THE BERRY-ESSEEN BOUND FOR STUDENTIZED U-STATISTICS*

by

Lin Zhengyan
Hangzhou University
P. R. China

Technical Report #97-19

Department of Statistics
Purdue University
West Lafayette, IN USA

October, 1997

*Project supported by NSFC(19571021) and NSFZP(195043), finished during the author's visit at Department of Statistics, Purdue University, U.S.A.

The Berry-Esseen Bound for Studentized U-Statistics*

LIN ZHENGYAN

Hangzhou University
Hangzhou 310028, P. R. China

ABSTRACT. We establish the Berry-Esseen bound $O(n^{-\delta/2})$ for a studentized U-statistic under the assumption of existence of the $2+\delta$-th moment.
Keywords: Studentized U-statistics, Berry-Esseen Bound.

1.Introduction and result.

Let $\{X_n, n \geq 1\}$ be a sequence of i.i.d. random variables with a common distribution $F(x)$, $h(x_1, x_2)$ a symmetric function. Define a U-statistic

$$U_n = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} h(X_i, X_j).$$

Arvesen(1969) studied convergence of normality for a studentized U- statistic

$$\sqrt{n}(U_n - \theta)/S_n,$$

where $\theta = Eh(X_1, X_2)$ and

$$S_n^2 = 4(n-1)(n-2)^{-2} \sum_{i=1}^{n} [(n-1)^{-1} \sum_{j \neq i, j=1}^{n} h(X_i, X_j) - U_n]^2.$$

Write the uniform distance

$$\Delta_n = \sup_x \left| P(\sqrt{n}(U_n - \theta)/S_n < x) - \Phi(x) \right|,$$

where $\Phi(x)$ is the standard normal distribution. Callaert and Veraverbeke(1981) established the Berry-Esseen bound $O(n^{-1/2})$ under the condition $E\left|h(x_1, x_2)\right|^{4.5} < \infty$. Zhao(1983) weakened the moment condition to $Eh(X_1, X_2)^4 < \infty$. Can this moment condition be weakened further? Our paper shows that it is possible and gives a sharp result, i.e. under the condition $E\left|h(X_1, X_2)\right|^{2+\delta} < \infty$ for some $0 < \delta \leq 1$, the Berry-Esseen bound $\Delta_n = O(n^{-\delta/2})$ will be established.

Let $g(X_1) = E(h(X_1, X_2) - \theta \,|X_1)$.

Theorem. *Suppose that* $E\left|h(X_1, X_2)\right|^{2+\delta} < \infty$ *for some* $0 < \delta \leq 1$ *and* $Eg(X_1)^2 > 0$. *Then*

$$\Delta_n = O(n^{-\delta/2}).$$

---

*) Project supported by NSFC(19571021) and NSFZP(195043), finished during the author's visit at Department of Statistics, Purdue University, U.S.

2. Proof

In order to prove our theorem, we need the following lemma.

Lemma(Cheng 1996). *Let H_n be a U-statistic of degree j ($j = 1$ or 2) based on a kernel $u(x_1, x_2)$. Then for $1 < r \le 2$*

$$E\left|H_n\right|^r \le \begin{cases} n^{-(r-1)}E\left|u(X_1,X_2)\right|^r, & j=1, \\ 4n^{-2(r-1)}E\left|u(X_1,X_2)\right|^r, & j=2; \end{cases}.$$

and for $r > 2$

$$E\left|H_n\right|^r \le \begin{cases} c_r n^{-r/2}E\left|u(X_1,X_2)\right|^r, & j=1, \\ c'_r n^{-r}E\left|u(X_1,X_2)\right|^r, & j=2, \end{cases}.$$

where c_r and c'_r are positive constants depending only on r.

Without loss of generality, assume that $\theta = 0$. Let

$$\varphi(X_i, X_j) = h(X_i, X_j) - g(X_i) - g(X_j),$$

$$g^*(X_i) = E[g(X_j)\varphi(X_i, X_j)\,|X_i].$$

For simplicity, put

$$h_{ij} = h(X_i, X_j),\ g_i = g(X_i),\ \varphi_{ij} = \varphi(X_i, X_j),\ g_i^* = g^*(X_i),\ \text{and}\ \sigma_g^2 = Eg_1^2.$$

Following the decomposition of S_n^2 in [2], write

$$S_n^2 = 4\sigma_g^2(1 + T_n + R_n) \tag{2.1}$$

with

$$T_n = \frac{1}{n\sigma_g^2}\sum_{i=1}^{n}(g_i^2 - \sigma_g^2 + 2g_i^*),\quad R_n = \frac{1}{4\sigma_g^2}\sum_{p=1}^{6}R_{np},$$

$$R_{n1} = -4\binom{n}{2}^{-1}\sum_{i<j}g_ig_j,\quad R_{n2} = 4\binom{n}{2}^{-1}\sum_{i<j}[(g_i+g_j)\varphi_{ij} - g_i^* - g_j^*],$$

$$R_{n3} = -\frac{8}{n}\sum_{i=1}^{n}\left[g_i\binom{n-1}{2}^{-1}\sum_{k<l}^{(i)}\varphi_{kl}\right],$$

$$R_{n4} = \frac{4}{n-2}\sum_{i=1}^{n}\left[\binom{n-1}{2}^{-1}\sum_{k<l}^{(i)}\varphi_{ik}\varphi_{il}\right],$$

$$R_{n5} = -\frac{4n(n-1)}{(n-2)^2}\left[\binom{n}{2}^{-1}\sum_{i<j}\varphi_{ij}\right]^2,\quad R_{n6} = \frac{4n}{(n-2)^2}\binom{n}{2}^{-1}\sum_{i<j}\varphi_{ij}^2,$$

where $\sum\limits_{i<j}$ and $\sum\limits_{k<l}^{(i)}$ stand for $\sum\limits_{1\le i<j\le n}$ and $\sum\limits_{1\le k<l\le n, k\ne i, l\ne i}$ respectively.

Now we can write

$$\frac{\sqrt{n}U_n}{S_n} = \frac{\sqrt{n}U_n}{2\sigma_g}\left(1 + T_n + R_n\right)^{-1/2}. \tag{2.2}$$

Noting that $n\sigma_g^2 T_n$ is a sum of i.i.d. random variables with mean zero, we have

$$P(|T_n| > \frac{1}{4}) \le 8E\,|T_n|^{1+\delta/2} \le cn^{-\delta/2}, \tag{2.3}$$

here, and in the sequel, c indicates a positive constant, whose value is irrelevant.

For R_n, we consider only R_{n4} as an example. Noting $\binom{n-1}{2}^{-1}\sum_{k<l}^{(i)}\varphi_{ik}\varphi_{il}$ is a degenerate U-statistic and using the lemma with $j=2$ and $r = 1+\delta/2$ we have

$$E\,|R_{n4}| \le (E\,|R_{n4}|^{1+\delta/2})^{2/(2+\delta)} \le cn^{-2\delta/(2+\delta)} \le cn^{-\delta/2}.$$

For the other R_{ni}, we have the same estimation. Hence $E\,|R_n| \le cn^{-\delta/2}$, and further

$$P(|R_n| > \frac{1}{4}) \le cn^{-\delta/2}. \tag{2.4}$$

Let $w : R \to R$ denotes a function which is infinitely differentiable with bounded derivation such that

$$\frac{2}{3} \le w(x) \le 2 \text{ and } w(x) = x^{-1/2} \text{ for } \frac{1}{2} \le |x| \le \frac{3}{2}.$$

By (2.3) and (2.4) we can consider

$$U_n^* := \frac{\sqrt{n}U_n}{2\sigma_g}w(1+T_n+R_n)$$

instead of $\sqrt{n}U_n/S_n$.

Now introduce truncation variables

$$g_{in} = g_i I(|g_i| \le \sqrt{n}),$$

$$g_{in}^* = g_i^* I(|g_i^*| \le n),$$

$$\varphi_{ijn} = \varphi_{ij} I(|\varphi_{ij}| \le n(\delta)),$$

where $n(\delta) = n^{(4+\delta)/2(2+\delta)}$. We can replace g_i by g_{in} in T_n and R_n since

$$P(\bigcup_{i=1}^{n}\{g_i \ne g_{in}\}) \le nP(|g_1| > \sqrt{n}) \le n^{-\delta/2}E\,|g_1|^{2+\delta}. \tag{2.5}$$

Similarly we can replace g_i^* (or φ_{ij}) by g_{in}^* (or φ_{ijn}). Moreover

$$|Eg_{1n}| = |Eg_1 I(|g_1| > \sqrt{n})| \le n^{-(1+\delta)/2}E\,|g_1|^{2+\delta}, \tag{2.6}$$

$$|E\varphi_{12n}| = |E\varphi_{12}I(|\varphi_{12}| > n(\delta))| \le n^{-\frac{(1+\delta)(4+\delta)}{2(2+\delta)}}E\,|\varphi_{12}|^{2+\delta} \tag{2.7}$$

and similarly

$$Eg_{1n}^2 \le n^{-\delta/2}E\,|g_1|^{2+\delta}, \quad E\,|g_{1n}^*| \le n^{-\delta/2}E\,|g_1^*|^{1+\delta/2}. \tag{2.8}$$

Furthermore let $g'_{in} = g_{in} - Eg_{in}$, $\varphi'_{ijn} = \varphi_{ijn} - E\varphi_{ijn}$, $\bar{g}_{in} = E(g'_{jn}\varphi'_{ijn}|X_i)$. Then, noting (2.5)-(2.8), we can anew define

$$T_n = \frac{1}{n\sigma_g^2}\sum_{i=1}^{n}(g_{in}^2 - Eg_{in}^2 + 2g_{in}^* - 2Eg_{in}^*) =: \frac{1}{n\sigma_g^2}\sum_{i=1}^{n} d_{in}$$

and R_n with g'_{in}, φ'_{ijn} and $\bar{g}_{in}$ instead of g_i, φ_{ij} and g_i^* respectively. Put $w_n = w(1+T_n+R_n)$, $Q_n = \frac{1}{\sqrt{n}\sigma_g}\sum_{i=1}^{n} g'_{in}$, $D_n = \frac{\sqrt{n}}{2\sigma_g}\binom{n}{2}^{-1}\sum_{1\le i<j\le n}\varphi'_{ijn}$. We can consider

$$U'_n := (Q_n + D_n)w_n$$

instead of U_n^*.

Denote the characteristic function of X by f_X. Applying the Berry-Esseen inequality for characteristic functions, we need only to show that

$$J_n := \int\limits_{C\le|t|\le\epsilon n^{\delta/2}} |t|^{-1}\left|f_{U'_n}(t) - e^{-t^2/2}\right| dt \le cn^{-\delta/2} \tag{2.9}$$

for any given small $\epsilon > 0$, where $C > 0$ is a constant. Write

$$\begin{aligned} J_n \le& \int\limits_{C\le|t|\le\epsilon n^{\delta/2}} |t|^{-1}\left|f_{(Q_n+D_n)w_n}(t) - f_{Q_n w_n}(t)\right| dt \\ &+ \int\limits_{C\le|t|\le\epsilon n^{\delta/2}} |t|^{-1}\left|f_{Q_n w_n}(t) - e^{-t^2/2}\right| dt \\ =:& J_{n1} + J_{n2}. \end{aligned} \tag{2.10}$$

Estimate J_{n2} first. Let

$$m := m(n,t) = [C_1 n t^{-2}\ln|t|] \quad for\ C \le |t| \le \epsilon n^{\delta/2}$$

where C_1 is a constant large enough. By choosing our constants, we may assume $2 \le m \le n/2$. Let $q_n = \frac{1}{\sqrt{n}\sigma_g}\sum_{i=1}^{m} g'_{in}$, $q'_n = Q_n - q_n$, $t_n = \frac{1}{n\sigma_g^2}\sum_{i=1}^{m} d_{in}$, $t'_n = T_n - t_n$. Furthermore, let r_{np} are R_{np}, $p = 1,\cdots,6$ respectively but with the summation being extended over $\{(i,j), 1 \le i < j \le m\}$ $(p = 1,2,5,6)$ or $\{(i,k,l) : 1 \le i \le m, 1 \le k < l \le m,\ k \ne i,\ l \ne i\}$ $(p = 3,4)$; $r'_{np} = R_{np} - r_{np}$. Put $r_n = \sum_{p=1}^{6} r_{np}$, $r'_n = \sum_{p=1}^{6} r'_{np}$. Consider $f_{Q_n w_n}(t) = E\exp(itQ_n w_n)$. Write

$$\begin{aligned} Q_n w_n =& q_n w(1+t'_n) + q_n(t_n+R_n)w'(1+t'_n+\theta(t_n+R_n)) \\ &+ q'_n w(1+t'_n+r'_n) + q'_n(t_n+r_n)w'(1+t'_n+r'_n) \\ &+ q'_n(t_n+r_n)^2 w''(1+t'_n+r'_n+\theta(t_n+r_n))/2, \end{aligned} \tag{2.11}$$

where $|\theta| \le 1$. We have

$$Et_n^2 \le cn^{-2}mE(g_{1n}^4 + g_{1n}^{*2}) \le cmn^{-1-\delta/2}.$$

Estimate Er_n^2. As an example, we consider Er_{n4}^2.

$$Er_{n4}^2 \le cn^{-4}m^2 E\varphi_{12n}'^4 \le cm^2n^{-4}n(\delta)^{2-\delta} = cm^2n^{-2-\frac{\delta(6+\delta)}{2(2+\delta)}}, \tag{2.12}$$

here the lemma is used. Noting that q_n' and $t_n + r_n$ are independent and the derivatives of w are bounded, we obtain

$$\begin{aligned} &E\left|q_n'(t_n+r_n)^2 w''(1+t_n'+r_n'+\theta(t_n+r_n))\right| \\ \le& c(Eq_n'^2)^{1/2}E(t_n+r_n)^2 \le cmn^{-1-\delta/2}. \end{aligned} \tag{2.13}$$

Thus we can remove the term $itq_n'(t_n+r_n)^2w''(1+t_n'+r_n'+\theta(t_n+r_n))/2$ in the exponent of itQ_nw_n (cf. (2.11)) since $|t|\, mn^{-1-\delta/2} = O(|t|^{-1}(\ln|t|)n^{-\delta/2})$. Deal with the second term in the right hand side of (2.11). To this end we estimate $E\left|R_n\right|^{(2+\delta)/(1+\delta)}$. As an example, we consider R_{n4}. By the lemma,

$$\begin{aligned} &E\left|R_{n4}\right|^{(2+\delta)/(1+\delta)} \le cn^{-2/(1+\delta)}E\left|\varphi_{12n}'\right|^{2(2+\delta)/(1+\delta)} \\ \le& cn^{-\frac{2}{1+\delta}}n(\delta)^{\frac{(1-\delta)(2+\delta)}{1+\delta}} = cn^{-\frac{\delta(3+\delta)}{2(1+\delta)}}. \end{aligned} \tag{2.14}$$

It follows that

$$\begin{aligned} &E\left|q_nR_nw'(1+t_n'+\theta(t_n+R_n))\right| \\ \le& c(E\left|q_n\right|^{2+\delta})^{\frac{1}{2+\delta}}(E\left|R_n\right|^{\frac{2+\delta}{1+\delta}})^{\frac{1+\delta}{2+\delta}} \le cm^{\frac{1}{2}}n^{-\frac{1}{2}-\frac{\delta(3+\delta)}{2(2+\delta)}}. \end{aligned} \tag{2.15}$$

Split the sum q_nt_n into its diagonal and non-diagonal parts.

$$q_nt_n = u_n + v_n, \quad u_n = n^{-3/2}\sigma_g^{-3}\sum_{i=1}^{m} g_{in}'d_{in}. \tag{2.16}$$

We have

$$E\left|u_n\right| \le cn^{-3/2}m(E\left|g_{1n}\right|^3 + E\left|g_{1n}g_{1n}^*\right|) \le cmn^{-1-\delta/2}, \tag{2.17}$$

which allows us to remove $itu_nw'(1+t_n'+\theta(t_n+R_n))$ in the exponent of $f_{Q_nw_n}(t)$. Furthermore

$$Ev_n^2 \le cn^{-3}m^2(Eg_{1n}'^2)E(g_{1n}^2+g_{1n}^*)^2 \le cm^2n^{-2-\delta/2}. \tag{2.18}$$

Write

$$v_nw'(1+t_n'+\theta(t_n+R_n)) = v_nw'(1+t_n') + \tau, \tag{2.19}$$

where, noting $ER_n^2 \le cn^{-7\delta/6}$, we have

(2.20)

$$\begin{aligned} E|\tau| &\le cE|v_n(t_n+R_n)| \le c(Ev_n^2)^{1/2}(E(t_n+R_n)^2)^{1/2} \\ &\le cmn^{-1-\delta/4}(mn^{-1-\delta/2}+n^{-7\delta/6})^{1/2} \le c(m^{3/2}n^{-(3+\delta)/2}+mn^{-1-5\delta/6}). \end{aligned}$$

Hence we can replace $f_{Q_n w_n}(t)$ in J_{n2} by

(2.21)
$$\begin{aligned}
&Eexp\{it[q_n w(1+t'_n) + q'_n w(1+t'_n+r'_n) + q'_n(t_n+r_n)w'(1+t'_n+r'_n)]\} \\
&+ itE\{v_n w'(1+t'_n)exp\{it[q_n w(1+t'_n) + q'_n w(1+t'_n+r'_n) \\
&\qquad\qquad + q'_n(t_n+r_n)w'(1+t'_n+r'_n)]\}\} \\
&=: f_{n1}(t) + f_{n2}(t).
\end{aligned}$$

Expanding in powers of $itq'_n(t_n+r_n)w'(1+t'_n+r'_n)$ in $f_{n2}(t)$ and estimating the remainder

$$ct^2 E|q'_n|E|v_n(t_n+r_n)| \le ct^2 m^{3/2} n^{-(3+\delta)/2}$$

(cf. (2.20) and (2.12)), we can replace f_{n2} by

$$f_{n3}(t) := itE\{v_n w'(1+t'_n)exp\{it[q_n w(1+t'_n) + q'_n w(1+t'_n+r'_n)]\}\}.$$

Let

$$\begin{aligned}
\xi_1(t) &= E_1\Big(\frac{g'_{1n}}{\sqrt{n}\sigma_g} exp\{it\frac{g'_{1n}}{\sqrt{n}\sigma_g} w(1+t'_n)\}\Big), \\
\xi_2(t) &= E_2\Big(\frac{d_{2n}}{n\sigma_g^2} exp\{it\frac{g'_{2n}}{\sqrt{n}\sigma_g} w(1+t'_n)\}\Big),
\end{aligned}$$

where E_i denotes the conditional expectation given all random variables independent of $X_i, i = 1, 2$. Then we have

$$\begin{aligned}
&|\xi_1(t)| \le c|t|E(g'^2_{1n}/(n\sigma_g^2)) \le c|t|n^{-1}, \\
(2.22)\qquad &|\xi_2(t)| \le c|t|E|(g'_{2n}(g^2_{2n} - Eg^2_{2n} + 2g^*_{2n} - 2Eg^*_{2n})/(n^{3/2}\sigma_g^3)| \\
&\qquad \le c|t|n^{-1-\delta/2},
\end{aligned}$$

and further

$$\begin{aligned}
|f_{n3}(t)| &= |t(m^2-m)E\{\frac{g'_{1n}}{\sqrt{n}\sigma_g}\frac{d_{2n}}{n\sigma_g^2} w'(1+t'_n) \\
&\qquad \cdot exp\{it[q_n w(1+t'_n) + q'_n w(1+t'_n+r'_n)]\}\}| \\
&\le c|t|^3 m^2 n^{-2-\delta/2}.
\end{aligned}$$

Hence, recalling (2.21), $f_{Q_n w_n}(t)$ can be replaced by $f_{n1}(t)$. Noting that by (2.12)

$$\begin{aligned}
E|tq'_n r_n w'(1+t'_n+r'_n)| &\le c|t|(Eq'^2_n)^{1/2}(Er^2_n)^{1/2} \\
&\le c|t|mn^{-1-7\delta/12},
\end{aligned}$$

we can remove the term $itq'_n r_n w'(1+t'_n+r'_n)$ in the exponent of $f_{n1}(t)$. By using $|exp(ix) - 1 - ix| \le |x|^{3/2}$ with $x = tq'_n t_n w'(1+t'_n+r'_n)$ and the bound

$$E|q'_n t_n|^{3/2} \le (Eq'^2_n)^{3/4} E|t_n|^{3/2} \le cmn^{-1-\delta/2},$$

$f_{n1}(t)$ can be replaced by

$$f_{n4}(t) = Eexp\{it[q_n w(1+t'_n) + q'_n w(1+t'_n+r'_n)]\}$$
$$+ itEq'_n t_n w'(1+t'_n+r'_n)exp\{it[q_n w(1+t'_n) + q'_n w(1+t'_n+r'_n)]\}$$
$$=: f_{n5}(t) + f_{n6}(t).$$

Let $\eta(t) = Eexp\{it(g'_{1n}/(\sqrt{n}\sigma_g))\}$ and conditional expectations

$$\zeta(t) = E_1 exp\{it(g'_{1n}/(\sqrt{n}\sigma_g))w(1+t'_n)\},$$

$$\kappa(t) = E_1([d_{1n}/(n\sigma_g^2)]exp\{it(g'_{1n}/(\sqrt{n}\sigma_g))w(1+t'_n)\}).$$

Then we can write

$$f_{n5}(t) = Eexp\{itq'_n w(1+t'_n+r'_n)\}\zeta^m(t),$$

$$f_{n6}(t) = itmEq'_n w'(1+t'_n+r'_n)exp\{itq'_n w(1+t'_n+r'_n)\}\zeta^{m-1}(t)\kappa(t).$$

From (2.22) we have

$$|\kappa(t)| \le c|t|n^{-1-\delta/2}.$$

The function $|\zeta(t)|^2$ is the characteristic function of a symmetric random variable ψ such that

$$n^{-1}w^2(1+t'_n) \le Var\psi \le 2n^{-1}w^2(1+t'_n)$$

and

$$E|\psi|^3 \le 8\sigma_g^{-3}n^{-3/2}E|g'_{1n}|^3|w(1+t'_n)|^3 \le cn^{-1-\delta/2}.$$

Therefore

$$|\zeta(t)| \le exp(-c_1 t^2/n) \qquad \text{for } |t| \le c_2 n^{\delta/2}, \tag{2.23}$$

here $c_i, i = 1, 2, \ldots$, are positive constants. Similarly, we have

$$|\eta(t)| \le exp(-c_3 t^2/n) \qquad \text{for } |t| \le c_4 n^{\delta/2}. \tag{2.24}$$

Then

$$|f_{n6}(t)| \le ct^2 mn^{-1-\delta/2}exp\{-c_1(m-1)t^2/n\}$$
$$\le ct^2 mn^{-1-\delta/2}|t|^{-c_1 C_1} \le cmn^{-1-\delta/2} \qquad \text{for } |t| \le c_2 n^{\delta/2},$$

provided we choose C_1 large enough in the definition of m. Hence f_{n4} can be replaced by f_{n5}. We can assume that $k = m/2$ is an integer and write

$$\zeta^k(t) = E_Y exp\{itYw(1+t'_n)\},$$

where $Y = \sum_{i=1}^k g'_{in}/(\sqrt{n}\sigma_g)$ and E_Y denotes the conditional expectation given all random variables but Y. Expanding, we have

$$\zeta^k(t) = E_Y exp\{itY(1 - \frac{t'_n}{2} + \frac{t'^2_n}{2}w"(1+\theta t'_n))\}$$
$$= Eexp(itY) - \frac{itt'_n}{2}EYexp(itY) + \epsilon_1,$$

where $|\theta| \le 1$,

$$|\epsilon_1| \le c|t|t_n'^2 E|Y| + ct^2(t_n'^2 + t_n'^4)EY^2$$
$$\le c|t|t_n'^2 + ct^2(t_n'^2 + t_n'^4).$$

However

$$\text{(2.25)} \qquad \begin{aligned} Et_n'^2 &\le cn^{-1}E(g_{1n}^4 + g_{1n}^{*2}) \le cn^{-\delta/2}, \\ Et_n'^4 &\le cn^{-3}E(g_{1n}^8 + g_{1n}^{*4}) + cn^{-2}((Eg_{1n}^4)^2 + (Eg_{1n}^{*2})^2) \le cn^{-\delta/2}. \end{aligned}$$

Hence, noting (2.23), which implies $|\zeta^k(t)| \le |t|^{-5}$ by choosing the constant C_1 , we may replace $\zeta^k(t)$ in f_{n5} by

$$Eexp(itY) - \frac{itt_n'}{2}EY exp(itY) = \eta^k(t) - \frac{imtt_n'}{2}\eta^{k-1}(t)\eta_1(t),$$

where $\eta_1(t) = E\{(g_{1n}'/(\sqrt{n}\sigma_g))exp(itg_{1n}'/(\sqrt{n}\sigma_g))\}$ satisfying

$$|\eta_1(t)| \le c|t|Eg_{1n}'^2/(n\sigma_g^2) \le c|t|n^{-1}.$$

Then

$$f_{n5}(t) = \eta^m(t)Eexp\{itq_n'w(1 + t_n' + r_n')\} + \epsilon_2,$$

where, by the symmetry and noting $|\eta^{m-1}(t)| \le |\eta^{m-2}(t)| \le t^{-7}$ provided C_1 is large enough (cf. (2.24)), and

$$\begin{aligned} |\epsilon_2| \le\ & ct^2mn^{-1}|\eta^{m-1}(t)Et_n'exp\{itq_n'w(1 + t_n' + r_n')\}| \\ &+ ct^4m^2n^{-2}|\eta^{m-2}(t)Et_n'^2exp\{itq_n'w(1 + t_n' + r_n')\}| \\ \le\ & c|t|^{-3}\{|Ed_{nn}exp\{itq_n'w(1 + t_n' + r_n')\}| + Et_n'^2\}, \end{aligned}$$

where $d_{nn} = g_{nn}^2 - Eg_{nn}^2 + 2g_{nn}^* - 2Eg_{nn}^*, Et_n'^2 \le cn^{-\delta/2}$ by (2.25) and

$$\begin{aligned} &|Ed_{nn}exp\{itq_n'w(1 + t_n' + r_n')\}| \\ =&|Ed_{nn}\{1 + it\frac{g_{nn}'}{\sqrt{n}\sigma_g}w(1 + t_n' + r_n')\theta\}exp\{itq_{n-1}'w(1 + t_n' + r_n')\}| \\ \le&|Ed_{nn}exp\{itq_{n-1}'[w(1 + t_{n-1}' + r_{n-1}') + ((n\sigma_g^2)^{-1}d_{nn} + r_n' - r_{n-1}') \\ &\cdot w'(1 + t_{n-1}' + r_{n-1}' + \theta((n\sigma_g^2)^{-1}d_{nn} + r_n' - r_{n-1}'))]\}| + c|t|n^{-1/2}E|d_{nn}g_{nn}'| \\ \le&c|t|E|d_{nn}q_{n-1}'(n^{-1}d_{nn} + r_n' - r_{n-1}')| + c|t|n^{-\delta/2} \\ \le&c|t|n^{-\delta/2}, \end{aligned}$$

here we use the fact that $E(r_n' - r_{n-1}')^2 \le cn^{-3}E\varphi_{12n}^4 \le cn^{-1-7\delta/6}$ by the elementary calculation. Then we have

$$|\epsilon_2| \le ct^{-2}n^{-\delta/2}$$

and hence $f_{n5}(t)$ can be replaced by

$$f_{n7}(t) := \eta^m(t)Eexp\{itq_n'w(1 + t_n' + r_n')\}.$$

Write

$$\begin{aligned}
&|Eexp\{itq_n'w(1+t_n'+r_n')\} - exp\{-\frac{1}{2}(1-\frac{m}{n})t^2\}| \\
&\le |Eexp\{itq_n'w(1+t_n'+r_n')\} - Eexp(itq_n')| \\
&\quad + |Eexp(itq_n') - exp\{-\frac{1}{2}(1-\frac{m}{n})t^2\}| \\
&=: I_1 + I_2.
\end{aligned}$$

By the well-known inequality (e.g., Ch 5, Lemma 1 in Petrov 1975),

$$I_2 \le 16(n\sigma_g^2)^{-3/2}(n-m)(1-\frac{m}{n})^{-3/2}E|g_{1n}'|^3 \le cn^{-\delta/2}. \tag{2.26}$$

Consider I_1. Let θ_1 and θ_2 be random variables uniformly distributed on [0, 1], independent of all other random variables, and E_θ stands for the conditional expectation given all random variables but θ_1 and θ_2. Expanding in powers of $t_n' + r_n'$, we obtain

$$Eexp\{itq_n'w(1+t_n'+r_n')\} = Eexp\{itq_n' + itE_\theta q_n'(t_n'+r_n')w'(1+\theta_1(t_n'+r_n'))\}.$$

Estimate $E|E_\theta q_n'(t_n'+r_n')w'(1+\theta_1(t_n'+r_n'))|$. We have

$$E|E_\theta q_n' r_n' w'(1+\theta_1(t_n'+r_n'))| \le cn^{-\delta/2}$$

(cf. (2.15)). Hence it suffices to consider $E|E_\theta q_n' t_n' w'(1+\theta_1(t_n'+r_n'))|$. Similarly to (2.16), write

$$q_n' t_n' = u_n' + v_n', \quad u_n' = n^{-3/2}\sigma_g^{-3}\sum_{i=m+1}^{n} g_{in}' d_{in}.$$

We have $E|u_n'| \le cn^{-\delta/2}$ (cf. (2.17)) and

$$\begin{aligned}
&E|E_\theta v_n' w'(1+\theta_1(t_n'+r_n'))| \\
&= E|E_\theta(-\frac{1}{2}v_n' + v_n'\theta_1(t_n'+r_n')w"(1+\theta_1\theta_2(t_n'+r_n')))| \\
&\le cE|v_n'(t_n'+r_n')| \le c(Ev_n'^2)^{1/2}(E(t_n'+r_n')^2)^{1/2} \le cn^{-\delta/2}
\end{aligned}$$

(cf. (2.18), (2.25) and (2.14)). Hence we obtain

$$E|E_\theta q_n'(t_n'+r_n')w'(1+\theta_1(t_n'+r_n'))| \le cn^{-\delta/2},$$

and further

$$I_1 \le c|t|n^{-\delta/2}.$$

Furthemore, similarly to (2.26), we have

$$|\eta^m(t) - exp(-\frac{m}{2n}t^2)| \le cn^{-\delta/2}.$$

Write

$$|f_{n7}(t)-e^{-t^2/2}|\le|\eta^m(t)|(I_1+I_2)+exp\{-\frac{1}{2}(1-\frac{m}{n})t^2\}|\eta^m(t)-exp(-\frac{m}{2n}t^2)|.$$

Then, combining the above estimations we obtain

$$J_{n2}\le cn^{-\delta/2}. \tag{2.27}$$

At last, we estimite J_{n1}. Let $m_1=[3\sqrt{n}\log n]$. Define

$$D_{n1}=\frac{\sqrt{n}}{2\sigma_g}\binom{n}{2}^{-1}\sum_{m_1\le i<j\le n}\varphi'_{ijn},\quad D_{n2}=D_n-D_{n1}.$$

Using the martingale method, we have

$$\begin{aligned}E|D_{n2}w_n|^{2+\delta}&\le 8E|D_{n2}|^{2+\delta}\le c(n-m_1)^{1+\delta/2}n^{-(2+\delta)}\\&\le cn^{-3(2+\delta)/4}(\log n)^{1+\delta/2},\end{aligned} \tag{2.28}$$

which implies that

$$P(|D_{n2}w_n|>n^{-\delta/2})\le cn^{-\delta/2}.$$

Hence we can replace D_n by D_{n1}.

Write

$$E_n(t):=f_{(Q_n+D_{n1})w_n}(t)-f_{Q_nw_n}(t)=Ee^{iQ_nw_nt}(e^{iD_{n1}w_nt}-1).$$

Instead of $e^{iQ_nw_nt}$ by $e^{iQ_nw_nt}(e^{iD_{n1}w_nt}-1)$ and by noting that $e^{iD_{n1}w_nt}-1$ is a bounded factor, some estimation steps for J_{n1} are similar to that for J_{n2}. We only investigate the steps that need special explanations. Recalling (2.21), we can replace $E_n(t)$ by $\bar{f}_{n1}(t)+\bar{f}_{n2}(t)$, where

$$\begin{aligned}\bar{f}_{n1}(t)&=E\{(e^{iD_{n1}w_nt}-1)exp\{it[q_nw(1+t'_n)+q'_nw((1+t'_n+r'_n)\\&\qquad+q'_n(t_n+r_n)w'(1+t'_n+r'_n)]\}\},\\ \bar{f}_{n2}(t)&=itE\{(e^{iD_{n1}w_nt}-1)v_nw'(1+t'_n)exp\{it[q_nw(1+t'_n)\\&\qquad+q'_nw(1+t'_n+r'_n)+q'_n(t_n+r_n)w'(1+t'_n+r'_n)]\}\}.\end{aligned}$$

We have

$$|\bar{f}_{n2}(t)|\le ct^2E|D_{n1}v_n|\le ct^2(ED_{n1}^2Ev_n^2)^{1/2}\le ct^2mn^{-\frac{3}{2}-\frac{\delta}{4}}.$$

Hence $E_n(t)$ can be replaced by $\bar{f}_{n1}(t)$.

Put $m'=m\vee m_1$ and let $D'_{n1}=\frac{\sqrt{n}}{2\sigma_g}\binom{n}{2}^{-1}\sum_{m'<i<j\le n}\varphi'_{ijn}$ and $D''_{n1}=D_{n1}-D'_{n1}$ if $m_1<m$. It is clear that D'_{n1} is indenpedent of t_n and r_n. We have

$$\begin{aligned}&E(|D'_{n1}||q'_n(t_n+r_n)|^{3/2})\\&\le c(E|D'_{n1}|^{2+\delta})^{\frac{1}{2+\delta}}(E|q'_n|^{\frac{3(2+\delta)}{2(1+\delta)}})^{\frac{1+\delta}{2+\delta}}(E|t_n|^{\frac{3}{2}}+E|r_n|^{\frac{3}{2}})\\&\le cn^{-\frac{1}{2}+(\frac{1}{4}-\frac{\delta}{2})\vee 0}(mn^{-1-\frac{\delta}{2}}+m^2n^{-2-\delta})\le cmn^{-\frac{3+\delta}{2}+(\frac{1}{4}-\frac{\delta}{2})\vee 0}\end{aligned} \tag{2.29}$$

and

(2.30)

$$
\begin{aligned}
&E(|D''_{n1}||q'_n(t_n+r_n)|^{3/2}) \\
&\le (E|D''_{n1}|^{2+\delta})^{\frac{1}{2+\delta}}(E|q'_n|^{\frac{3(2+\delta)}{2(1+\delta)}}(E|t_n|^{\frac{3(2+\delta)}{2(1+\delta)}}+E|r_n|^{\frac{3(2+\delta)}{2(1+\delta)}}))^{\frac{1+\delta}{2+\delta}} \\
&\le c(m^{\frac{2+\delta}{2}}n^{-(2+\delta)})^{\frac{1}{2+\delta}}n^{(\frac14-\frac{\delta}{2})\vee 0}(m^{\frac{3(2+\delta)}{4(1+\delta)}}n^{-\frac{2+\delta}{2}}+m^{\frac{3(2+\delta)}{2(1+\delta)}}n^{-(2+\delta)})^{\frac{1+\delta}{2+\delta}} \\
&\le cm^{\frac54}n^{-\frac{3+\delta}{2}+(\frac14-\frac{\delta}{2})\vee 0}
\end{aligned}
$$

if $m_1 < m$, here the estimate for $E|D''_{n1}|^{2+\delta}$ is similar to that for $E|D_{n2}|^{2+\delta}$. Combining (2.29) and (2.30) implies that $\bar{f}_{n1}(t)$ can be replaced by $\bar{f}_{n3}(t)+\bar{f}_{n4}(t)$, where

$$
\begin{aligned}
\bar{f}_{n3}(t) &= E\{(e^{iD_{n1}w_nt}-1)exp\{it[q_nw(1+t'_n)+q'_nw(1+t'_n+r'_n)]\}\}, \\
\bar{f}_{n4}(t) &= itE\{q'_n(t_n+r_n)w'(1+t'_n+r'_n)(e^{iD_{n1}w_nt}-1)exp\{it[q_nw(1+t'_n) \\
&\qquad + q'_nw(1+t'_n+r'_n)]\}\}.
\end{aligned}
$$

From the facts

$$
\begin{aligned}
&E|D'_{n1}q'_n(t_n+r_n)| \le (ED'^2_{n1}Eq'^2_n)^{1/2}E|t_n+r_n| \\
&\le cn^{-1/2}(mn^{-1}+m^2n^{-2}) \le cmn^{-3/2}, \\
&E|D''_{n1}q'_n(t_n+r_n)| \le (ED''^2_{n1})^{1/2}[(Eq'^2_nEt^2_n)^{1/2}+(Eq'^2_nEr^2_n)^{1/2}] \\
&\le c(mn^{-2})^{1/2}(mn^{-1-\delta/2}+m^2n^{-2-\delta})^{1/2} \le cmn^{-3/2-\delta/4},
\end{aligned}
$$

we can remove $\bar{f}_{n4}(t)$ but the case of $\delta = 1$. When $\delta = 1$, from the above estimates, $\bar{f}_{n4}(t)$ can be replaced by

$$
\begin{aligned}
\bar{f}_{n5}(t) := itE\{&q'_nt_nw'(1+t'_n+r'_n)(e^{iD'_{n1}w_nt}-1) \\
&\cdot exp\{it[q_nw(1+t'_n)+q'_nw(1+t'_n+r'_n)]\}\}
\end{aligned}
$$

Let r''_{np} are R_{np} but with the summation being extended over $\{(i,j), m < i < j \le n\}(p = 1,2,5,6)$ or $\{(i,k,l) : m < i \le n, m < k < l \le n, k \ne i, l \ne i\}(p = 3,4)$ and let $r'''_{np} = r'_{np} - r''_{np}$, e.g.

$$
\begin{aligned}
r'''_{n4} =& \frac{4}{n-2}\binom{n-1}{2}^{-1}\{\sum_{i=1}^{m}\sum_{m+1\le l\le n}^{(i)}\sum_{1\le k\le l-1}^{(i)}\varphi'_{ikn}\varphi'_{iln} \\
&+\sum_{i=m+1}^{n}\sum_{1\le l\le n}^{(i)}\sum_{1\le k\le (l-1)\wedge m}^{(i)}\varphi'_{ikn}\varphi'_{iln}.
\end{aligned}
$$

Let $r''_n = \sum_{p=1}^{6} r''_{np}, r'''_n = \sum_{p=1}^{6} r'''_{np}$. Obviously, r''_n is independent of the σ-field $\sigma(X_1,\ldots,X_m)$. Using the martingale method, we have for $\delta = 1$

(2.31) $$Er'''^2_{n4} \le cn^{-6}(m^2n^2+n^2mn)E\varphi'^4_{12n} \le cmn^{-3}n(1) \le cmn^{-13/6}.$$

For $Er_n'''^2$ we have the same bound. Hence

$$E|q_n'^2 t_n r_n'''| \le (Eq_n'^4 Et_n^2)^{1/2}(Er_n'''^2)^{1/2} \le cmn^{-19/12}.$$

Moreover

$$\begin{aligned}
&E|q_n' t_n r_n'''| \le (Eq_n'^2 Et_n^2)^{1/2}(Er_n'''^2)^{1/2} \le cmn^{-11/6},\\
&E|q_n' t_n D_{n1}'(t_n + r_n + r_n''')|\\
&\le (Eq_n'^2 ED_{n1}'^2)^{1/2}(Et_n^2 + (Et_n^2 Er_n^2)^{1/2}) + (Eq_n'^4 ED_{n1}'^4)^{1/4}(Et_n^2)^{1/2}(Er_n'''^2)^{1/2}\\
&\le cmn^{-2}.
\end{aligned}$$

Thus $\bar{f}_{n5}(t)$ can be replaced by

$$\begin{aligned}
\bar{f}_{n6}(t) := itE\{q_n' t_n w'(1 + t_n' + r_n'')(e^{iD_{n1}' w(1+t_n'+r_n'')t} - 1)exp\{it[q_n w(1 + t_n')\\
+ q_n' w(1 + t_n' + r_n'')]\}\}.
\end{aligned}$$

Then, using the notations in estimating f_{n4} we can write

$$\begin{aligned}
\bar{f}_{n6}(t) =& itmEq_n' w'(1 + t_n' + r_n'')(e^{iD_{n1}' w(1+t_n'+r_n'')t} - 1)\\
&\cdot exp\{itq_n' w(1 + t_n' + r_n'')\}\zeta^{m-1}(t)\kappa(t).
\end{aligned}$$

Recalling the estimate of $f_{n6}(t)$ and noting that $E|D_{n1}'| \le cn^{1/2}$, we can remove $\bar{f}_{n6}(t)$, and further it suffices to consider $\bar{f}_{n3}(t)$. Combining the fact that

$$\begin{aligned}
&(ED_{n1}''^2)^{1/2} \le cm^{1/2}n^{-1}, \quad E|q_n' r_n'''| \le cm^{\frac{1}{2}}n^{-\frac{1}{2}-\frac{7\delta}{12}},\\
&E|D_{n1}'(t_n + r_n + r_n''')| \le c(m^{\frac{1}{2}}n^{-\frac{3}{2}} + mn^{-\frac{3}{2}-\frac{7\delta}{12}} + m^{1/2}n^{-1-\frac{7\delta}{12}}),
\end{aligned}$$

we can replace $\bar{f}_{n3}(t)$ by

$$\begin{aligned}
\bar{f}_{n7}(t) :&= E\{(e^{iD_{n1}' w(1+t_n'+r_n'')t} - 1)exp\{it[q_n w(1 + t_n') + q_n' w(1 + t_n' + r_n'')]\}\}\\
&= E\{(e^{iD_{n1}' w(1+t_n'+r_n'')t} - 1)exp\{itq_n' w(1 + t_n' + r_n'')\}\zeta^m(t)\}
\end{aligned}$$

(cf. $f_{n5}(t)$). Recalling the estimate of $\zeta^m(t)$ and noting $E|D_{n1}'| \le cn^{-1/2}$ we obtain

$$J_{n1} \le cn^{-\delta/2}.$$

Combining it with and (2.27) completes the proof of the theorem.

References

[1] Arvesen, J. N. (1969). Jackkinifing U-statistics. *Ann. Math. Statist.* **40** 2076-2100.

[2] Callaert, H. and Veraverbeke, N. (1981). The order of the normal approximation for a studentized U-statistic. *Ann. Statist.* **9** 194-200.

[3] Callaert, H. and Janssen, P. (1978). The Berry-Esseen theorem for U-statistics. *Ann. Statist.* **6** 417-421.

[4] Cheng, P. (1996). Some inequalities on U-statistics. (preprint).

[5] Petrov, V. V. (1975). *Sums of Independent Random Variables.* Springer, Berlin.

[6] Zhao, L. C. (1983). The rate of the normal approximation for a studentized U-statistics. *Sei. Explo.* **3**(2) 45-52.

Department of Mathematics, Hangzhou University, Hangzhou 310028, P. R. China