

Student U-统计量的 Berry-Esseen 界*

林正炎

(浙江大学西溪校区数学系, 杭州 310028)

摘要 在 $2 + \delta$ 阶矩存在的条件下, 建立了 Student U-统计量的 Berry-Esseen 界 $O(n^{-\delta/2})$.

关键词 Student U-统计量 Berry-Esseen 界

设 $\{X_n, n \geq 1\}$ 是一 i.i.d. 随机变量序列, 具有共同的分布函数 $F(x)$, 又设 $h(x_1, x_2)$ 是一对称函数. 定义 U-统计量

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

Arvesen^[1]研究了 Student U-统计量

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

的正态收敛性, 其中 $\theta = Eh(X_1, X_2)$,

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

引入一致距离

$$\Delta_n = \sup_x |P(\sqrt{n}(U_n - \theta)/S_n < x) - \Phi(x)|, \quad n > 2,$$

其中 $\Phi(x)$ 是标准正态分布函数, Callaert 等人^[2]在 $E|h(X_1, X_2)|^{4.5} < \infty$ 的条件下建立了 Berry-Esseen 界 $O(n^{-1/2})$; 赵林城^[3]减弱上述矩条件到 $E(X_1, X_2)^4 < \infty$. 这个条件还能进一步减弱吗? 本文证明这是可能的, 而且得到了一个最佳的结论, 即, 如果对某个 $\delta > 0$, 条件 $E|h(X_1, X_2)|^{2+\delta} < \infty$ 成立, 那么就有 Berry-Esseen 界 $\Delta_n = O(n^{-\delta/2})$.

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

定理 假设对某个 $0 < \delta \leq 1$, $E|h(X_1, X_2)|^{2+\delta} < \infty$ 且 $Eg(X_1)^2 > 0$, 那么,

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

为了证明我们的定理, 需要下列引理:

引理 设 H_n 是一核为 $u(x_1, x_2)$ 的 j 阶 ($j = 1$ 或 2) U-统计量, 满足 $Eu(X_1, X_2) = 0$, 那么对 $1 \leq r < 2$,

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

对 $r \geq 2$,

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

其中 c_r 和 c'_r 是仅与 r 有关的正常数.

这是陈希孺^[4]的一个结论(对 $r \geq 2$ 的情形, Serfling^[5]也提到上述结果).

不失一般性, 可设 $\theta = 0$. 记

$$\begin{aligned} \varphi(X_i, X_j) &= h(X_i, X_j) - g(X_i) - g(X_j), \\ g^*(X_i) &= E[g(X_j)\varphi(X_i, X_j) | X_i]. \end{aligned}$$

为简单, 记

$$h_{ij} = h(X_i, X_j), g_i = g(X_i), \varphi_{ij} = \varphi(X_i, X_j), g_i^* = g^*(X_i), \sigma_g^2 = E g_1^2.$$

根据文献[2]中关于 S_n^2 的分解, 可写

$$S_n^2 = 4\sigma_g^2(1 + T_n + R_n), \quad (1)$$

其中

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

而

$$\begin{aligned} R_{n1} &= -4 \binom{n}{2}^{-1} \sum_{i < j} g_i g_j, \quad R_{n2} = 4 \binom{n}{2}^{-1} \sum_{i < j} [(g_i + g_j)\varphi_{ij} - g_i^* - g_j^*], \\ R_{n3} &= -\frac{8}{n} \sum_{i=1}^n \left[g_i \binom{n-1}{2}^{-1} \sum_{k < l}^{(i)} \varphi_{kl} \right], \\ R_{n4} &= \frac{4}{n-2} \sum_{i=1}^n \left[\binom{n-1}{2}^{-1} \sum_{k < l}^{(i)} \varphi_{ik} \varphi_{il} \right], \\ R_{n5} &= -\frac{4n(n-1)}{(n-2)^2} \left[\binom{n}{2}^{-1} \sum_{i < j} \varphi_{ij} \right]^2, \quad R_{n6} = \frac{4n}{(n-2)^2} \binom{n}{2}^{-1} \sum_{i < j} \varphi_{ij}^2, \end{aligned}$$

其中 $\sum_{i < j}$ 和 $\sum_{k < l}^{(i)}$ 分别表示 $\sum_{1 \leq i < j \leq n}$ 和 $\sum_{1 \leq k < l \leq n, k \neq i, l \neq i}$. 于是有

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

注意到 $n\sigma_g^2 T_n$ 是均值为 0 的 i.i.d. 随机变量之和, 有

$$P\left(|T_n| > \frac{1}{4}\right) \leq 8E|T_n|^{1+\delta/2} \leq cn^{-\delta/2}, \quad (3)$$

此处以及以后, c 表示一正常数, 不同处可取不同的值.

对于 R_n , 只考察 R_{n4} . 对于其他的项可完全类似地进行. 注意到 $\binom{n-1}{2}^{-1} \sum_{k < l}^{(i)} \varphi_{ik} \varphi_{il}$ 是一个退化的 U-统计量, 利用引理(取 $j=2, r=1+(\delta/2)$), 有

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

对于其他的 R_{ni} , 我们有相同的估计, 因此 $E|R_n| \leq cn^{-\delta/2}$. 进一步又有

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

令 w 是 $\mathbb{R} \rightarrow \mathbb{R}$ 的函数, 它具有有界导数, 无限次可微且满足

$$\frac{2}{3} \leq w(x) \leq 2; \quad w(x) = x^{-1/2}, \quad \frac{1}{2} \leq |x| \leq \frac{3}{2}.$$

由(3)和(4)式, 可考虑 $U_n^* := \frac{\sqrt{n}U_n}{2\sigma_g} w(1 + T_n + R_n)$ 代替 $\sqrt{n}U_n/S_n$. 记

$$g_{in} = g_i I(|g_i| \leq \sqrt{n}), \quad g_{in}^* = g_i^* I(|g_i^*| \leq n), \\ \varphi_{ijn} = \varphi_{ij} I(|\varphi_{ij}| \leq n(\delta)),$$

其中 $n(\delta) = n^{(4+\delta)/2(2+\delta)}$. 因为

$$P(\bigcup_{i=1}^n \{g_i \neq g_{in}\}) \leq nP(|g_1| > \sqrt{n}) \leq n^{-\delta/2} E|g_1|^{2+\delta}, \quad (5)$$

所以在 T_n 和 R_n 中可以 g_{in} 代替 g_i . 类似地可以用 g_{in}^* 和 φ_{ijn} 分别代替 g_i^* 和 φ_{ij} . 此外

$$|Eg_{1n}| = |Eg_1 I(|g_1| > \sqrt{n})| \leq n^{-(1+\delta)/2} E|g_1|^{2+\delta}, \quad (6)$$

$$|E\varphi_{12n}| = |E\varphi_{12} I(|\varphi_{12}| > n(\delta))| \leq n^{-\frac{(1+\delta)(4+\delta)}{2(2+\delta)}} E|\varphi_{12}|^{2+\delta}; \quad (7)$$

类似地

$$E(g_1 - g_{1n})^2 \leq n^{-\delta/2} E|g_1|^{2+\delta}, \quad |E(g_1^* - g_{1n}^*)| \leq n^{-\delta/2} E|g_1^*|^{1+\delta/2}. \quad (8)$$

进一步记 $g'_{in} = g_{in} - Eg_{in}$, $\varphi'_{ijn} = \varphi_{ijn} - E\varphi_{ijn}$, $\bar{g}_{in} = E(g'_{jn}\varphi'_{ijn} | X_i)$. 由(5)~(8)式, 可以修改 T_n 的定义为

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

类似地也可以 g'_{in} , φ'_{ijn} 和 $\bar{g}_{in}$ 分别代替 g_i , φ_{ij} 和 g_i^* 去定义 R_n . 记

$$w_n = w(1 + T_n + R_n), \quad Q_n = \frac{1}{\sqrt{n}\sigma_g} \sum_{i=1}^n g'_{in}, \quad D_n = \frac{\sqrt{n}}{2\sigma_g} \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} \varphi'_{ijn}.$$

可以考虑以 $U'_n := (Q_n + D_n)w_n$ 代替 U_n^* .

用 f_X 表示 X 的特征函数. 为了证明 Berry-Esseen 不等式, 只需证明对任意给定的 $\varepsilon > 0$,

$$J_n := \int_{C \leq |t| \leq \varepsilon n^{\delta/2}} |t|^{-1} |f_{U'_n}(t) - e^{-t^2/2}| dt \leq cn^{-\delta/2}, \quad (9)$$

其中 $C > 0$ 是常数. 写

$$J_n \leq \int_{C \leq |t| \leq \varepsilon n^{\delta/2}} |t|^{-1} |f_{(Q_n+D_n)w_n}(t) - f_{Q_n w_n}(t)| dt \\ + \int_{C \leq |t| \leq \varepsilon n^{\delta/2}} |t|^{-1} |f_{Q_n w_n}(t) - e^{-t^2/2}| dt =: J_{n1} + J_{n2}. \quad (10)$$

首先估计 J_{n2} . 记

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

其中 C_1 是一个足够大的常数. 通过对常数的选择, 可以假设 $2 \leq m \leq n/2$. 令 $q_n =$

$\frac{1}{\sqrt{n}\sigma_g} \sum_{i=1}^m g'_{in}$, $q'_n = Q_n - q_n$, $t'_n = \frac{1}{n\sigma_g^2} \sum_{i=1}^m d_{in}$, $t'_n = T_n - t_n$. 进一步, 对 $p=1, \dots, 6$, 令 r_{np} 是 R_{np} , 但是其中的求和范围做如下的改变: 对 $p=1, 2, 5, 6$, 在 $\{(i, j), 1 \leq i < j \leq m\}$ 上求和, 对 $p=3, 4$, 在 $\{(i, k, l), 1 \leq i \leq m, 1 \leq k < l \leq m, k \neq i, l \neq i\}$ 上求和. 又令 $r'_{np} = R_{np} - r_{np}$, 并记 $r_n = \sum_{p=1}^6 r_{np}$, $r'_n = \sum_{p=1}^6 r'_{np}$. 考察 $f_{Q_n w_n}(t) = E \exp(itQ_n w_n)$. 写

$$\begin{aligned} Q_n w_n &= q_n w(1 + t'_n) + q_n(t_n + R_n)w'(1 + t'_n + \theta(t_n + R_n)) \\ &\quad + q'_n w(1 + t'_n + r'_n) + q'_n(t_n + r_n)w'(1 + t'_n + r'_n) \\ &\quad + q'_n(t_n + r_n)^2 w''(1 + t'_n + r'_n + \theta(t_n + r_n))/2, \end{aligned} \quad (11)$$

其中 $|\theta| \leq 1$. 有

$$E t_n^2 \leq c n^{-2} m E(g_{1n}^4 + g_{1n}^{*2}) \leq c m n^{-1-\delta/2}.$$

我们来估计 $E r_n^2$. 作为一个例子, 考察 $E r_{n4}^2$. 利用引理,

$$E r_{n4}^2 \leq c n^{-4} m^2 E \varphi_{12n}^4 \leq c m^2 n^{-4} n(\delta)^{2-\delta} = c m^2 n^{-2-\frac{\delta(6+\delta)}{2(2+\delta)}}. \quad (12)$$

注意到 q'_n 和 $t_n + r_n$ 是独立的, 而且 w 的导数是有界的, 我们有

$$E |q'_n(t_n + r_n)^2 w''(1 + t'_n + r'_n + \theta(t_n + r_n))| \leq c(E q_n'^2)^{1/2} E(t_n + r_n)^2 \leq c m n^{-1-\delta/2}. \quad (13)$$

由于 $|t| m n^{-1-(\delta/2)} = O(|t|^{-1}(\ln |t|) n^{-\delta/2})$, 因此可以在 $i t Q_n w_n$ 的展开式中移去项 $i t q'_n(t_n + r_n)^2 w''(1 + t'_n + r'_n + \theta(t_n + r_n))/2$ (参见 (11) 式). 下面来处理 (11) 式右边的第 2 项. 为此估计 $E |R_n|^{(2+\delta)/(1+\delta)}$. 作为一个例子, 考察 R_{n4} . 由引理,

$$\begin{aligned} E |R_{n4}|^{(2+\delta)/(1+\delta)} &\leq c n^{-2/(1+\delta)} E |\varphi'_{12n}|^{2(2+\delta)/(1+\delta)} \\ &\leq c n^{-\frac{2}{1+\delta} n(\delta)^{\frac{(1-\delta)(2+\delta)}{1+\delta}}} = c n^{-\frac{\delta(3+\delta)}{2(1+\delta)}}. \end{aligned} \quad (14)$$

由此得

$$\begin{aligned} E |q_n R_n w'(1 + t'_n + \theta(t_n + R_n))| \\ \leq c(E |q_n|^{2+\delta})^{\frac{1}{2+\delta}} (E |R_n|^{\frac{2+\delta}{1+\delta}})^{\frac{1+\delta}{2+\delta}} \leq c m^{\frac{1}{2}} n^{-\frac{1}{2}-\frac{\delta(3+\delta)}{2(2+\delta)}}. \end{aligned} \quad (15)$$

将和式 $q_n t_n$ 分解为对角线和非对角线两部分:

$$q_n t_n = u_n + v_n, \quad u_n = n^{-3/2} \sigma_g^{-3} \sum_{i=1}^m g'_{in} d_{in}. \quad (16)$$

有

$$E |u_n| \leq c n^{-3/2} m (E |g_{1n}|^3 + E |g_{1n} g_{1n}^*|) \leq c m n^{-1-\delta/2}. \quad (17)$$

因此可以在 $f_{Q_n w_n}(t)$ 的展开式中移去 $i t u_n w'(1 + t'_n + \theta(t_n + R_n))$. 关于 v_n , 有

$$E v_n^2 \leq c n^{-3} m^2 (E g_{1n}^2) E (g_{1n}^2 + g_{1n}^*)^2 \leq c m^2 n^{-2-\delta/2}. \quad (18)$$

写

$$v_n w'(1 + t'_n + \theta(t_n + R_n)) = v_n w'(1 + t'_n) + \tau. \quad (19)$$

注意到 $E R_n^2 \leq c n^{-7\delta/6}$, 对上式中的 τ 成立

$$\begin{aligned} E |\tau| &\leq c E |v_n(t_n + R_n)| \leq c (E v_n^2)^{1/2} (E(t_n + R_n)^2)^{1/2} \\ &\leq c m n^{-1-(\delta/4)} (m n^{-1-(\delta/2)} + n^{-7\delta/6})^{1/2} \\ &\leq c (m^{3/2} n^{-(3+\delta)/2} + m n^{-1-(5\delta/6)}). \end{aligned} \quad (20)$$

因此可以用下式代替 J_{n2} 中的 $f_{Q_n w}(t)$:

$$\begin{aligned} & \text{Eexp} \{it[q_n w(1+t'_n) + q'_n w(1+t'_n+r'_n) + q'_n(t_n+r_n)w'(1+t'_n+r'_n)]\} \\ & + it \text{E} \{v_n w'(1+t'_n) \exp \{it[q_n w(1+t'_n) + q'_n w(1+t'_n+r'_n) \\ & + q'_n(t_n+r_n)w'(1+t'_n+r'_n)]\}\} =: f_{n1}(t) + f_{n2}(t). \end{aligned} \quad (21)$$

展开 $f_{n2}(t)$ 中的指数为 $itq'_n(t_n+r_n)w'(1+t'_n+r'_n)$ 的因子, 相应的剩余项的界是

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

(参见(20)和(12)式), 因此可以用

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

代替 $f_{n2}(t)$. 令

$$\begin{aligned} \xi_1(t) &= \left(\text{E}_1 \frac{g'_{1n}}{\sqrt{n}\sigma_g} \exp \left\{ it \frac{g'_{1n}}{\sqrt{n}\sigma_g} w(1+t'_n) \right\} \right), \\ \xi_2(t) &= \text{E}_2 \left(\frac{d_{2n}}{n\sigma_g^2} \exp \left\{ it \frac{g'_{2n}}{\sqrt{n}\sigma_g} w(1+t'_n) \right\} \right), \end{aligned}$$

其中 E_i 表示给定所有与 $X_i (i=1, 2)$ 独立的随机变量的条件下的条件期望. 我们有

$$\begin{aligned} |\xi_1(t)| &\leq c |t| \text{E} (g_{1n}^2 / (n\sigma_g^2)) \leq c |t| n^{-1}, \\ |\xi_2(t)| &\leq c |t| \text{E} |g'_{2n} (g_{2n}^2 - \text{E} g_{2n}^2 + 2g_{2n}^* - 2\text{E} g_{2n}^*) / (n^{3/2} \sigma_g^3)| \\ &\leq c |t| n^{-1-(\delta/2)}, \end{aligned} \quad (22)$$

进一步

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

因此并回顾(21)式, $f_{Q_n w}(t)$ 可以用 $f_{n1}(t)$ 替代. 因由(12)式成立着

$$\text{E} |tq'_n r_n w'(1+t'_n+r'_n)| \leq c |t| (\text{E} (q_n'^2)^{1/2}) (\text{E} r_n^2)^{1/2} \leq c |t|^3 m n^{-1-7\delta/12},$$

故可以在 $f_{n1}(t)$ 的指数中移去项 $itq'_n r_n w'(1+t'_n+r'_n)$. 利用不等式 $|\exp(ix) - 1 - ix| \leq |x|^{3/2}$

于 $x = tq'_n t_n w'(1+t'_n+r'_n)$, 并注意到

$$\text{E} |q'_n t_n|^{3/2} \leq (\text{E} q_n'^2)^{3/4} \text{E} |t_n|^{3/2} \leq c m n^{-1-(\delta/2)},$$

$f_{n1}(t)$ 可以用 $f_{n4}(t)$ 代替:

$$\begin{aligned} f_{n4}(t) &= \text{Eexp} \{it[q_n w(1+t'_n) + q'_n w(1+t'_n+r'_n)]\} \\ &\quad + it \text{E} q'_n t_n w(1+t'_n+r'_n) \exp \{it[q_n w(1+t'_n) + q'_n w(1+t'_n+r'_n)]\} \\ &=: f_{n5}(t) + f_{n6}(t). \end{aligned}$$

记 $\eta(t) = \text{Eexp} \{it(g'_{1n}/(\sqrt{n}\sigma_g))\}$ 并令条件期望

$$\zeta(t) = \text{E}_1 \exp \{it(g'_{1n}/(\sqrt{n}\sigma_g)) w(1+t'_n)\},$$

$$\kappa(t) = \text{E}_1 ([d_{1n}/(n\sigma_g^2)] \exp \{it(g'_{1n}/(\sqrt{n}\sigma_g)) w(1+t'_n)\}).$$

于是可写

$$\begin{aligned} f_{n5}(t) &= \text{Eexp}\{itq'_n w(1+t'_n+r'_n)\} \zeta^m(t), \\ f_{n6}(t) &= itm \text{E} q'_n w'(1+t'_n+r'_n) \exp\{itq'_n w(1+t'_n+r'_n)\} \zeta^{m-1}(t) \kappa(t). \end{aligned}$$

由(22)式有

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

函数 $|\zeta(t)|^2$ 是某一对称随机变量 ψ 的特征函数, ψ 满足关系

$$\begin{aligned} n^{-1}w^2(1+t'_n) &\leq \text{Var}\psi \leq 2n^{-1}w^2(1+t'_n), \\ \text{E}|\psi|^3 &\leq 8\sigma_g^{-3}n^{-3/2}\text{E}|g'_{1n}|^3|w(1+t'_n)|^3 \leq cn^{-1-(\delta/2)}. \end{aligned}$$

因此对 $|t| \leq c_1 n^{\delta/2}$,

$$|\zeta(t)| \leq \exp(-c_2 t^2/n), \quad (23)$$

此处及下文中的 $c_i (i=1, 2, \dots)$ 是正常数. 类似地有

$$|\eta(t)| \leq \exp(-c_3 t^2/n), \quad |t| \leq c_4 n^{\delta/2}. \quad (24)$$

于是对 $|t| \leq c_1 n^{\delta/2}$, 只要在 m 的定义中的 c_1 足够大, 就有

$$\begin{aligned} |f_{n6}(t)| &\leq ct^2 mn^{-1-(\delta/2)} \exp\{-c_2(m-1)t^2/n\} \\ &\leq ct^2 mn^{-1-(\delta/2)} |t|^{c_2 c_1} \leq cmn^{-1-(\delta/2)}. \end{aligned}$$

因此 f_{n4} 可以用 f_{n5} 代替. 不妨假设 $k = m/2$ 是一个整数, 并写

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

其中 $Y = \sum_{i=1}^k g'_{in}/(\sqrt{n}\sigma_g)$, E_Y 表示在给定除了 Y 之外的所有随机变量的条件下的条件期望. 利用 Taylor 公式有

$$\begin{aligned} \zeta^k(t) &= \text{E}_Y \exp\left\{itY\left(1 - \frac{t'_n}{2} + \frac{t_n'^2}{2} w''(1+\theta'_n)\right)\right\} \\ &= \text{Eexp}(itY) - \frac{itt'_n}{2} \text{E}_Y \exp(itY) + \varepsilon_1, \end{aligned}$$

其中 $|\theta| \leq 1$,

$$|\varepsilon_1| \leq c|t|t_n'^2 \text{E}|Y| + ct^2(t_n'^2 + t_n'^4) \text{E}Y^2 \leq c|t|t_n'^2 + ct^2(t_n'^2 + t_n'^4).$$

而

$$\begin{aligned} \text{E}t_n'^2 &\leq cn^{-1} \text{E}(g_{1n}^4 + g_{1n}^{*2}) \leq cn^{-\delta/2}, \\ \text{E}t_n'^4 &\leq cn^{-3} \text{E}(g_{1n}^8 + g_{1n}^{*4}) + cn^{-2}((\text{E}g_{1n}^4)^2 + (\text{E}g_{1n}^{*2})^2) \leq cn^{-\delta/2}. \end{aligned} \quad (25)$$

此外由(23)式, 通过适当选取常数 c_1 , 有 $|\zeta^k(t)| \leq |t|^{-5}$. 因此, 可以用

$$\text{Eexp}(itY) - \frac{itt'_n}{2} \text{E}_Y \exp(itY) = \eta^k(t) - \frac{imut'_n}{2} \eta^{k-1}(t) \eta_1(t)$$

代替 f_{n5} 中的 $\zeta^k(t)$, 其中 $\eta_1(t) = \text{E}\{(g'_{1n}/(\sqrt{n}\sigma_g)) \exp(itg'_{1n}/(\sqrt{n}\sigma_g))\}$ 满足

$$|\eta_1(t)| \leq c|t| \text{E}g_{1n}'^2/(n\sigma_g^2) \leq c|t|n^{-1}.$$

于是可写

$$f_{n5}(t) = \eta^m(t) \text{Eexp}\{itq'_n w(1+t'_n+r'_n)\} + \varepsilon_2.$$

对其中的 ε_2 , 由对称性并注意到通过选取足够大的 c_1 , $|\eta^{m-1}(t)| \leq |\eta^{m-2}(t)| \leq t^{-7}$ (参见

(24)式),有

$$\begin{aligned} |\varepsilon_2| &\leq ct^2 mn^{-1} |\eta^{m-1}(t) E t'_n \exp\{it'_n q'_n w(1+t'_n+r'_n)\}| \\ &\quad + ct^4 m^2 n^{-2} |\eta^{m-2}(t) E t_n'^2 \exp\{it'_n q'_n w(1+t'_n+r'_n)\}| \\ &\leq c|t|^{-3} \{ |E d_{nn} \exp\{it'_n q'_n w(1+t'_n+r'_n)\}| + E t_n'^2 \}, \end{aligned}$$

其中 $d_{nn} = g_{nn}^2 - E g_{nn}^2 + 2g_{nn}^* - 2E g_{nn}^*$. 由(25)式, $E t_n'^2 \leq cn^{-\delta/2}$, 又

$$\begin{aligned} &|E d_{nn} \exp\{it'_n q'_n w(1+t'_n+r'_n)\}| \\ &= \left| E d_{nn} \left\{ 1 + it \frac{g'_{nn}}{\sqrt{n}\sigma_g} w(1+t'_n+r'_n)\theta \right\} \exp\{it'_{n-1} q'_{n-1} w(1+t'_{n-1}+r'_{n-1})\} \right| \\ &\leq |E d_{nn} \exp\{it'_{n-1} q'_{n-1} [w(1+t'_{n-1}+r'_{n-1}) + ((n\sigma_g^2)^{-1} d_{nn} + r'_n - r'_{n-1}) \\ &\quad \cdot w'(1+t'_{n-1}+r'_{n-1} + \theta((n\sigma_g^2)^{-1} d_{nn} + r'_n - r'_{n-1}))]\}| + c|t|n^{-1/2} E |d_{nn} g'_{nn}| \\ &\leq c|t| E |d_{nn} q'_{n-1} (n^{-1} d_{nn} + r'_n - r'_{n-1})| + c|t|n^{-\delta/2} \\ &\leq c|t|n^{-\delta/2}, \end{aligned}$$

这里利用了下列不等式: $E(r'_n - r'_{n-1})^2 \leq cn^{-3} E \varphi_{12n}^4 \leq cn^{-1-(7\delta/6)}$, 它可通过初等计算证明. 因此有

$$|\varepsilon_2| \leq ct^{-2} n^{-\delta/2},$$

且由此 $f_{n5}(t)$ 可用

$$f_{n7}(t) := \eta^m(t) E \exp\{it'_n q'_n w(1+t'_n+r'_n)\}$$

代替.

写

$$\begin{aligned} &\left| E \exp\{it'_n q'_n w(1+t'_n+r'_n)\} - \exp\left\{-\frac{1}{2}\left(1-\frac{m}{n}\right)t^2\right\} \right| \\ &\leq |E \exp\{it'_n q'_n w(1+t'_n+r'_n)\} - E \exp(it'_n q'_n)| + \left| E \exp(it'_n q'_n) - \exp\left\{-\frac{1}{2}\left(1-\frac{m}{n}\right)t^2\right\} \right| \\ &=: I_1 + I_2. \end{aligned}$$

由熟知的不等式(例如参见文献[6]第5章引理1)知

$$I_2 \leq 16(n\sigma_g^2)^{-3/2}(n-m) \left(1-\frac{m}{n}\right)^{-3/2} E |g'_{1n}|^3 \leq cn^{-\delta/2}. \quad (26)$$

考察 I_1 . 令 θ_1 和 θ_2 是与其他一切随机变量都独立的 $[0,1]$ 上均匀分布的随机变量, E_θ 表示给定除了 θ_1 和 θ_2 之外的所有随机变量条件下的条件期望. 由 Taylor 公式可得

$$E \exp\{it'_n q'_n w(1+t'_n+r'_n)\} = E \exp\{it'_n q'_n + it E_\theta q'_n(t'_n+r'_n) w'(1+\theta_1(t'_n+r'_n))\}.$$

我们来估计 $E |E_\theta q'_n(t'_n+r'_n) w'(1+\theta_1(t'_n+r'_n))|$. 因有

$$E |E_\theta q'_n r'_n w'(1+\theta_1(t'_n+r'_n))| \leq cn^{-\delta/2}$$

(参见(15)式), 所以只需考虑 $E |E_\theta q'_n t'_n w'(1+\theta_1(t'_n+r'_n))|$. 类似于(16)式, 可写

$$q'_n t'_n = u'_n + v'_n, \quad u'_n = n^{-3/2} \sigma_g^{-3} \sum_{i=m+1}^n g'_{in} d_{in}.$$

有 $E |u'_n| \leq cn^{-\delta/2}$ (参见(17)式)和

$$\begin{aligned} & E | E_{\theta} v'_n w' (1 + \theta_1 (t'_n + r'_n)) | \\ &= E \left| E_{\theta} \left(-\frac{1}{2} v'_n + v'_n \theta_1 (t'_n + r'_n) w'' (1 + \theta_1 \theta_2 (t'_n + r'_n)) \right) \right| \\ &\leq c E | v'_n (t'_n + r'_n) | \leq c (E v_n'^2)^{1/2} (E (t'_n + r'_n)^2)^{1/2} \leq c n^{-\delta/2} \end{aligned}$$

(参见(18)、(25)和(14)式), 因此得证

$$E | E_{\theta} q'_n (t'_n + r'_n) w' (1 + \theta_1 (t'_n + r'_n)) | \leq c n^{-\delta/2},$$

进一步又有

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

此外, 类似于(26)式,

$$\left| \eta^m(t) - \exp\left(-\frac{m}{2n} t^2\right) \right| \leq c n^{-\delta/2}.$$

令

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

综合上面的估计得

$$J_{n2} \leq c n^{-\delta/2}. \quad (27)$$

最后估计 J_{n1} . 令 $m_1 = [3\sqrt{n} \log n]$. 定义

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

利用鞅论方法有

$$\begin{aligned} E | D_{n2} w_n |^{2+\delta} &\leq 8E | D_{n2} |^{2+\delta} \leq c(n - m_1)^{1+(\delta/2)} n^{-(2+\delta)} \\ &\leq c n^{-3(2+\delta)/4} (\log n)^{1+(\delta/2)}. \end{aligned} \quad (28)$$

由此即有 $P(|D_{n2} w_n| > n^{-\delta/2}) \leq c n^{-\delta/2}$. 因此可以 D_{n1} 代替 D_n .

写

$$E_n(t) := f_{(Q_n + D_{n1})w_n}(t) - f_{Q_n w_n}(t) = E e^{iQ_n w_n t} (e^{iD_{n1} w_n t} - 1).$$

用 $e^{iQ_n w_n t} (e^{iD_{n1} w_n t} - 1)$ 代替 $e^{iQ_n w_n t}$, 并注意到 $e^{iD_{n1} w_n t} - 1$ 是一个有界的因子, 对 J_{n1} 的估计的某些步骤是与对 J_{n2} 的估计的相应步骤类似的, 只考虑需要特别说明的那些步骤. 回顾(21)式, 可以用 $\tilde{f}_{n1}(t)$ 和 $\tilde{f}_{n2}(t)$ 代替 $E_n(t)$, 其中

$$\begin{aligned} \tilde{f}_{n1}(t) &= E \{ (e^{iD_{n1} w_n t} - 1) \exp \{ i t [q_n w (1 + t'_n) + q'_n w (1 + t'_n + r'_n) \\ &\quad + q'_n (t_n + r_n) w' (1 + t'_n + r'_n)] \} \}, \\ \tilde{f}_{n2}(t) &= i t E \{ (e^{iD_{n1} w_n t} - 1) v_n w' (1 + t'_n) \exp \{ i t [q_n w (1 + t'_n) \\ &\quad + q'_n w (1 + t'_n + r'_n) + q'_n (t_n + r_n) w' (1 + t'_n + r'_n)] \} \}. \end{aligned}$$

有

$$|\tilde{f}_{n2}(t)| \leq c t^2 E |D_{n1} v_n| \leq c t^2 (E D_{n1}^2 E v_n^2)^{1/2} \leq c t^2 m n^{-\frac{3}{2} - \frac{\delta}{4}}.$$

因此可以用 $\tilde{f}_{n1}(t)$ 代替 $E_n(t)$.

记 $m' = m \vee m_1$, 且当 $m_1 < m$ 时, 令 $D'_{n1} = \frac{\sqrt{n}}{2\sigma_g} \binom{n}{2}^{-1} \sum_{m' < i < j \leq n} \varphi'_{ijn}$ 和 $D'_{n2} = D_{n1} - D'_{n1}$. 显然

D'_{n1} 与 t_n 和 r_n 独立. 我们有

$$\begin{aligned} & E(|D'_{n1}| | q'_n(t_n + r_n) |^{3/2}) \\ & \leq c(E(|D'_{n1}|^{2+\delta})^{\frac{1}{2+\delta}}(E|q'_n|^{\frac{3(2+\delta)}{2(1+\delta)}})^{\frac{1+\delta}{2+\delta}}(E|t_n|^{\frac{3}{2}} + E|r_n|^{\frac{3}{2}})) \\ & \leq cn^{-\frac{1}{2} + (\frac{1}{4} - \frac{\delta}{2})} \vee 0 (mn^{-1-\frac{\delta}{2}} + m^2 n^{-2-\delta}) \leq cmn^{-\frac{3+\delta}{2} + (\frac{1}{4} - \frac{\delta}{2})} \vee 0, \end{aligned} \quad (29)$$

$$\begin{aligned} & E(|D''_{n1}| | q'_n(t_n + r_n) |^{3/2}) \\ & \leq (E(|D''_{n1}|^{2+\delta})^{\frac{1}{2+\delta}}(E|q'_n|^{\frac{3(2+\delta)}{2(1+\delta)}})(E|t_n|^{\frac{3(2+\delta)}{2(1+\delta)}} + E|r_n|^{\frac{3(2+\delta)}{2(1+\delta)}}))^{\frac{1+\delta}{2+\delta}} \\ & \leq c(m^{\frac{2+\delta}{2}} n^{-(2+\delta)})^{\frac{1}{2+\delta}} n^{(\frac{1}{4} - \frac{\delta}{2})} \vee 0 (m^{\frac{3(2+\delta)}{4(1+\delta)}} n^{-\frac{2+\delta}{2}} + m^{\frac{3(2+\delta)}{2(1+\delta)}} n^{-(2+\delta)})^{\frac{1+\delta}{2+\delta}} \\ & \leq cm^{\frac{5}{4}} n^{-\frac{3+\delta}{2} + (\frac{1}{4} - \frac{\delta}{2})} \vee 0, \end{aligned} \quad (30)$$

其中对于 $E|D''_{n1}|^{2+\delta}$ 的估计类似于对 $E|D_{n2}|^{2+\delta}$ 的估计. 结合(29)和(30)式可知, $\bar{f}_{n1}(t)$ 可以 $\bar{f}_{n3}(t) + \bar{f}_{n4}(t)$ 代替, 其中

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

由不等式

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

对 $\delta < 1$ 的情形, 我们可以移去 $\bar{f}_{n4}(t)$. 当 $\delta = 1$ 时, 从上面的估计, 可以用

$$\begin{aligned} \bar{f}_{n5}(t) &:= itE\{q'_n t_n w'(1+t'_n + r'_n)(e^{iD'_{n1}w_n t} - 1) \\ & \quad \cdot \exp\{it[q_n w(1+t'_n) + q'_n w(1+t'_n + r'_n)]\}\} \end{aligned}$$

代替 $\bar{f}_{n4}(t)$. 令 r''_{np} 是 R_{np} , 但是其中的求和范围做如下的改变: 对 $p = 1, 2, 5, 6$, 在 $\{(i, j), m < i < j \leq n\}$ 上求和; 对 $p = 3, 4$, 在 $\{(i, k, l), m < i \leq n, m < k < l \leq n, k \neq i, l \neq i\}$ 上求和. 又令 $r'''_{np} = r'_{np} - r''_{np}$, 例如

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

令 $r''_n = \sum_{p=1}^6 r''_{np}$, $r'''_n = \sum_{p=1}^6 r'''_{np}$. 显然 r''_n 与 σ 域 $\sigma(X_1, \dots, X_m)$ 是独立的. 利用鞅论方法对于 $\delta = 1$ 的情形有

$$Er''^2_{n4} \leq cn^{-6}(m^2 n^2 + n^2 mn)E\varphi'^4_{12n} \leq cmn^{-3}n(1) \leq cmn^{-13/6}. \quad (31)$$

对于 Er''^2_{n5} 有相同的界. 因此

$$E|q'^2_n t_n r'''_n| \leq (Eq'^4_n Et^2_n)^{1/2}(Er''^2_n)^{1/2} \leq cmn^{-19/12}.$$

此外

$$\begin{aligned}
E |q'_n t_n r'''_n| &\leq (E q'^2_n E t_n^2)^{1/2} (E r'''^2_n)^{1/2} \leq cmn^{-11/6}, \\
E |q'_n t_n D'_{n1}(t_n + r_n + r'''_n)| \\
&\leq (E q'^2_n E D'^2_{n1})^{1/2} (E t_n^2 + (E t_n^2 E r_n^2)^{1/2}) + (E q'^4_n E D'^4_{n1})^{1/4} (E t_n^2)^{1/2} (E r'''^2_n)^{1/2} \\
&\leq cmn^{-2}.
\end{aligned}$$

因此可以用

$$\begin{aligned}
\bar{f}_{n6}(t) &:= i t E \{ q'_n t_n w' (1 + t'_n + r''_n) (e^{i D'_{n1} w (1 + t'_n + r''_n) t} - 1) \exp \{ i t [q_n w (1 + t'_n) \\
&\quad + q'_n w (1 + t'_n + r''_n)] \} \}
\end{aligned}$$

代替 $\bar{f}_{n5}(t)$. 利用估计 f_{n4} 时的记号, 可写

$$\begin{aligned}
\bar{f}_{n6}(t) &= i t m E q'_n w' (1 + t'_n + r''_n) (e^{i D'_{n1} w (1 + t'_n + r''_n) t} - 1) \\
&\quad \cdot \exp \{ i t q'_n w (1 + t'_n + r''_n) \} \zeta^{m-1}(t) \kappa(t).
\end{aligned}$$

回顾对 $f_{n6}(t)$ 的估计并注意到 $E |D'_{n1}| \leq cn^{1/2}$, 我们可以移去 $\bar{f}_{n6}(t)$, 因此只需考察 $\bar{f}_{n3}(t)$ 就够了. 综合以下事实:

$$\begin{aligned}
(E D'^2_{n1})^{1/2} &\leq cm^{1/2} n^{-1}, \quad E |q'_n r'''_n| \leq cm^{\frac{1}{2}} n^{-\frac{1}{2} - \frac{7\delta}{12}}, \\
E |D'_{n1}(t_n + r_n + r'''_n)| &\leq c(m^{\frac{1}{2}} n^{-\frac{3}{2}} + mn^{-\frac{3}{2} - \frac{7\delta}{12}} + m^{\frac{1}{2}} n^{-1 - \frac{7\delta}{12}}),
\end{aligned}$$

可以用

$$\begin{aligned}
\bar{f}_{n7}(t) &:= E \{ (e^{i D'_{n1} w (1 + t'_n + r''_n) t} - 1) \exp \{ i t [q_n w (1 + t'_n) + q'_n w (1 + t'_n + r''_n)] \} \} \\
&= E \{ (e^{i D'_{n1} w (1 + t'_n + r''_n) t} - 1) \exp \{ i t q'_n w (1 + t'_n + r''_n) \} \zeta^m(t) \}
\end{aligned}$$

代替 $\bar{f}_{n3}(t)$ (参见 $f_{n5}(t)$). 回顾对 $\zeta^m(t)$ 的估计并注意到 $E |D'_{n1}| \leq cn^{-1/2}$, 得

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

将它与(27)式结合即得证定理.

参 考 文 献

- Arvesen J N. Jackknifing U-statistics. Ann Math Statist, 1969, 40(6): 2 076 ~ 2 100
- Callaert H, Veraverbeke N. The order of the normal approximation for a studentized U-statistics. Ann Statist, 1981, 9(1): 194 ~ 200
- 赵林城. 学生氏 U-统计量正态逼近的速度. 科学探索, 1983, 3(2): 45 ~ 52
- 陈希孺. 关于 U-统计量和 von-Mises 统计量的极限性质. 中国科学, 1980, (6): 522 ~ 532
- Serfling R J. Approximation Theorems of Mathematical Statistics. New York: Wiley, 1980
- Petov V V. Sums of Independent Random Variables. Berlin: Springer, 1975