EDGEWORTH EXPANSION FOR TWO-SAMPLE U-STATISTICS

EDGEWORTH EXPANSION FOR TWO-SAMPLE U-STATISTICS*

Yoshihiko Maesono**

(Received Sep. 10, 1985)

Abstract

Formal Edgeworth expansion with remainder term $o(N^{-1})$ is established for two-sample U-statistics. And the conditions which ensure the validity of the expansion are also discussed.

1. Introduction

Let X_1, X_2, $\cdots$, X_m and Y_1, Y_2, $\cdots$, Y_n be independent random samples from distributions with c. d. f.'s(cumulative distribution functions) F(x) and G(y), respectively. Let $h(x_1, \cdots, x_r; y_1, \cdots, y_s)$ be symmetric in its x_i components and separately symmetric in its y_j components, satisfying $E[h(X_1, \cdots, X_r; Y_1, \cdots, Y_s)]=0$ with $r \leqq m$ and $s \leqq n$. h is called a kernel and (r, s) are called its degree. We shall define a two-sample U-statistic with a kernel of degree (r, s), h, by

$$U_{m,n} = \binom{m}{r}^{-1} \binom{n}{s}^{-1} \sum_{Cm,r} \sum_{Cn,s} h(X_{i_1}, \cdots, X_{i_r}; Y_{j_1}, \cdots, Y_{j_s})$$

where $\sum_{Cm,r}$ indicates that the summation is over $1 \leqq i_1 < \cdots < i_r \leqq n$.

In this paper, putting $N = m+n$, we shall discuss an asymptotic expansion under the assumption

(A) $$0 < \lambda = \lim_{N \to \infty} m/N < 1.$$

This assumption means that $m = 0(N)$ and $n = 0(N)$.

Callaert, Janssen and Veraverbeke[1] have obtained the asymptotic expansion of one-sample U-statistic with a kernel of degree two. And Maesono[6] has obtained it with a kernel of arbitrary degree.

In Section 2 we state a representation for $U_{m,n}$ in terms of forward martingales and get the bounds of absolute moments of martingales. In Section 3, using the martingale representation of $U_{m,n}$, we obtain formal Edgeworth expansion of $U_{m,n}$ with remainder term $o(N^{-1})$. In Section 4, we discuss the conditions for the valid expansion.

2. Preliminaries

We shall represent $U_{m,n}$ in terms of forward martingales. Hoeffding[4] (cf.

Serfling[7] p178) has obtained the similar representation for one-sample U-statistics. Under the assumption $E|h(X_1, \cdots, X_r; Y_1, \cdots, Y_s)|<\infty$, let us define the following notations:

for $0\leqq a\leqq r$ and $0\leqq b\leqq s$

$$w_{a,b}(x_1, \cdots, x_a; y_1, \cdots, y_b)$$
$$=E[h(X_1, \cdots, X_r; Y_1, \cdots Y_s)| X_1=x_1, \cdots, X_a=x_a, Y_1=y_1, \cdots, Y_b=y_b],$$

$$g_{0,0}=0, \quad g_{1,0}(x_1)=w_{1,0}(x_1), \quad g_{0,1}(y_1)=w_{0,1}(y_1),$$

$$g_{2,0}(x_1, x_2)=w_{2,0}(x_1, x_2)-\sum_{i=1}^{2}g_{1,0}(x_i), \quad g_{1,1}(x_1; y_1)=w_{1,1}(x_1; y_1)-g_{1,0}(x_1)-g_{0,1}(y_1),$$

$$g_{0,2}(y_1, y_2)=w_{0,2}(y_1, y_2)-\sum_{j=1}^{2}g_{0,1}(y_j),$$
$$\vdots$$

$$g_{r,s}(x_1, \cdots, x_r; y_1, \cdots, y_s)=w_{r,s}(x_1, \cdots, x_r; y_1, \cdots, y_s)$$
$$-\sum_{b=0}^{s-1}\sum_{Cs,b}g_{r,b}(x_1, \cdots, x_r; y_{j_1}, \cdots, y_{j_b})-\sum_{a=0}^{r-1}\sum_{Cr,a}g_{a,s}(x_{i_1}, \cdots, x_{i_a}; y_1, \cdots, y_s)$$
$$-\sum_{a=0}^{r-1}\sum_{b=0}^{s-1}\sum_{Cr,a}\sum_{Cs,b}g_{a,b}(x_{i_1}, \cdots, x_{i_a}; y_{j_1}, \cdots, y_{j_b}),$$

for $0\leqq a\leqq r$ and $0\leqq b\leqq s$
$$A_{a,b}=\sum_{Cm,a}\sum_{Cn,b}g_{a,b}(X_{i_1}, \cdots, X_{i_a}; Y_{j_1}, \cdots, Y_{j_b}).$$

Then $U_{m,n}$ can be rewritten as

$$U_{m,n}=\binom{m}{r}^{-1}\binom{n}{s}^{-1}\sum_{a=0}^{r}\sum_{b=0}^{s}\binom{m-a}{r-a}\binom{n-b}{s-b}A_{a,b}.$$

It is shown in the proof of Lemma 2.3 that $A_{a,b}$ is a forward martingale for each a and b $(a=0, 1, \cdots, r; b=0, 1, \cdots, s)$.

By the definition of $g_{a,b}$, Lemma 2.1 follows.

LEMMA 2.1. *Assume that* $E|h(X_1, \cdots, X_r; Y_1, \cdots, Y_s)|<\infty$. *If one of* $\{i_1, \cdots, i_a\}$ *is not contained in* $\{p_1, \cdots, p_c\}$, *or one of* $\{j_1, \cdots, j_b\}$ *is not contained in* $\{q_1, \cdots, q_d\}$, *then*

(2.1) $$E[g_{a,b}(X_{i_1}, \cdots, X_{i_a}; Y_{j_1}, \cdots, Y_{j_b})| X_{p_1}, \cdots, X_{p_c}; Y_{q_1}, \cdots, Y_{q_d}]=0.$$

Proof. By double induction on a and b we can prove (2.1) directly.

Using Lemma 2.1, we can prove the useful two lemmas.

LEMMA 2.2. *Assume the assumptions in Lemma 2.1. Then for any function f which satisfies* $E|fg_{a,b}|<\infty$, *we have*

(2.2) $$E[f(X_{p_1}, \cdots, X_{p_c}; Y_{q_1}, \cdots, Y_{q_d})g_{a,b}(X_{i_1}, \cdots, X_{i_a}; Y_{j_1}, \cdots, Y_{j_d})]=0.$$

Proof. Taking the conditional expectation, we have the desired result from (2.1).

Before describing the next lemma, we prepare notations. For $1\leqq m_1<\cdots<m_a\leqq m$, $1\leqq n_1<\cdots<n_b\leqq n$, $0\leqq a\leqq r$ and $0\leqq b\leqq s$, let us define

$$B_{a,\,b}(m_1,\,\cdots,\,m_a\,;\,n_1,\,\cdots,\,n_b)$$

$$=\sum_{i_1=1}^{m_1}\cdots\sum_{i_a=i_{a-1}+1}^{m_a}\sum_{j_1=1}^{n_1}\cdots\sum_{j_b=j_{b-1}+1}^{n_b}g_{a,\,b}(X_{i_1},\,\cdots,\,X_{i_a}\,;\,Y_{j_1},\,\cdots,\,Y_{j_b}).$$

Then we have the upper bound of the pth absolute moment of $B_{a,\,b}$.

LEMMA 2.3. *Given the existence of the pth $(p\geqq 2)$ absolute moment of kernel h, there exist a positive constant C such that*

$$(2.3) \qquad \mathrm{E}\,|\,B_{a,\,b}(m_1,\,\cdots,\,m_a\,;\,n_1,\,\cdots,\,n_b)\,|^p\leqq C(\prod_{i=1}^{a}m_i\prod_{j=1}^{b}n_j)^{\frac{p}{2}}.$$

If the second moment of kernel h is finite, the inequality (2.3) also holds with $p=1$.

Proof. The latter part of the lemma immediately follows from the former. Therefore we consider the case $p\geqq 2$. By double induction on c and d, we shall prove the following inequality : for $0\leqq c\leqq a$, $0\leqq d\leqq b$, $1\leqq u_1<\cdots<u_c<i_{c+1},\,\cdots,\,i_a$, and $1\leqq q_1<\cdots<q_d<j_{d+1},\,\cdots,\,j_b$,

$$\mathrm{E}\,|\,\sum_{i_1=1}^{u_1}\cdots\sum_{i_c=i_{c-1}+1}^{u_c}\sum_{j_1=1}^{q_1}\cdots\sum_{j_d=j_{d-1}+1}^{q_d}$$

$$(2.4) \qquad g_{a,\,b}(X_{i_1},\,\cdots,\,X_{i_c},\,\cdots,\,X_{i_a}\,;\,Y_{j_1},\,\cdots,\,Y_{j_d},\,\cdots,\,Y_{j_b})\,|^p$$

$$\leqq (C_p)^{c+d}\mathrm{E}\,|\,g_{a,\,b}(X_1,\,\cdots,\,X_a\,;\,Y_1,\,\cdots,\,Y_b)\,|^p\;(\prod_{i=1}^{c}u_i\prod_{j=1}^{d}q_j)^{\frac{p}{2}}$$

where $C_p=\{8(p-1)\max(1,\,2^{p-3})\}^p$.

When $c=1$ and $d=0$, let $Z_k=\sum_{i_1=1}^{k}g_{a,\,b}(X_{i_1},\,\cdots,\,X_{i_a}\,;\,Y_{j_1},\,\cdots,\,Y_{j_b})$ for $k=1,\,2,\,\cdots,\,u_1$. Then we have $Z_k-Z_{k-1}=g_{a,\,b}(X_k,\,X_{i_2},\,\cdots,\,X_{i_a}\,;\,Y_{j_1},\,\cdots,\,Y_{j_b})$ and $k<i_2,\,\cdots,\,i_a$, where $Z_0=0$. Since $Z_1,\,\cdots,\,Z_{k-1}$ are functions of $X_1,\,\cdots,\,X_{k-1},\,X_{i_2},\,\cdots,\,X_{i_a}$ and $Y_{j_1},\,\cdots,\,Y_{j_b}$, we find from lemma 2.1 that

$$\mathrm{E}[Z_k-Z_{k-1}\,|\,Z_1,\,\cdots,\,Z_{k-1}]=\mathrm{E}[\mathrm{E}\{g_{a,\,b}(X_k,\,X_{i_2},\,\cdots,\,X_{i_a}\,;\,Y_{j_1},\,\cdots,\,Y_{j_b})$$

$$|\,X_1,\,\cdots,\,X_{k-1},\,X_{i_2},\,\cdots,\,X_{i_a},\,Y_{j_1},\,\cdots,\,Y_{j_b}\}|\,Z_1,\,\cdots,\,Z_{k-1}]=0.$$

Therefore $\{Z_k\}_{0\leqq k\leqq u_1}$ is a forward martingale. Applying an upper bound for moments of martingale obtained by Dharmadhikari, Fabian and Jogdeo[2], we have the inequality (2.4), when $c=1$ and $d=0$.

Using equation (2.1), the rest of the proof is similarly obtained.

3. Formal Edgeworth expansion

Let us define the following notations :

$$\sigma_{m,\,n}^2=\mathrm{Var}(U_{m,\,n}),$$

for $0\leqq a\leqq r$ and $0\leqq b\leqq s$

$$k_{a,\,b}(m,\,n)=\binom{m}{r}^{-1}\binom{n}{s}^{-1}\sigma_{m,\,n}^{-1}\binom{m-a}{r-a}\binom{n-b}{s-b},$$

$$\xi_{a,\,b}^2=\mathrm{E}[g_{a,\,b}(X_1,\,\cdots,\,X_a\,;\,Y_1,\,\cdots,\,Y_b)]^2,$$

$$\tau_{m,\,n}^2 = \frac{r^2}{m}\xi_{1,\,0}^2 + \frac{s^2}{n}\xi_{0,\,1}^2,$$

$$\eta(t)=\mathrm{E}[\exp\{it g_{1,\,0}(X_1)\}], \quad \nu(t)=\mathrm{E}[\exp\{it g_{0,\,1}(Y_1)\}].$$

Note that from the equation (2.2) in Lemma 2.2,

$$\sigma_{m,\,n}^2 = \binom{m}{r}^{-2}\binom{n}{s}^{-2}\sum_{a=0}^{r}\sum_{b=0}^{s}\binom{m-a}{r-a}^2\binom{n-b}{s-b}^2\binom{m}{a}\binom{n}{b}\xi_{a,\,b}^2.$$

In this section we assume the following condition

(C1) $\mathrm{E}\,|\,h(X_1,\,\cdots,\,X_r\,;\,Y_1,\,\cdots,\,Y_s)\,|^5 < \infty.$

Before we obtain the expansion, we prepare the useful lemma.

LEMMA 3.1. *If (C1) is satisfied, then there exist positive constants ε and δ such that for* $0 \leq t \leq \varepsilon N^{\frac{1}{2}}$ *and fixed integers* u *and* v,

$$(3.1) \qquad \left|\, I_{u,\,v} - \eta^{m-u}\!\left(\frac{rt}{m\sigma_{m,\,n}}\right)\nu^{n-v}\!\left(\frac{st}{n\sigma_{m,\,n}}\right)\,\right| \leq o(N^{-1})P_1(t)e^{-\delta t^2},$$

where $P_1(t)$ is a polynomial in t and

$$\begin{aligned}
I_{u,\,v} = e^{-\frac{t^2}{2}}\Bigg[&1 - \frac{(it)^2}{2\tau_{m,\,n}^2}\left\{\frac{r^2 s^2}{mn}\xi_{1,\,1}^2 + \frac{[r(r-1)]^2}{2m^2}\xi_{2,\,0}^2 + \frac{[s(s-1)]^2}{2n^2}\xi_{0,\,2}^2\right.\\
&\left. + \frac{ur^2}{m^2}\xi_{1,\,0}^2 + \frac{vs^2}{n^2}\xi_{0,\,1}^2\right\} + \frac{(it)^3}{6\tau_{m,\,n}^3}\left\{\frac{r^3}{m^2}\mathrm{E}[g_{1,\,0}^3(X_1)] + \frac{s^3}{n^2}\mathrm{E}[g_{0,\,1}^3(Y_1)]\right\}\\
&+ \frac{(it)^4}{24\tau_{m,\,n}^4}\left\{\frac{r^4}{m^3}(\mathrm{E}[g_{1,\,0}^4(X_1)] - 3\xi_{1,\,0}^4) + \frac{s^4}{n^3}(\mathrm{E}[g_{0,\,1}^4(Y_1)] - 3\xi_{0,\,1}^4)\right\}\\
&+ \frac{(it)^6}{72\tau_{m,\,n}^6}\left\{\frac{r^3}{m^2}\mathrm{E}[g_{1,\,0}^3(X_1)] + \frac{s^3}{n^2}\mathrm{E}[g_{0,\,1}^3(Y_1)]\right\}^2\Bigg].
\end{aligned}$$

Proof. By the same way of Lemma 2 as Callaert et al.[1], we have the desired result easily.

In order to obtain a formal Edgeworth expansion, we shall find the function $\tilde{\chi}_{m,\,n}(t)$ which satisfies

$$(3.2) \qquad \int_0^{N^{\frac{1}{4}}/\log N} t^{-1}\,|\,\chi_{m,\,n}(t) - \tilde{\chi}_{m,\,n}(t)\,|\,dt = o(N^{-1})$$

where $\chi_{m,\,n}(t) = \mathrm{E}[\exp(it\,\sigma_{m,\,n}^{-1}U_{m,\,n})]$. (See Callaert et al. [1] and Maesono[6].)

Note that

$$\sigma_{m,\,n}^{-1}U_{m,\,n} = \sum_{a=0}^{r}\sum_{b=0}^{s}k_{a,\,b}(m,\,n)A_{a,\,b}.$$

Let

$$_1\chi_N(t) = \mathrm{E}[\exp\{it\sum_{1\leq a+b\leq 3}k_{a,\,b}(m,\,n)A_{a,\,b}\}].$$

Then from Lemma 2.3 and the fact $k_{a,\,b}(m,\,n) = O(N^{\frac{1}{2}-(a+b)})$, we have

$$\int_0^{N^{\frac{1}{4}}/\log N} t^{-1}|\chi_{m,\,n}(t)-{}_1\chi_N(t)|\,dt = o(N^{-1}).$$

Furthermore let us define

$$\chi_N^*(t)=\mathrm{E}[\exp\{it(k_{1,\,0}A_{1,\,0}+k_{0,\,1}A_{0,\,1})\}]+itk_{1,\,1}\mathrm{E}[A_{1,\,1}\exp\{it(k_{1,\,0}A_{1,\,0}+k_{0,\,1}A_{0,\,1})\}]$$

$$+\mathrm{E}[\exp\{it(k_{1,\,0}A_{1,\,0}+k_{0,\,1}A_{0,\,1})\}\{itk_{2,\,0}A_{2,\,0}+itk_{0,\,2}A_{0,\,2}+(it)^2k_{1,\,1}A_{1,\,1}k_{2,\,0}A_{2,\,0}$$

$$(3.3)\qquad +(it)^2k_{1,\,1}A_{1,\,1}k_{0,\,2}A_{0,\,2}+(it)^2k_{2,\,0}A_{2,\,0}k_{0,\,2}A_{0,\,2}$$

$$+\frac{(it)^2}{2}k_{1,\,1}^2A_{1,\,1}^2+\frac{(it)^2}{2}k_{2,\,0}^2A_{2,\,0}^2+\frac{(it)^2}{2}k_{0,\,2}^2A_{0,\,2}^2$$

$$+itk_{2,\,1}A_{2,\,1}+itk_{1,\,2}A_{1,\,2}+itk_{3,\,0}A_{3,\,0}+itk_{0,\,3}A_{0,\,3}\}],$$

where $k_{a,\,b}$ is an abbreviation of $k_{a,\,b}(m,\,n)$. Then by the similar way of Callaert et al. [1] and Maesono[6], we have

$$\int_0^{N^{\frac{1}{4}}/\log N} t^{-1}|{}_1\chi_N(t)-\chi_N^*(t)|\,dt = o(N^{-1}).$$

Since $k_{1,\,0}(m,\,n)=r/(m\sigma_{m,\,n})$ and $k_{0,\,1}(m,\,n)=s/(n\sigma_{m,\,n})$, from Lemma 3.1 the approximation of the first term of (3.3) is $I_{0,\,0}$. And the second term of (3.3) is approximated as follows. From Lemma 2.2, we have

$$itk_{1,\,1}\mathrm{E}[A_{1,\,1}\exp\{it(k_{1,\,0}A_{1,\,0}+k_{0,\,1}A_{0,\,1})\}]$$

$$=itk_{1,\,1}mn\eta^{m-1}\Big(\frac{rt}{m\sigma_{m,\,n}}\Big)\nu^{n-1}\Big(\frac{st}{n\sigma_{m,\,n}}\Big)$$

$$\times\mathrm{E}[g_{1,\,1}(X_1;Y_1)\exp\{it(k_{1,\,0}g_{1,\,0}(X_1)+k_{0,\,1}g_{0,\,1}(Y_1))\}]$$

which will be denoted by $itk_{1,\,1}mnI_{1,\,1}^*E_1^*$. From Lemma 3.1, $I_{1,\,1}^*$ is approximated by $I_{1,\,1}$. Taking the first few terms of the Taylor series for approximating E_1^* and using Lemma 2.2, we have an approximation E_1 such that

$$E_1=\frac{(it)^2rs}{mn\sigma_{m,\,n}^2}\mathrm{E}[g_{1,\,0}(X_1)g_{0,\,1}(Y_1)g_{1,\,1}(X_1;Y_1)]$$

$$+\frac{(it)^3}{6\sigma_{m,\,n}^3}\Big\{\frac{3r^2s}{m^2n}\mathrm{E}[g_{1,\,0}^2(X_1)g_{0,\,1}(Y_1)g_{1,\,1}(X_1;Y_1)]+\frac{3rs^2}{mn^2}\mathrm{E}[g_{1,\,0}(X_1)g_{0,\,1}^2(Y_1)g_{1,\,1}(X_1;Y_1)]\Big\}.$$

Since $\sigma_{m,\,n}^2=\tau_{m,\,n}^2(1+O(N^{-1}))$ and $k_{1,\,1}(m,\,n)=rs/(mn\tau_{m,\,n})(1+O(N^{-1}))$, we can obtain an approximation $\rho(t)$ of the second term as follows:

$$\rho(t)=e^{-\frac{1}{2}t^2}\Big(\frac{r^2s^2(it)^3}{mn\tau_{m,\,n}^3}\mathrm{E}[g_{1,\,0}(X_1)g_{0,\,1}(Y_1)g_{1,\,1}(X_1;Y_1)]$$

$$+\frac{(it)^4}{6\tau_{m,\,n}^4}\Big\{\frac{3r^3s^2}{m^2n}\mathrm{E}[g_{1,\,0}^2(X_1)g_{0,\,1}(Y_1)g_{1,\,1}(X_1;Y_1)]+\frac{3r^2s^3}{mn^2}\mathrm{E}[g_{1,\,0}(X_1)g_{0,\,1}^2(Y_1)g_{1,\,1}(X_1;Y_1)]\Big\}$$

$$+\frac{r^2s^2(it)^6}{6mn\tau_{m,\,n}^6}\mathrm{E}[g_{1,\,0}(X_1)g_{0,\,1}(Y_1)g_{1,\,1}(X_1;Y_1)]\Big[\frac{r^3}{m^2}\mathrm{E}[g_{1,\,0}^3(X_1)]+\frac{s^3}{n^2}\mathrm{E}[g_{0,\,1}^3(Y_1)]\Big]\Big).$$

From the condition (C1) and $\sigma_{m,\,n}^2=O(N^{-1})$, we have

$$|E_1^*-E_1|\leq P_2(t)O(N^{-2}),$$

where $P_2(t)$ is a Polynomial in t. Then from (3.1) and above inequality, we have

$$\int_0^{N^{\frac{1}{4}}/\log N} t^{-1} | it k_{1,1} m n I^*_{1,1} E^*_1 - \rho(t)| dt = o(N^{-1}).$$

Similarly we can obtain the approximations of the rest terms of (3.3). Hence we have an approximation of $\chi^*_N(t)$ as follows :

$$\tilde{\chi}_{m,n}(t) = e^{-\frac{1}{2}t^2}\left\{1 + \frac{K_3}{6}(it)^3 + \frac{K_4}{24}(it)^4 + \frac{K_3^2}{72}(it)^6\right\}$$

where

$$K_3 = \frac{1}{\tau_{m,n}^3}\left\{\frac{r^3}{m^2}E[g_{1,0}^3(X_1)] + \frac{s^3}{n^2}E[g_{0,1}^3(Y_1)] + \frac{6r^2s^2}{mn}E[g_{1,0}(X_1)g_{0,1}(Y_1)g_{1,1}(X_1 ; Y_1)]\right.$$

$$+ \frac{3r^3(r-1)}{m^2}E[g_{1,0}(X_1)g_{1,0}(X_2)g_{2,0}(X_1, X_2)]$$

$$\left. + \frac{3s^3(s-1)}{n^2}E[g_{0,1}(Y_1)g_{0,1}(Y_2)g_{0,2}(Y_1, Y_2)]\right\}$$

and

$$K_4 = \frac{1}{\tau_{m,n}^4}\left\{\frac{r^4}{m^3}\left(E[g_{1,0}^4(X_1)] - 3\xi_{1,0}^4\right) + \frac{s^4}{n^3}\left(E[g_{0,1}^4(Y_1)] - 3\xi_{0,1}^4\right)\right.$$

$$+ \frac{12r^3s^2}{m^2n}E[g_{1,0}^2(X_1)g_{0,1}(Y_1)g_{1,1}(X_1 ; Y_1)] + \frac{12r^2s^3}{mn^2}E[g_{1,0}(X_1)g_{0,1}^2(Y_1)g_{1,1}(X_1 ; Y_1)]$$

$$+ \frac{12r^4(r-1)}{m^3}E[g_{1,0}^2(X_1)g_{1,0}(X_2)g_{2,0}(X_1, X_2)] + \frac{12s^4(s-1)}{n^3}E[g_{0,1}^2(Y_1)g_{0,1}(Y_2)g_{0,2}(Y_1, Y_2)]$$

$$+ \frac{12r^3(r-1)s^2}{m^2n}E[g_{1,0}(X_1)g_{1,0}(X_2)g_{0,1}(Y_1)g_{2,1}(X_1, X_2 ; Y_1)] + \frac{12r^2s^3(s-1)}{mn^2}E[g_{1,0}(X_1)g_{0,1}(Y_1)$$

$$\times g_{0,1}(Y_2)g_{1,2}(X_1 ; Y_1, Y_2)] + \frac{4r^4(r-1)(r-2)}{m^3}E[g_{1,0}(X_1)g_{1,0}(X_2)g_{1,0}(X_3)g_{3,0}(X_1, X_2, X_3)]$$

$$+ \frac{4s^4(s-1)(s-2)}{n^3}E[g_{0,1}(Y_1)g_{0,1}(Y_2)g_{0,1}(Y_3)g_{0,3}(Y_1, Y_2, Y_3)] + \frac{24r^3(r-1)s^2}{m^2n}E[g_{1,0}(X_2)g_{0,1}(Y_1)$$

$$\times g_{1,1}(X_1 ; Y_1)g_{2,0}(X_1, X_2)] + \frac{24r^2s^3(s-1)}{mn^2}E[g_{1,0}(X_1)g_{0,1}(Y_2)g_{1,1}(X_1 ; Y_1)g_{0,2}(Y_1, Y_2)]$$

$$+ \frac{12r^2s^4}{mn^2}E[g_{0,1}(Y_1)g_{0,1}(Y_2)g_{1,1}(X_1 ; Y_1)g_{1,1}(X_1 ; Y_2)] + \frac{12r^4s^2}{m^2n}E[g_{1,0}(X_1)g_{1,0}(X_2)$$

$$\times g_{1,1}(X_1 ; Y_1)g_{1,1}(X_2 ; Y_1)] + \frac{12r^4(r-1)^2}{m^3}E[g_{1,0}(X_2)g_{1,0}(X_3)g_{2,0}(X_1, X_2)g_{2,0}(X_1, X_3)]$$

$$\left. + \frac{12s^4(s-1)^2}{n^3}E[g_{0,1}(Y_2)g_{0,1}(Y_3)g_{0,2}(Y_1, Y_2)g_{0,2}(Y_1, Y_3)]\right\}.$$

This $\tilde{\chi}_{m,n}(t)$ satisfies the equation (3.2). Inverting $\tilde{\chi}_{m,n}(t)$, we have a formal Edgeworth expansion $Q_{m,n}(x)$ such that

$$Q_{m,n}(x) = \Phi(x) - \phi(x)\left[\frac{K_3}{6}(x^2-1) + \frac{K_4}{24}(x^3-3x) + \frac{K_3^2}{72}(x^5-10x^3+15x)\right]$$

where $\Phi(x)$ and $\phi(x)$ denote the distribution function and density of the standard normal

distribution. Note that $K_3 = O(N^{-\frac{1}{2}})$ and $K_4 = O(N^{-1})$. Then $Q_{m,n}(x)$ is the expansion with remainder term $o(N^{-1})$.

4. Conditions for the valid expansion

In order to prove the validity of the formal Edgeworth expansion $Q_{m,n}(x)$, we shall apply Esseen's smoothing lemma[3]. From smoothing lemma, we have

$$\sup_x |P(\sigma_{m,n}^{-1} U_{m,n} \leq x) - Q_{m,n}(x)| \leq \frac{1}{\pi} \int_{-N\log N}^{N\log N} |t|^{-1} |\chi_{m,n}(t) - \tilde{\chi}_{m,n}(t)| \, dt + o(N^{-1})$$

where $Q_{m,n}(x)$, $\chi_{m,n}(t)$ and $\tilde{\chi}_{m,n}(t)$ are defined in the previous section. Since the proof for the negative part of t is similar to that for positive one, we shall find the conditions which ensure

$$\int_0^{N\log N} t^{-1} |\chi_{m,n}(t) - \tilde{\chi}_{m,n}(t)| \, dt = o(N^{-1}).$$

Let us define

$$_2\chi_N(t) = E[\exp\{ it \sum_{1 \leq a+b \leq 4} k_{a,b}(m, n) A_{a,b} \}]$$

and

$$_3\chi_N(t) = E[\exp\{ it \sum_{1 \leq a+b \leq 5} k_{a,b}(m, n) A_{a,b} \}].$$

Then from Lemma 2.3 and $k_{a,b}(m, n) = O(N^{\frac{1}{2}-(a+b)})$, we have

$$\int_{N^{\frac{1}{4}}/\log N}^{N^{\frac{3}{4}}/\log N} t^{-1} |\chi_{m,n}(t) - {}_2\chi_N(t)| \, dt = o(N^{-1})$$

and

$$\int_{N^{\frac{3}{4}}/\log N}^{N\log N} t^{-1} |\chi_{m,n}(t) - {}_3\chi_N(t)| \, dt = o(N^{-1}).$$

Then putting

$$(\mathrm{I}) = \int_0^{N^{\frac{1}{4}}/\log N} t^{-1} |\chi_{m,n}(t) - \tilde{\chi}_{m,n}(t)| \, dt,$$

$$(\mathrm{II}) = \int_{N^{\frac{1}{4}}/\log N}^{N^{\frac{3}{4}}/\log N} t^{-1} |{}_2\chi_N(t)| \, dt, \qquad (\mathrm{III}) = \int_{N^{\frac{3}{4}}/\log N}^{N\log N} t^{-1} |{}_3\chi_N(t)| \, dt$$

and

$$(\mathrm{IV}) = \int_{\log N}^{\infty} t^{-1} |\tilde{\chi}_{m,n}(t)| \, dt,$$

we have

$$\int_0^{N\log N} t^{-1} |\chi_{m,n}(t) - \tilde{\chi}_{m,n}(t)| \, dt \leq (\mathrm{I}) + (\mathrm{II}) + (\mathrm{III}) + (\mathrm{IV}) + o(N^{-1}).$$

In Section 2, we have proved that (I) is $o(N^{-1})$ under the condition (C1). It immediately follows that $(\mathrm{IV}) = o(N^{-1})$ under the same condition (C1).

To obtain an order of (II), we shall consider following decomposition. For $0 \leq u \leq m-a+1$ and $0 \leq v \leq n-b+1$, let us define

$$D_{a,b}(u,\ v)=B_{a,b}(u,\ m-a+2,\ \cdots,\ m\ ;\ v,\ n-b+2,\ \cdots,\ n),$$

$$H_{a,b}(u,\ v)=\sum_{i_1=1}^{u}\sum_{i_2=i_1+1}^{m-a+2}\cdots\sum_{i_a=i_{a-1}+1}^{m}\sum_{j_1=v+1}^{n-b+1}\sum_{j_2=j_1+1}^{n-b+2}\cdots\sum_{j_b=j_{b-1}+1}^{n}g_{a,b}(X_{i_1},\ \cdots,\ X_{i_a}\ ;\ Y_{j_1},\ \cdots,\ Y_{j_b}),$$

$$L_{a,b}(u,\ v)=\sum_{i_1=u+1}^{m-a+1}\sum_{i_2=i_1+1}^{m-a+2}\cdots\sum_{i_a=i_{a-1}+1}^{m}\sum_{j_1=1}^{v}\sum_{j_2=j_1+1}^{n-b+2}\cdots\sum_{j_b=j_{b-1}+1}^{n}g_{a,b}(X_{i_1},\ \cdots,\ X_{i_a}\ ;\ Y_{j_1},\ \cdots,\ Y_{j_b})$$

and

$$R_{a,b}(u,\ v)=\sum_{i_1=u+1}^{m-a+1}\sum_{i_2=i_1+1}^{m-a+2}\cdots\sum_{i_a=i_{a-1}+1}^{m}\sum_{j_1=v+1}^{n-b+1}\sum_{j_2=j_1+1}^{n-b+2}\cdots\sum_{j_b=j_{b-1}+1}^{n}g_{a,b}(X_{i_1},\ \cdots,\ X_{i_a}\ ;\ Y_{j_1},\ \cdots,\ Y_{j_b}).$$

Then $A_{a,b}=D_{a,b}(u,\ v)+H_{a,b}(u,\ v)+L_{a,b}(u,\ v)+R_{a,b}(u,\ v)$. Here $H_{a,b}(u,\ v)$ and $\{Y_1,\ \cdots,\ Y_v\}$ are independent. Similarly $L_{a,b}(u,\ v)$ and $\{X_1,\ \cdots,\ X_u\}$ are independent. Furthermore $R_{a,b}(u,\ v)$ and $\{X_1,\ \cdots,\ X_u,\ Y_1,\ \cdots,\ Y_v\}$ are also independent. From Lemma 2.3, we have

$$E|D_{a,b}(u,\ v)|^p\leq 0((uv)^{\frac{p}{2}}N^{\frac{p}{2}(a+b-2)}).$$

And we get

$$E|H_{a,b}(u,\ v)|^p\leq E|H_{a,b}(u,\ v)+D_{a,b}(u,\ v)-D_{a,b}(u,\ v)|^p$$

$$\leq 2^{p-1}\{E|H_{a,b}(u,\ v)+D_{a,b}(u,\ v)|^p+E|D_{a,b}(u,\ v)|^p\}$$

$$=0(u^{\frac{p}{2}}N^{\frac{p}{2}(a+b-1)}).$$

Similarly we have

$$E|L_{a,b}(u,\ v)|^p\leq 0(v^{\frac{p}{2}}N^{\frac{p}{2}(a+b-1)}).$$

Hence we can obtain an appropriate upper bound for $_2\chi_N(t)$ by the similar way of Lemma 4 as Callaert et al. [1] and Lemma 3 as Maesono[6]. The bound and the proof of it are rather complicated, and will be omitted here.

In addition to (C1), we assume that

(C2) $\lim_{|t|\to\infty}|\eta(t)|<1$ and $\lim_{|t|\to\infty}|\nu(t)|<1.$

Then by the same arguments which have been described in Callaert et al. [1] pp308–309, we can prove that (Ⅱ) is $o(N^{-1})$ under the conditions (C1) and (C2).

Let us define

$$\zeta(x_1,\ x_2)=w_{2,0}(x_1,\ x_2)-\{(r-2)/(r-1)\}[w_{1,0}(x_1)+w_{1,0}(x_2)]$$

$$\mu(y_1,\ y_2)=w_{0,2}(y_1,\ y_2)-\{(s-2)/(s-1)\}[w_{0,1}(y_1)+w_{0,1}(y_2)].$$

Then if we assume the following complicated condition, it will be possible to show that (Ⅲ) is $o(N^{-1})$.

(C3) There exist positive constants $c_1<1$ and $c_2<1$ such that for $u=[m^\alpha]$ and $v=[n^\beta]$, where $0<\alpha<1/8$ and $0<\beta<1/8$,

$$P\Big(\Big|E[\exp\{it(k_{2,0}\sum_{j=u+1}^{m}\zeta(X_1,\ X_j)$$

$$+ k_{1,1} \sum_{j=v+1}^{n} g_{1,1}(X_1 ; Y_j))\} \mid X_{u+1}, \cdots, X_m, Y_{v+1}, \cdots, Y_n] \mid \leq c_1,$$

$$or, \ |E[\exp\{ it(k_{0,2} \sum_{j=v+1}^{n} \mu(Y_1, Y_j)$$

$$+ k_{1,1} \sum_{j=u+1}^{m} g_{1,1}(X_j ; Y_1))\} \mid X_{u+1}, \cdots, X_m, Y_{v+1}, \cdots, Y_n] \mid \leq c_2\Big)$$

$$\geq 1 - o\Big(\frac{1}{N \log N}\Big)$$

uniformly for all $t \in [N^{\frac{3}{4}}/\log N, \ N \log N]$.

This is an extension of the conditions which are given in Callaert et al. [1] and Maesono[6]. It may be hard to check the validity of (C3) in most of the examples encountered in statistics. Then it is desirable to obtain simple condition which ensures $(\text{III}) = o(N^{-1})$.

From the discussion above, we have

THEOREM.   If the conditions (C1), (C2) and (C3) are satisfied, then

$$\sup_x | P(\sigma_{m,n}^{-1} U_{m,n} \leq x) - Q_{m,n}(x)| = o(N^{-1}).$$

REMARK.   Instead of condition (C1), the asymptotic expansion may be valid under the existence of a fourth moment of kernel h. Lin[5] has proved it in the case of one-sample U-statistics with kernel of degree two.

References

[1] H. Callaert, P. Janssen and N. Veraverbeke, An Edgeworth expansion for U-statistics, Ann. Statist., 8 (1980), 299-312.

[2] S. W. Dharmadhikari, V. Fabian and K. Jogdeo, Bounds of moments of martingales, Ann. Math. Statist., 39 (1968), 1719-1723.

[3] C. F. Esseen, Fourier analysis of distribution functions : A mathematical study of the Laplace-Gaussian law, Acta Math., 77 (1945), 1-125.

[4] W. Hoeffding, The strong law of large numbers for U-statistics, Univ. of North Carolina Institute of Statistics Mimeo Series, No. 302, (1961).

[5] Lin Zhengyan, A note on the asymptotic expansion for U-statistics (in Chinese), Acta Mathematica Sinica, 27 (1984), 595-598.

[6] Y. Maesono, Edgeworth expansion for one-sample U-statistics, submitted.

[7] R. J. Serfling, Approximation Theorems of Mathematical Statistics, Wiley, New York, 1980.