第4卷 第1期 1984年1月
数学研究与评论
JOURNAL OF MATHEMATICAL RESEARCH AND EXPOSITION
Vol.4 No.1 Jan. 1984

随机足标U-统计量逼近正态的阶*

林正炎

（杭州大学）

随机变量随机和的收敛性问题无论在理论上还是实用上都是有重要意义的。关于随机和的中心极限定理已有相当一般的结果。近十年来又有一系列讨论收敛速度的文章(如Landers和Rogge[1],Sreehari[2]和Prakasa Rao[3])。关于U-统计量，它的随机中心极限定理已在Sproule[4]中给出。近年来对U-统计量的Berry-Esseen不等式也有相当深入的结果(如赵林城[5]，林正炎[6])。本文进一步讨论U-统计量的随机中心极限定理的收敛速度。

假设$\{X_n\}$是独立同分布随机变量序列。$\varphi(x_1,x_2)$是二元对称函数。不失一般性，设$E\varphi(X_1,X_2)=0$。定义U-统计量

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

记 $\psi(X_i)=E\{\varphi(X_i,X_j)\mid X_i\}$，$\hat{U}_n=\frac{2}{n}\sum_{i=1}^{n}\psi(X_i)$，$\sigma_n^2=EU_n^2$，$\hat{\sigma}_n^2=E\hat{U}_n^2$，$\sigma_\psi^2=E\psi^2(X_1)$。易知

$$\hat{\sigma}_n^2=4\sigma_\psi^2n^{-1},\quad \sigma_n^2=4\sigma_\psi^2n^{-1}+O(n^{-2}). \tag{1}$$

又记 $Y_{ij}=\varphi(X_i,X_j)-\psi(X_i)-\psi(X_j)$，$\Delta_n=(2\sigma_\psi)^{-1}n^{\frac{1}{2}}\binom{n}{2}^{-1}\sum_{1\le i<j\le n}Y_{ij}$。

另设ν_n是与$\{X_n\}$定义在同一概率空间上的取正整值的随机变量。首先我们考虑$\frac{\nu_n}{n}\xrightarrow{P}\tau$(正常数)的情形。

定理1 假设U-统计量$\{U_n\}$的核φ满足$E\varphi(X_1,X_2)=0$，$E|\varphi(X_1,X_2)|^{2+\delta}<\infty\ (0<\delta\le1)$，$\sigma_\psi^2>0$；又设正整值随机变量序列$\{\nu_n\}$满足

$$P\left(\left|\frac{\nu_n}{n}-\tau\right|>n^{-\alpha}\right)\le O(n^{-\frac{\delta}{2}}) \tag{2}$$

此处 $\alpha=\begin{cases}\delta, & 0<\delta\le\sqrt{3}-1,\\ \delta+\dfrac{\delta}{2+\delta}, & \sqrt{3}-1<\delta\le1.\end{cases}$

则有

$$\sup_x|P(\sigma_{\nu_n}^{-1}U_{\nu_n}<x)-\phi(x)|=O(n^{-\frac{\delta}{2}}),$$

其中$\phi(x)$是标准正态分布。

证 记$A_n=\left\{\left|\frac{\nu_n}{n}-\tau\right|\le n^{-\alpha}\right\}$，$n_0=[n\tau]$，$n_1=[n\tau-n^{1-\alpha}]$，$n_2=[n\tau+n^{1-\alpha}]$。

*1982年1月9日收到。

i) $0<\delta\leqslant\sqrt{3}-1$。由(1)，在 A_n 中

$$\left|\frac{\sigma_{\nu_n}}{2\sigma_\psi \nu_n^{-\frac{1}{2}}}-1\right|\leqslant\left|\frac{\sigma_{\nu_n}^2}{4\sigma_\psi^2\nu_n^{-1}}-1\right|=o(\nu_n^{-1})=o(n^{-1}). \tag{3}$$

故有

$$\sup_x|P(\sigma_{\nu_n}^{-1}U_{\nu_n}<x)-\phi(x)|\leqslant\sup_x|P(\sigma_{\nu_n}^{-1}U_{\nu_n}<x,A_n)-\phi(x)|+P(\bar{A}_n)$$
$$\leqslant\sup_x|P((2\sigma_\psi)^{-1}\nu_n^{\frac{1}{2}}U_{\nu_n}<x,\ A_n)-\phi(x)|+O(n^{-\frac{\delta}{2}})$$
$$\leqslant\sup_x|P((2\sigma_\psi)^{-1}\nu_n^{\frac{1}{2}}\hat{U}_{\nu_n}<x,A_n)-\phi(x)|+P(|\Delta_{\nu_n}|>n^{-\frac{\delta}{2}},\ A_n)+O(n^{-\frac{\delta}{2}}). \tag{4}$$

据[1](该文的结论虽是在二阶矩有限的条件下给出的，但不难写出 $2+\delta$ 阶矩有限时的完全类似的形式)，

$$\sup_x|P((2\sigma_\psi)^{-1}\nu_n^{\frac{1}{2}}\hat{U}_{\nu_n}<x,A_n)-\phi(x)|$$
$$\leqslant\sup_x|P((2\sigma_\psi)^{-1}\nu_n^{\frac{1}{2}}\hat{U}_{\nu_n}<x)-\phi(x)|+O(n^{-\frac{\delta}{2}})=O(n^{-\frac{\delta}{2}}). \tag{5}$$

对(4)式右端的第二项，

$$P(|\Delta_{\nu_n}|>n^{-\frac{\delta}{2}},A_n)\leqslant P(\max_{n_1\leqslant k\leqslant n_2}|\Delta_k|>n^{-\frac{\delta}{2}})$$
$$\leqslant P\left(\max_{n_1\leqslant k\leqslant n_2}\left|\sum_{j=2}^{k}\sum_{i=1}^{j-1}Y_{ij}\right|<cn^{\frac{3-\delta}{2}}\right) \tag{6}$$

(c 表常数，在不同的位置可取不同的值)。由于 $\zeta_j=\sum_{i=1}^{j-1}Y_{ij}(j=2,3,\cdots)$ 为一鞅差序列，故由关于半鞅的 Rényi-Hájek-Chow 不等式[7]，上式右端不超过

$$cE\left|\sum_{j=2}^{n_2}\sum_{i=1}^{j-1}Y_{ij}\right|^{2+\delta}n^{-\frac{1}{2}(3-\delta)(2+\delta)}.$$

又因对固定的 j，$W_k=\sum_{i=1}^{k}Y_{ij}(k=1,\cdots,j-1)$ 也有鞅性，所以由关于鞅的矩不等式

$$E\left|\sum_{j=2}^{n_2}\sum_{i=1}^{j-1}Y_{ij}\right|^{2+\delta}\leqslant cn_2^{2+\delta}\leqslant cn^{2+\delta}. \tag{7}$$

从而

$$P(|\Delta_{\nu_n}|>n^{-\frac{\delta}{2}},A_n)<cn^{-1+\frac{\delta}{2}+\frac{\delta^2}{2}}\leqslant cn^{-\frac{\delta}{2}} \tag{8}$$

最后的一个关系式是因为 $0<\delta\leqslant\sqrt{3}-1$。将(5)和(8)代入(4)即得待证之结论。

ii) $\sqrt{3}-1<\delta\leqslant1$。注意到(2)、(3)等式，我们有

$$\sup_x|P(\sigma_{\nu_n}^{-1}U_{\nu_n}<x)-\phi(x)|$$
$$\leqslant\sup_x|P((2\sigma_\psi)^{-1}(n\tau)^{\frac{1}{2}}U_{\nu_n}<x,A_n)-\phi(x)|+O(n^{-\frac{\delta}{2}})$$
$$\leqslant\sup_x|P((2\sigma_\psi)^{-1}(n\tau)^{\frac{1}{2}}U_{n_0}<x,A_n)-\phi(x)|$$
$$+P((2\sigma_\psi)^{-1}(n\tau)^{\frac{1}{2}}\max_{n_1\leqslant k\leqslant n_2}|U_k-U_{n_0}|>n^{-\frac{\delta}{2}})+O(n^{-\frac{\delta}{2}}). \tag{9}$$

由[5]并应用条件(2)，式右第一项为$O(n^{-\frac{\delta}{2}})$。对于第二项，写

$$U_k-U_{n_0}=\binom{k}{2}^{-1}\{\sum_{1\leq i<j\leq k}\varphi(X_i,X_j)-\sum_{1\leq i<j\leq n_0}\varphi(X_i,X_j)\}+\{\binom{k}{2}^{-1}-\binom{n_0}{2}^{-1}\}\sum_{1\leq i<j\leq n_0}\varphi(X_i,X_j),$$

并注意到

$$\max_{n_1\leq k\leq n_2}\left|\binom{k}{2}^{-1}-\binom{n_0}{2}^{-1}\right|<cn^{-2-\alpha},$$

我们有

$$P((2\sigma_\varphi)^{-1}(n\tau)^{\frac{1}{2}}\max_{n_1\leq k\leq n_2}|U_k-U_{n_0}|>n^{-\frac{\delta}{2}})$$

$$\leq P\left(\max_{n_1\leq k\leq n_0}\left|\sum_{j=k+1}^{n_0}\sum_{i=1}^{j-1}\varphi(X_i,X_j)\right|>cn^{\frac{3-\delta}{2}}\right)$$

$$+P\left(\max_{n_0<k\leq n_2}\left|\sum_{j=n_0+1}^{k}\sum_{i=1}^{j-1}\varphi(X_i,X_j)\right|>cn^{\frac{3-\delta}{2}}\right)+P\left(\left|\sum_{j=2}^{n_0}\sum_{i=1}^{j-1}\varphi(X_i,X_j)\right|>cn^{\frac{3-\delta}{2}+\alpha}\right).\quad(10)$$

我们来考虑上式右端的第二项。

$$P\left(\max_{n_0<k\leq n_2}\left|\sum_{j=n_0+1}^{k}\sum_{i=1}^{j-1}\varphi(X_i,X_j)\right|>cn^{\frac{3-\delta}{2}}\right)\leq P\left(\max_{n_0<k\leq n_2}\left|\sum_{j=n_0+1}^{k}\sum_{i=1}^{j-1}Y_{ij}\right|>cn^{\frac{3-\delta}{2}}\right)$$

$$+P\left(\max_{n_0<k\leq n_2}\left|(k-n_0)\sum_{i=1}^{n_0}\psi(X_i)+(k-1)\sum_{i=n_0+1}^{k}\psi(X_i)\right|>cn^{\frac{3-\delta}{2}}\right).\quad(11)$$

对于其后第一项，仿照 i)中(6)-(8)，并注意到此时$\alpha=\delta+\dfrac{\delta}{2+\delta}$。

$$P\left(\max_{n_0<k\leq n_2}\left|\sum_{j=n_0+1}^{k}\sum_{i=1}^{j-1}Y_{ij}\right|>cn^{\frac{3-\delta}{2}}\right)\leq cE\left|\sum_{j=n_0+1}^{n_2}\sum_{i=1}^{j-1}Y_{ij}\right|^{2+\delta}n^{-\frac{1}{2}(3-\delta)(2+\delta)}$$

$$\leq c(n_2-n_0)^{1+\frac{\delta}{2}}n_2^{1+\frac{\delta}{2}}n^{-\frac{1}{2}(3-\delta)(2+\delta)}\leq cn^{-1-\delta}\leq cn^{-\frac{\delta}{2}};\quad(12)$$

类似地对(11)右端的第二项，

$$P\left(\max_{n_0<k\leq n_2}\left|(k-n_0)\sum_{i=1}^{n_0}\psi(X_i)+(k-1)\sum_{i=n_0+1}^{k}\psi(X_i)\right|>cn^{\frac{3-\delta}{2}}\right)$$

$$\leq cE\left|(n_2-n_0)\sum_{i=1}^{n_0}\psi(X_i)+(n_2-1)\sum_{i=n_0+1}^{n_2}\psi(X_i)\right|^{2+\delta}n^{-\frac{1}{2}(3-\delta)(2+\delta)}$$

$$\leq c(n^{(1-\alpha)(2+\delta)}\cdot n^{1+\frac{\delta}{2}}+n^{2+\delta}\cdot n^{(1-\alpha)(1+\frac{\delta}{2})})\cdot n^{-\frac{1}{2}(3-\delta)(2+\delta)}=cn^{-\frac{\delta}{2}}.\quad(13)$$

将(12)、(13)代入(11)即得(10)式右端第二项的估计$O(n^{-\frac{\delta}{2}})$；对(10)式右端一、三两项，应用完全类似的方法也可得到同样的估计。定理证毕。

我们进一步讨论$\dfrac{\nu_n}{n}\xrightarrow{P}\lambda$(正值随机变量)的情形。

定理 2 U-统计量如定理1所设，但限制$0<\delta<\dfrac{1}{3}$。正整值随机变量序列$\{\nu_n\}$满足

$$P\left(\left|\frac{\nu_n}{n}-\lambda\right|>n^{-\alpha}\right)\leq O(n^{-\tau}),\quad(14)$$

其中$\dfrac{3}{2}\delta<\alpha\leq\dfrac{1}{2}$，$\tau=\delta(1-\alpha)(2-3\delta)^{-1}$，正值随机变量$\lambda$满足

$$P(\lambda\leqslant n^{\frac{3\delta-2\alpha}{2-3\delta}})+P(\lambda>\log n)=O(n^{-\tau}),\tag{15}$$

而且存在 $\beta>o$，使得

$$\sup_{A\in\mathscr{D}_n^\infty,B\in F}|P(A|B)-P(A)|\leqslant\theta(n)\downarrow 0,\tag{16}$$

式中 $\theta(n)=O(e^{-\beta n})$，$\mathscr{D}_n^\infty$ 和 F 分别表示由$\{X_m,m\geqslant n\}$和 λ 产生的 σ-域。则有

$$\sup_x|P(\sigma_{\nu_n}^{-1}U_{\nu_n}<x)-\phi(x)|=O(n^{-\min(\frac{2\beta\tau}{\delta},\tau)}).$$

证 参照[2]、[3]，我们有

$$\sup_x|P(\sigma_{\nu_n}^{-1}\hat{U}_{\nu_n}<x)-\phi(x)|=O(n^{-\min(\frac{2\beta\tau}{\delta},\tau)}).$$

因此类似于定理1证明的第一部分中指出的，下列估计对于我们的结论是足够的：

$$P(|\Delta_{\nu_n}|>n^{-\tau})=O(n^{-\tau}).\tag{17}$$

为证 (17)，记 $m_n=n^{3\tau}$，$M_n=n^a\log n$，$A_n=\left\{\left|\frac{\nu_n}{n}-\lambda\right|\leqslant n^{-a}\right\}$，

$B_{ni}=\{in^{-a}<\lambda<(i+1)n^{-a}\}$，$C_n=\{\lambda\leqslant n^{3\tau-a}\}\cup\{\lambda>\log n\}$，$a_{ni}=[(i-1)n^{1-a}]$
$b_{ni}=[(i+2)n^{1-a}]$。于是可写

$$\begin{aligned}P(|\Delta_{\nu_n}|>n^{-\tau})&\leqslant P(|\Delta_{\nu_n}|>n^{-\tau},A_n)+P(\bar{A}_n)\\&\leqslant\sum_{i=m_n}^{M_n}P(|\Delta_{\nu_n}|>n^{-\tau},A_n,B_{ni})+P(C_n)+P(\bar{A}_n).\end{aligned}\tag{18}$$

由条件(14)、(15)，

$$P(C_n)+P(\bar{A}_n)=O(n^{-\tau}).\tag{19}$$

而据(16)，又有

$$\begin{aligned}P(|\Delta_{\nu_n}|>n^{-\tau},A_n,B_{ni})&\leqslant P(\max_{a_{ni}\leqslant k\leqslant b_{ni}}|\Delta_k|>n^{-\tau},B_{ni})\\&\leqslant P(B_{ni})\{\theta(a_{ni})+P(\max_{a_{ni}\leqslant k\leqslant b_{ni}}|\Delta_k|>n^{-\tau})\}.\end{aligned}\tag{20}$$

记 $t(n)=(2\sigma_\psi)^{-1}n^{\frac{1}{2}}\binom{n}{2}^{-1}$，$T_n=\sum_{1\leqslant i<j\leqslant n}Y_{ij}$，因此 $\Delta_n=t(n)T_n$。也利用鞅差性，仿照定理1中的论证

$$\begin{aligned}P(\max_{a_{ni}\leqslant k\leqslant b_{ni}}|\Delta_k|>n^{-\tau})&\leqslant P(\max_{a_{ni}\leqslant k\leqslant b_{ni}}|T_k|>n^{-\tau}t^{-1}(a_{ni}))\\&\leqslant E|T_{b_{ni}}|^{2+\delta}(n^{-\tau}t^{-1}(a_{ni}))^{-(2+\delta)}\leqslant cb_{ni}^{2+\delta}(n^{-\tau}a_{ni}^{\frac{3}{2}})^{-(2+\delta)}\\&\leqslant ci^{(2+\delta)-\frac{3}{2}(2+\delta)}n^{(1-a)(2+\delta)-(\frac{3}{2}(1-a)-\tau)(2+\delta)}\leqslant cn^{-\frac{1}{2}(1-a+\tau)(2+\delta)}<cn^{-\tau},\end{aligned}\tag{21}$$

最后第二个不等号在于 $i\geqslant m_n$。此外由(16)，显然有

$$\theta(a_{ni})\leqslant\theta(a_{nm_n})\leqslant cn^{-\tau}.\tag{22}$$

将(21)、(22)代入(20)并结合(18)、(19)即得(17)式，定理证毕。

参考文献

[1] Landers, D. and Rogge, L., The exact approximation order in the central-limit-theorem for random summation, Z. *Wahrscheinlichkeitstheorie verw. Geb.*, 36(1976), 269-283.

[2] Sreehari, M., An estimate of the residual in the central limit theorem for sums of random numbers of random variables. University of Baroda, Unpublished manuscript.

[3] Prakasa Rao, B. L. S., On the rate of convergence in the random central limit theorem for martingale, *Bull. Acad. Polon. Sci. Sêr. Sci. Math. Astronom. Phys.*, 22(1974), 1255-1260.

[4] Sproule, R. N., Asymptotic properties of U-statistics, *Trans. Amer. Math. Soc.* 199(1974), 55-64.

[5] 赵林城，U-统计量的收敛速度．科学探索，1(1981)，89-94．

[6] 林正炎．关于U-统计量的 Berry-Esseen 不等式的推广，应用数学学报，6(1983)，468—475．

[7] Chow, Y. S., A martingale inequality and the law of large numbers. *Proc. Amer. Math. Soc.* 11 (1960), 107-111.

The Order Approximating to Normal Distribution for U-Statistics with Random Indexes

Lin Zhengyan

Abstract

In this paper, we consider the rate of convergence in the random central limit theorem for U-statistics, and get two results:

Theorem 1 Suppose that the kernel φ of U-statistics $\{U_n\}$ satisfies $E\,\varphi(X_1, X_2)=0$ $E\,|\varphi(X_1, X_2)|^{2+\delta}<\infty\ (0<\delta\leqslant 1), E^2\{E(\varphi(X_1, X_2)\,|\,X_1)\}>0$. Suppose also that $\{\nu_n\}$ is a sequence of positive integer-valued r. v. satisfing

$$P\left(\left|\frac{\nu_n}{n}-\tau\right|>n^{-a}\right)\leqslant O(n^{-\frac{\delta}{2}})$$

here τ is a positive constant, $a=\begin{cases}\delta, & 0<\delta\leqslant\sqrt{3}-1,\\ \delta+\delta(2+\delta)^{-1}, & \sqrt{3}-1<\delta\leqslant 1.\end{cases}$ Then

$$\sup_x |P(\sigma_{\nu_n}^{-1}U_{\nu_n}<x)-\phi(x)|=O(n^{-\frac{\delta}{2}}),$$

where $\sigma_n^2=EU_n^2$.

Theorem 2 Define U-statisrics as theorem 1, with $\delta\in(0, 1/3)$, suppose that a sequence $\{\nu_n\}$ of positive integer-valued r. v. a satisfies

$$P\left(\left|\frac{\nu_n}{n}-\lambda\right|>n^{-a}\right)\leqslant O(n^{-\tau})$$

here $\frac{3}{2}\delta<a\leqslant\frac{1}{2}, \tau=\delta(1-a)(2-3\delta)^{-1}$, suppose also that λ is a positive valued r. v. a satisfying

$$P(\lambda\leqslant n^{\frac{3\delta-2a}{2-3\delta}})+P(\lambda>\log\ n)=O(n^{-\tau}),$$

and
$$\sup_{A\in\mathscr{B}_n^\infty, B\in F}|P(A|B)-P(A)|\leqslant\theta(n)\downarrow 0,$$

where $\theta(n)=O(e^{-\beta n})$ for some $\beta>0$, F and $\mathscr{B}_n^\infty$ are the σ- fields generated by λ and $\{X_m, m\geqslant n\}$ respectively. Then

$$\sup_x |P(\sigma_{\nu_n}^{-1}U_{\nu_n}<x)-\phi(x)|=O(n^{-\min\left(\frac{2\beta\tau}{\delta}, \tau\right)}).$$