & + \frac{p(p+1)}{4} - \frac{np}{2} - \sum_{i=1}^{p}\log\Gamma\left(\frac{n+1-i}{2}\right) \\ & + \frac{p(n-p-1)}{2}\left\{\log 2 + \frac{\Gamma'(n/2)}{\Gamma(n/2)}\right\} \\ & - \frac{n-p-1}{2}\sum_{i=1}^{p-1}(p-i)\Delta_{n+1,i}. \end{aligned} \tag{3.24}$$

Now we shall calculate the upper bound of the Kullback-Leibler information $I(\boldsymbol{B} : \boldsymbol{B}^*)$ for $n > p$. By lemma 3.3

$$\begin{aligned} \sum_{i=1}^{p}\log\Gamma\left(\frac{n+1-i}{2}\right) > & \sum_{i=1}^{p}\left\{\frac{1}{2}\log 2\pi + \frac{n-i}{2}\log\frac{n+1-i}{2} - \frac{n+1-i}{2}\right. \\ & \left. -\frac{1}{64}\left(\frac{n+1-i}{2}\right)^{-2}\left(\frac{n+1-i}{2}+1\right)^{-1}\right\} \end{aligned}$$

$$> \frac{p}{2}\log 2\pi + \sum_{i=1}^{p} \frac{n-i}{2} \log \frac{n+1-i}{2} + \frac{p(p-1)}{4} - \frac{np}{2} - \frac{p}{8(n+1-p)^3}.$$

Here, using the inequalities $x/(x-1) \le \log(1-x)$, $(0 \le x < 1)$, and $1+x \le (1-x)^{-1}$, $(x < 1)$, we have

$$\begin{aligned}
\sum_{i=1}^{p} \frac{n-i}{2} \log \frac{n+1-i}{2} &= \sum_{i=1}^{p} \frac{n-i}{2} \left\{ \log \frac{n}{2} + \log \left(1 - \frac{i-1}{n}\right) \right\} \\
&\ge \sum_{i=1}^{p} \frac{n-i}{2} \left(\log \frac{n}{2} - \frac{i-1}{n+1-i} \right) \\
&\ge \sum_{i=1}^{p} \frac{n-i}{2} \log \frac{n}{2} - \frac{1}{2(n+1-p)} \sum_{i=1}^{p} (n-i)(i-1) \\
&\ge \frac{p(2n-p-1)}{4} \log \frac{n}{2} - \frac{p(p+1)}{4} \left(1 - \frac{p}{n+1}\right)^{-1} \\
&\quad + \frac{p}{2} \left(1 + \frac{p-1}{n}\right) + \frac{p(p+1)(2p+1)}{12(n+1-p)}.
\end{aligned}$$

Therefore

$$\sum_{i=1}^{p} \log \Gamma \left(\frac{n+1-i}{2} \right) > \frac{p}{2} \log 2\pi + \frac{p(2n-p-1)}{4} \log \frac{n}{2} - \frac{p(p^2-1)}{12(n+1-p)} + \frac{p(p-1)}{2n} - \frac{np}{2} - \frac{p}{8(n+1-p)^3}. \tag{3.25}$$

By lemma 3.4

$$\log 2 + \frac{\Gamma'(n/2)}{\Gamma(n/2)} < \log n - \frac{1}{n} - \frac{1}{18n(n-2)}, \tag{3.26}$$

and it is clear that

$$\Delta_{n+1,i} > \frac{1}{n}. \tag{3.27}$$

Substituting (3.25), (3.26) and (3.27) into (3.24), we obtain

$$I(\boldsymbol{B} : \boldsymbol{B}^*) < \frac{p(p^2+3)}{4n} + \frac{p(p^2-1)}{12(n+1-p)} + \frac{p(p+1)}{36n(n-2)} - \frac{p}{36(n-2)} + \frac{p}{8(n+1-p)^3}$$

which tends to zero as $p^3/n \to 0$ or equivalently $p^3/N \to 0$.

Thus, we have proved the following theorem.

Theorem 3.1 *The Wishart distribution (3.1) is asymptotically equivalent in the sense of type* $(\mathcal{B})_d$ *to the* $s = p(p+1)/2$*-dimensional normal variable* $\boldsymbol{A}^*$ *with mean vector* $n\boldsymbol{\lambda}$ *and variance-covariance matrix* $n\boldsymbol{\Omega}$, *where*

$$\boldsymbol{\lambda} = (\lambda_{11}, \lambda_{12}, \cdots, \lambda_{1p}, \lambda_{22}, \cdots, \lambda_{2p}, \cdots, \lambda_{pp})',$$

and

$$\Omega = \begin{pmatrix} 2\lambda_{11}^2 & & & \lambda_{ik}\lambda_{jl} + \lambda_{il}\lambda_{jk} \\ & \lambda_{12}^2 + \lambda_{11}\lambda_{22} & & \\ & & \ddots & \\ * & & & 2\lambda_{pp}^2 \end{pmatrix}_{s\times s}$$

provided that $p^3/n \to 0$ *as* $n \to \infty$. *An upper bound to the uniform error (3.3) is given by*

$$\begin{aligned} \delta_d(\boldsymbol{A}, \boldsymbol{A}^* : \mathcal{B}_{(s)}) \leq & \left\{ \frac{p(p^2+3)}{8n} + \frac{p(p^2-1)}{24(n+1-p)} + \frac{p(p+1)}{72n(n-2)} \right. \\ & \left. - \frac{p}{72(n-2)} + \frac{p}{16(n+1-p)^3} \right\}^{1/2}. \end{aligned}$$

Remark. The condition $p^3/n \to 0$ is the best possible for which the Kullback-Leibler information converges to zero. This can be shown in the following:

$$\begin{aligned} (3.28) \quad \sum_{i=1}^{p-1}(p-i)\Delta_{n+1,i} &= \sum_{i=1}^{p-1}(p-i)\sum_{r=0}^{\infty}\frac{1}{r+1}\prod_{t=0}^{r}\frac{1+2t}{n+1-i+2t} \\ &= \frac{1}{n}\sum_{i=1}^{p-1}(p-i)\left(1-\frac{i-1}{n}\right)^{-1}\left\{1+\frac{3}{2n}\left(1-\frac{i-3}{n}\right)^{-1}\right. \\ &\quad \left. + \frac{5}{n^2}\left(1-\frac{i-3}{n}\right)^{-1}\left(1-\frac{i-5}{n}\right)^{-1}\right\} + O\left(\frac{p^2}{n^4}\right) \\ &= \frac{p(p-1)}{2n} + \frac{p(2p^2+3p-5)}{12n^2} + O\left(\frac{p^4}{n^3}\right). \end{aligned}$$

By Stirling's formula and Taylor's expansion,

$$\begin{aligned} (3.29) \quad \sum_{i=1}^{p}\log\Gamma\left(\frac{n+1-i}{2}\right) &\sim \sum_{i=1}^{p}\left\{\frac{n-i}{2}\log\frac{n+1-i}{2} - \frac{n+1-i}{2} + \frac{1}{2}\log 2\pi\right. \\ &\quad \left. + \frac{1}{6(n+1-i)} - \frac{1}{90(n+1-i)^2}\right\} + O\left(\frac{p}{n^3}\right) \\ &= \frac{p(2p^2+3p-1)}{24n} + \frac{p(2n-p-1)}{4}\log\frac{n}{2} - \frac{np}{2} \\ &\quad + \frac{p}{2}\log 2\pi + O\left(\frac{p^4}{n^2}\right), \end{aligned}$$

and also

$$\log 2 + \frac{\Gamma'(n/2)}{\Gamma(n/2)} \sim \log n - \frac{1}{n} - \frac{1}{3n^2} + O\left(\frac{1}{n^4}\right). \tag{3.30}$$

Substituting (3.28), (3.29) and (3.30) into (3.24), we obtain

$$I(\boldsymbol{B} : \boldsymbol{B}^*) \sim \frac{p^3}{12n} + \frac{p^2}{4n} + \frac{p}{3n} + O\left(\frac{p^4}{n^2}\right).$$

Now, let φ be a measurable mapping from $(\boldsymbol{R}_{(s)}, \boldsymbol{\mathcal{B}}_{(s)})$ to another measurable space $(\boldsymbol{R}_{(t)}, \boldsymbol{\mathcal{B}}_{(t)})$. Then we have the following corollary by the lemma 3.1 and the theorem above.

Corollary 3.1 *Let* $\boldsymbol{A}$ *and* $\boldsymbol{A}^*$ *be the same situation as in the theorem above. Then, it holds that*

$$\varphi(\boldsymbol{A}) \sim \varphi(\boldsymbol{A}^*)\,(\boldsymbol{\mathcal{B}})_d, \quad (n \to \infty).$$

Chapter 4

Uniform Asymptotic Normality of the Dirichlet Distribution

4.1 Introduction

The theory of the uniform asymptotic equivalence of probability distributions has been developed by Ikeda ([11], [12]). However, so far as the author knows, applying the uniform approximation theory to concrete multivariate distributions has not been done sufficiently. In Chapter 3, the author proved the uniform (or type $(\mathcal{B})_d$) asymptotic normality of the Wishart distribution by giving an upper bound of the uniform error based on the Kullback-Leibler information. In this chapter, we shall consider the same problem for the Dirichlet distribution which is often used to calculate quantity connected with order statistics.

A k-dimensional continuous distribution with probability density function

$$(4.1)\qquad f(\boldsymbol{x}_{(k)}) = \frac{\Gamma(\alpha)}{\prod_{i=1}^{k}\Gamma(\alpha_i)\cdot\Gamma(\alpha-\sum_{i=1}^{k}\alpha_i)}\prod_{i=1}^{k}x_i^{\alpha_i-1}\cdot\left(1-\sum_{i=1}^{k}x_i\right)^{\alpha-\sum_{i=1}^{k}\alpha_i-1}$$

on the domain $D = \{\boldsymbol{x}_{(k)} = (x_1, x_2, \cdots, x_k); x_i > 0, \sum_{i=1}^{k} x_i < 1\}$, where $\alpha = \sum_{i=1}^{k+1}\alpha_i$, $\alpha_i > 0$ $(i = 1, 2, \cdots, k+1)$ are parameters, is called a Dirichlet distribution. In case $k = 1$, this becomes a beta distribution. It is straightforward to show that the marginal distribution X_i is given by beta distribution

$$(4.2)\qquad f_i(x_i) = \frac{\Gamma(\alpha)}{\Gamma(\alpha_i)\Gamma(\alpha-\alpha_i)}x_i^{\alpha_i-1}(1-x_i)^{\alpha-\alpha_i-1},\ (0 < x < 1),$$

for $i = 1, 2, \cdots, k, (k+1)$. Mean vector and variance-covariance matrix of $\boldsymbol{X}_{(k)}$ are given by

$$E(\boldsymbol{X}_{(k)}) = \left(\frac{\alpha_1}{\alpha}, \frac{\alpha_2}{\alpha}, \cdots, \frac{\alpha_k}{\alpha}\right) \equiv \boldsymbol{\mu}_{(k)}$$

and

$$(4.3) \quad Var(\boldsymbol{X}_{(k)}) = \frac{1}{\alpha^2(\alpha+1)} \begin{pmatrix} \alpha_1(\alpha-\alpha_1) & -\alpha_1\alpha_2 & \cdots & -\alpha_1\alpha_k \\ -\alpha_1\alpha_2 & \alpha_2(\alpha-\alpha_2) & \cdots & -\alpha_2\alpha_k \\ \vdots & \vdots & \ddots & \vdots \\ -\alpha_1\alpha_k & -\alpha_2\alpha_k & \cdots & \alpha_k(\alpha-\alpha_k) \end{pmatrix} \equiv \Sigma_{(k)}.$$

We investigate the uniform asymptotic normality of $\boldsymbol{X}_{(k)}$ under the following conditions

case (1) $\alpha_1, \alpha_2, \cdots, \alpha_{k+1} \to \infty$, k is fixed,

case (2) $k \to \infty$, $\alpha_1, \alpha_2, \cdots, \alpha_{k+1} \to \infty$.

The proof of the assertion above is given by checking the convergence of the Kullback-Leibler information. Throughout the present chapter, lemma 3.3 and lemma 3.4 in Chapter 3 play fundamental roles for the calculations of the information.

Let f and g be the probability density functions of $\boldsymbol{X}_{(k)}$ and $\boldsymbol{Y}_{(k)}$ respectively. The Kullback-Leibler information is defined by

$$I(\boldsymbol{X}_{(k)} : \boldsymbol{Y}_{(k)}) = \int_{\boldsymbol{R}_{(k)}} f \log \frac{f}{g} d\boldsymbol{x}_{(k)}.$$

Also we define

$$(4.4) \quad \delta_d(\boldsymbol{X}_{(k)}, \boldsymbol{Y}_{(k)} : \mathcal{B}_{(k)}) = \sup_{E \in \mathcal{B}_{(k)}} | P^{X_{(k)}}(E) - P^{Y_{(k)}}(E) | = \frac{1}{2} \int_{\boldsymbol{R}_{(k)}} | f - g | \, d\boldsymbol{x}_{(k)},$$

where $\mathcal{B}_{(k)}$ is the usual Borel field in the k-dimensional Euclidean space $\boldsymbol{R}_{(k)}$. The error estimation is given by

$$\delta_d(\boldsymbol{X}_{(k)}, \boldsymbol{Y}_{(k)} : \mathcal{B}_{(k)}) \leq \sqrt{I(\boldsymbol{X}_{(k)} : \boldsymbol{Y}_{(k)})/2}.$$

4.2 Upper bound of the Kullback-Leibler information

Let $\boldsymbol{X}_{(k)}$ and $\boldsymbol{Y}_{(k)}$ be k-dimensional random vector, whose probability density functions are the Dirichlet distribution (4.1) and the normal density $g(\boldsymbol{y}_{(k)})$ with same mean vector and variance-covariance matrix as (4.1), respectively. Thus,

$$(4.5) \qquad g(\boldsymbol{y}_{(k)}) = (2\pi)^{-k/2} \mid \Sigma_{(k)} \mid^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{y}_{(k)} - \boldsymbol{\mu}_{(k)})\Sigma_{(k)}^{-1}(\boldsymbol{y}_{(k)} - \boldsymbol{\mu}_{(k)})'\right\}.$$

In order to investigate the uniform(or type $(\mathcal{B})_d$) asymptotic normality of the Dirichlet distribution (4.1), we try to find conditions under which the Kullback-Leibler information converges to zero.

By (4.1) and (4.5), the Kullback-Leibler information $I(\boldsymbol{X}_{(k)} : \boldsymbol{Y}_{(k)})$ becomes

$$\begin{aligned}(4.6) \quad I(\boldsymbol{X}_{(k)} : \boldsymbol{Y}_{(k)}) &= E_{X_{(k)}}\left[\log \frac{f(\boldsymbol{X}_{(k)})}{g(\boldsymbol{X}_{(k)})}\right] \\ &= \log \frac{\Gamma(\alpha)(2\pi)^{k/2} \mid \Sigma_{(k)} \mid^{1/2}}{\prod_{i=1}^k \Gamma(\alpha_i) \cdot \Gamma\left(\alpha - \sum_{i=1}^k \alpha_i\right)} + \sum_{i=1}^k (\alpha_i - 1) E_{X_{(k)}}[\log X_i] \\ &\quad + \left(\alpha - \sum_{i=1}^k \alpha_i - 1\right) E_{X_{(k)}}\left[\log\left(1 - \sum_{i=1}^k X_i\right)\right] \\ &\quad + \frac{1}{2} E_{X_{(k)}}\left[(\boldsymbol{X}_{(k)} - \boldsymbol{\mu}_{(k)})\Sigma_{(k)}^{-1}(\boldsymbol{X}_{(k)} - \boldsymbol{\mu}_{(k)})'\right].\end{aligned}$$

By (4.3), it is easy to verify that

$$(4.7) \qquad \mid \Sigma_{(k)} \mid = \frac{1}{\alpha^{k+1}(\alpha+1)^k} \prod_{i=1}^k \alpha_i \cdot \left(\alpha - \sum_{i=1}^k \alpha_i\right).$$

As was seen in (4.2), under $\boldsymbol{X}_{(k)}$, the variable $1 - \sum_{i=1}^k X_i = X_{k+1}$ is distributed as the beta distribution $B(\alpha - \sum_{i=1}^k \alpha_i, \sum_{i=1}^k \alpha_i)$.

For the exact calculation of this information, we shall prepare a formula of integral calculus.

Lemma 4.1 *For positive real numbers p and q,*

$$\int_0^1 x^{p-1}(1-x)^{q-1} \log(1-x)dx = -\frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}\{\psi(p+q) - \psi(q)\}$$

where $\psi(x) = d\log\Gamma(x)/dx$.

For the third member of the right-hand side of (4.6), by using the lemma, we obtain

$$E_{X_{(k)}}\left[\log\left(1-\sum_{i=1}^{k}X_i\right)\right] = -\left\{\psi(\alpha)-\psi\left(\alpha-\sum_{i=1}^{k}\alpha_i\right)\right\} \tag{4.8}$$

and for the second term we have

$$E_{X_{(k)}}[\log X_i] = -\{\psi(\alpha)-\psi(\alpha_i)\}. \tag{4.9}$$

We also have

$$E_{X_{(k)}}\left[(\boldsymbol{X}_{(k)}-\boldsymbol{\mu}_{(k)})\Sigma_{(k)}^{-1}(\boldsymbol{X}_{(k)}-\boldsymbol{\mu}_{(k)})'\right] = k. \tag{4.10}$$

Substituting (4.7), (4.8), (4.9) and (4.10) into (4.6) we obtain

$$\begin{aligned} I(\boldsymbol{X}_{(k)}:\boldsymbol{Y}_{(k)}) &= \frac{k}{2}\log 2\pi + \log\Gamma(\alpha) + \frac{1}{2}\log\frac{1}{\alpha^{k+1}(\alpha+1)^k}\prod_{i=1}^{k+1}\alpha_i \\ &\quad -\sum_{i=1}^{k+1}\log\Gamma(\alpha_i) - \sum_{i=1}^{k+1}(\alpha_i-1)\{\psi(\alpha)-\psi(\alpha_i)\} + \frac{k}{2}. \end{aligned} \tag{4.11}$$

Now we shall calculate an upper bound of the Kullback-Leibler information $I(\boldsymbol{X}_{(k)} : \boldsymbol{Y}_{(k)})$ for $\alpha_i > 1$ $(i = 1, 2, \cdots, k+1)$. By lemma 3.3 in Chapter 3, we have

$$\log\Gamma(\alpha) < \frac{1}{2}\log 2\pi + \left(\alpha-\frac{1}{2}\right)\log\alpha - \alpha, \tag{4.12}$$

$$\sum_{i=1}^{k+1}\log\Gamma(\alpha_i) > \frac{k+1}{2}\log 2\pi + \sum_{i=1}^{k+1}\left(\alpha_i-\frac{1}{2}\right)\log\alpha_i - \alpha - \sum_{i=1}^{k+1}\frac{1}{64\alpha_i^2(\alpha_i+1)}. \tag{4.13}$$

By lemma 3.4 in Chapter 3, we also have

$$\begin{aligned} \sum_{i=1}^{k+1}(\alpha_i-1)\{\psi(\alpha)-\psi(\alpha_i)\} &> \sum_{i=1}^{k+1}(\alpha_i-1)(\log\alpha-\log\alpha_i) - \frac{35}{72}\sum_{i=1}^{k+1}\frac{1}{\alpha_i} \\ &\quad + \frac{k(6\alpha-5)}{12\alpha(\alpha-1)} + \frac{5}{12\alpha} + \frac{k}{2}. \end{aligned} \tag{4.14}$$

Substituting (4.12), (4.13) and (4.14) into (4.11), and using the inequality $2/(2x+1) \le \log(1+1/x)$, $(x>0)$, we have

$$I(\boldsymbol{X}_{(k)}:\boldsymbol{Y}_{(k)}) \le \frac{k}{2}\log\frac{\alpha}{\alpha+1} + \frac{35}{72}\sum_{i=1}^{k+1}\frac{1}{\alpha_i} - \frac{k(6\alpha-5)}{12\alpha(\alpha-1)} \tag{4.15}$$

$$
\begin{aligned}
&-\frac{5}{12\alpha}+\sum_{i=1}^{k+1}\frac{1}{64\alpha_i^2(\alpha_i+1)} \\
\leq\ & \frac{35}{72}\sum_{i=1}^{k+1}\frac{1}{\alpha_i}-\frac{k(24\alpha^2-16\alpha-5)}{12\alpha(\alpha-1)(2\alpha+1)}-\frac{5}{12\alpha} \\
&+\sum_{i=1}^{k+1}\frac{1}{64\alpha_i^2(\alpha_i+1)} \quad (\equiv \overline{I}(\boldsymbol{X}_{(k)}:\boldsymbol{Y}_{(k)})).
\end{aligned}
$$

Thus, if $k\to\infty$ (or k is fixed), $\alpha_1,\alpha_2,\cdots,\alpha_{k+1}\to\infty$, then the value above $\overline{I}(\boldsymbol{X}_{(k)}:\boldsymbol{Y}_{(k)})$ tends to zero provided that $\sum_{i=1}^{k+1}1/\alpha_i\to 0$ (hence $k/\alpha\to 0$).

Thus, we have proved the following theorem.

Theorem 4.1 *Under the limiting $k\to\infty$ (or k is fixed), $\alpha_1,\alpha_2,\cdots,\alpha_{k+1}\to\infty$, the Dirichlet distribution (4.1) is asymptotically normal, uniformly over the Borel field, provided that $\sum_{i=1}^{k+1}1/\alpha_i\to 0$. An upper bound of the uniform error (4.4) is given by*

$$
\begin{aligned}
\delta_d(\boldsymbol{X}_{(k)},\boldsymbol{Y}_{(k)}:\boldsymbol{\mathcal{B}}_{(k)}) \leq\ & \left\{\frac{35}{144}\sum_{i=1}^{k+1}\frac{1}{\alpha_i}-\frac{k(24\alpha^2-16\alpha-5)}{24\alpha(\alpha-1)(2\alpha+1)}-\frac{5}{24\alpha}\right. \\
&\left.+\sum_{i=1}^{k+1}\frac{1}{128\alpha_i^2(\alpha_i+1)}\right\}^{1/2}.
\end{aligned}
$$

The following corollary is immediate from asymptotically equivalent $(\boldsymbol{\mathcal{B}})_d$ property.

Corollary 4.1 (i) *For any measurable transformation t from $\boldsymbol{R}_{(k)}$ into $\boldsymbol{R}_{(m)}$, $m\leq k$, the transformd variable*

$$
t(\boldsymbol{X}_{(k)})=\boldsymbol{Z}_{(k)}
$$

is asymptotically equivalent $(\boldsymbol{\mathcal{B}})_d$ to $t(\boldsymbol{Y}_{(k)})=\boldsymbol{U}_{(k)}$ under the condition in the theorem above, where t may be dependent on the parameters involved.

(ii) *In particular, if t is a linear transformation and*

$$
\boldsymbol{Z}_{(k)}=t(\boldsymbol{X}_{(k)})=\boldsymbol{X}_{(k)}A_{(k)},
$$

then $\boldsymbol{Z}_{(k)}$ is asymptotically equivalent $(\boldsymbol{\mathcal{B}})_d$ to the normal $N(\boldsymbol{\nu}_{(k)},\Lambda_{(k)})$ with $\boldsymbol{\nu}_{(k)}=\boldsymbol{\mu}_{(k)}A_{(k)}$ and $\Lambda_{(k)}=A'_{(k)}\Sigma_{(k)}A_{(k)}$, where of course $A_{(k)}$ may be dependent on the parameters.

Some values of $\overline{I}(\boldsymbol{X}_{(k)} : \boldsymbol{Y}_{(k)})$ in case (1) (*i.e.*, k is fixed) are tabulated in the following table 4.1. Similarly, in case (2) (*i.e.*, $k \to \infty$), the values of $\overline{I}(\boldsymbol{X}_{(k)} : \boldsymbol{Y}_{(k)})$ for some values of k and α_i where $\sum_{i=1}^{k+1} 1/\alpha_i$ tends to zero are tabulated in table 4.2.

Table 4.1.

Values of $\overline{I}(\boldsymbol{X}_{(k)} : \boldsymbol{Y}_{(k)})$ for $\delta(n) = (\alpha_1, \alpha_2, \cdots, \alpha_{k+1}) = (n, n, \cdots, n)$.

$k \setminus \delta(n)$	$\delta(10)$	$\delta(20)$	$\delta(50)$	$\delta(100)$	$\delta(500)$
1	0.0268	0.0133	0.0053	0.0026	0.0005
2	0.0657	0.0327	0.0131	0.0065	0.0013
3	0.1094	0.0546	0.0218	0.0109	0.0022
4	0.1551	0.0774	0.0310	0.0155	0.0031
5	0.2017	0.1008	0.0403	0.0201	0.0040
10	0.4403	0.2201	0.0880	0.0440	0.0088
20	0.9240	0.4619	0.1847	0.0924	0.0185

Table 4.2.

Values of $\overline{I}(\boldsymbol{X}_{(k)} : \boldsymbol{Y}_{(k)})$ for $\delta(k) = (\alpha_1, \alpha_2, \cdots, \alpha_{k+1}) =$ $((6/5)^k, (6/5)^{2k}, \cdots, (6/5)^{(k+1)k})$.

$\delta(3)$	$\delta(4)$	$\delta(5)$	$\delta(6)$	$\delta(7)$	$\delta(8)$
0.4148	0.3812	0.3123	0.2431	0.1883	0.1475

Acknowledgements

The author would like to express his sincere thanks to Professor S. Shirahata of Osaka University for his warm support, kind encouragement and for reading the manuscript.

The author is also grateful to Professor N. Inagaki, Professor M. Goto and the members of Department of Applied Mathematics, Faculty of Engineering Science, Osaka University, for their warm hospitality at the seminar.

The author would like to express his sincere thanks to Professor S. Ikeda of Soka University for his kind guidance, many useful suggestions and for constant encouragement throughout the present work.

Finally, the author would like to thanks Professor T. Matsunawa of The Institute of Statistical Mathematics for many helpful comments and for useful discussions and Professor H. Nishida of Kanazawa Gakuin University for his kind encouragement.

Bibliography

[1] Y. Akimoto, On the stable vector bundles of rank 2 on P_C^3, *TRU Math.*, **16**(1980), 1–12.

[2] Y. Akimoto, A note on the implication relation between two types of asymptotic equivalence, *TRU Math.*, **24**(1988), 65–73.

[3] Y. Akimoto, The uniform asymptotic normality of the Wishart distribution based on the K-L information, *J. Japan Statist. Soc.*, **24**(1994), 73–81.

[4] Y. Akimoto, A note on uniform asymptotic normality of Dirichlet distribution, *Math. Japonica*, **44**(1996), 25–30.

[5] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis* (2nd ed.), John Wiley, 1984.

[6] P. Billingsley, *Convergence of Probability Measures*, John Wiley, 1968.

[7] H. Cramér, *Mathematical Methods of Statistics*, Princeton Univ. Press, Princeton, 1946.

[8] P. R. Halmos, *Measure Theory*, van Nostrand, 1950.

[9] S. Ikeda, Necessary conditions for the convergence of Kullback-Leiblers' mean informations, *Ann. Inst. Statist. Math.*, **14**(1962), 107–118.

[10] S. Ikeda, On characterization of the Kullback-Leibler mean information for continuous probability distributions, *Ann. Inst. Statist. Math.*, **14**(1962), 73–79.

[11] S. Ikeda, Asymptotic equivalence of probability distributions with applications to some problems of asymptotic independence, *Ann. Inst. Statist. Math.*, **15**(1963), 87–116.

[12] S. Ikeda, Asymptotic equivalence of real probability distributions, *Ann. Inst. Statist. Math.*, **20**(1968), 339–362.

[13] S. Ikeda, Some criteria for uniform asymptotic equivalence of real probability distributions, *Ann. Inst. Statist. Math.*, **27**(1975), 421–428.

[14] S. Ikeda and T. Matsunawa, Asymptotic equivalence of probability distributions and its statistical applications, Part I, General Theory, *Research Memorandum* No. **222**(1982), Inst. Statist. Math., Tokyo.

[15] S. Ikeda and T. Matsunawa, Asymptotic equivalence of probability distributions and its statistical applications, Part II, Theory to Real Case, *Research Memorandum* No. **231**(1982), Inst. Statist. Math., Tokyo.

[16] S. Ikeda and T. Matsunawa, Asymptotic equivalence of probability distributions and its statistical applications, Part III, Applications, *Research Memorandum* No. **227**(1982), Inst. Statist. Math., Tokyo.

[17] N. L. Johnson and S. Kotz, *Distributions in Statistics: Continuous Multivariate Distributions*, John Wiley, 1972.

[18] O. Kraft and N. Schmitz, A note on Hoeffding's inequality, *J. Amer. Statist. Ass.*, **64**(1969), 907–912.

[19] S. Kullback, *Information Theory and Statistics*, John Wiley, 1959.

[20] T. Matsunawa, Some inequalities based on inverse factorial series, *Ann. Inst. Statist. Math.*, **28**(1976), 291–305.

[21] T. Matsunawa, On the Stirling formula, *Proc. Inst. Statist. Math.*, **25**(1978), 107–116 (in Japanese).

[22] Y. Nonaka, A note on asymptotic normal approximation to Wishart distribution - I(Canonical case), *Technical Report* No. **16**(1982), Inst. Inform. Sci., Soka Univ.

[23] J. Ogawa and S. Ikeda, On the asymptotic null distribution of the F-statistic for testing a partial null hypothesis in a randomized $PBIB$ design with m associate classes under the Neyman model, *Canad. J. Statist.*, **1**(1973), 1-23.

[24] T. Ono and Y. Akimoto, A space curve of genus 7 and degree 8 whose normal bundle is semi-stable, *TRU Math.*, **22**(1986), 31–37.

[25] R.R. Rao, Relations between weak and uniform convergence of measures with applications, *Ann. Math. Statist.*, **33**(1962), 659–680.

[26] J. Wishart, The generalized product moment distribution in samples from a normal multivariate population, *Biometrika*, **20A**(1928), 32–52.