1982年 第2卷 第4期
Vol.2, No.4, 1982

数学研究与评论
JOURNAL OF MATHEMATICAL RESEARCH & EXPOSITION

U-统计量的分布的非一致性收敛速度*

赵林城 陈希孺

(中国科技大学数学系)

设$\{X_n\}$为一串iid.随机变量，$h(x,y)$为两个变元x,y的对称函数，则称

$$U_n=\binom{n}{2}^{-1}\sum_{1\le j<k\le n}h(X_j,X_k) \tag{1}$$

为以$h(x,y)$为核的U-统计量．不失一般性，以下设$Eh(X_1,X_2)=0$，并记

$$\sigma_n^2=Var(U_n),\quad \sigma_g^2=Eg^2(X_1),\quad \Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{u^2}{2}}du,$$

其中，$g(x)=Eh(x,X_2)$．

U-统计量是一个很重要的统计量．探讨它的大样本性质，一直是国内外数理统计学者兴趣之所在．近几年来，它的大样本理论的研究取得了很大的进展．在依分布收敛于正态的速度方面，1978年，H. Callaert和P. Janssen集一系列研究之大成，在[1]中证得：若$E|h(X_1,X_2)|^3<\infty$，且$\sigma_g^2>0$，则

$$\sup_x\left|P\left(\frac{U_n}{\sigma_n}\leqslant x\right)-\Phi(x)\right|=O(n^{-\frac{1}{2}}). \tag{2}$$

这一结果后来为赵林城[2]所改进．赵在$E|h(X_1,X_2)|^{2+\delta}<\infty\ (0<\delta\leqslant1)$，且$\sigma_g^2>0$的条件下，得到了理想的收敛速度$O(n^{-\frac{\delta}{2}})$．

这些结果平行于独立随机变量和的经典理论．近年来，在独立和的情形发展了非一致性收敛速度的深刻理论（参看[3]，第五、六章），其难度远远超出了“一致性”的理论．这个理论断言，若$X_1,X_2,\cdots$ iid.，且$EX_1=0$，$E|X_1|^3<\infty$，$\sigma^2=Var(X_1)>0$，则对$\overline{X}_n=$

1981年9月7日收到．

$\frac{1}{n}\sum_{j=1}^{n}X_j$，有

$$\left|P\left(\frac{\overline{X}_n}{\sqrt{Var(\overline{X}_n)}}\leqslant x\right)-\Phi(x)\right|\leqslant C n^{-\frac{1}{2}}(1+|x|)^{-3}，对所有x\in(-\infty,\infty)， \tag{3}$$

其中C是一个与n及x都无关的常数，仅与X_1的分布(实际上仅与$\rho=E|X_1|^3/\sigma^3$)有关．如所周知，独立和情形具有的一些大样本性质，有不少已对U_-统计量证实了．截至目前为止，还没有见到一个反例，说明独立和具有的某种大样本性质，对U_-统计量不成立，只是有一些性质，由于U_-统计量本身的难度而未被证实罢了．将U_-统计量与独立和类比，自然会提出这样的问题：对独立和建立的非一致性收敛速度，是否可以推广到U_-统计量？这当然是人们关心的一个问题．

作为这个方向上的一个尝试，赵林城在[4]中证明了一个初步结果，即若$E|h(X_1,X_2)|^3<\infty$，且$\sigma_g^2>0$，则存在常数C，使对所有x及充分大的n，有

$$\left|P\left(\frac{U_n}{\sigma_n}\leqslant x\right)-\Phi(x)\right|\leqslant C n^{-\frac{1}{2}}(1+|x|)^{-2}. \tag{4}$$

这个结果与独立和情形的相应结果仍有一定的差距．最近，我们在[5]中将(4)式右方改进为$Cn^{-\frac{1}{2}}(1+|x|)^{-3}$，从而得到了最终结果：

定理1 设$E|h(X_1,X_2)|^3<\infty$，且$\sigma_g^2=Eg^2(X_1)>0$，则存在常数C，使对所有x及充分大的n，有

$$\left|P\left(\frac{U_n}{\sigma_n}\leqslant x\right)-\Phi(x)\right|\leqslant Cn^{-\frac{1}{2}}(1+|x|)^{-3}. \tag{5}$$

注意，此处的C不依赖于n和x，它只依赖于核函数h，以及X_1的分布．

易见，这里给出的结果大大改进了(2)式给出的结果．

不言而喻，这个结果不难推广到核函数h为m个变元$x_1,x_2,\cdots,x_m$的对称函数的情形．而且，这里提供的结果和方法，可用来得出von-Mises统计量的非一致性收敛速度．如果有兴趣的话，这个结果还可作两方面的推广．一是可以推广到$E|h(X_1,X_2)|^{2+\delta}<\infty$ $(0<\delta<1)$的情形．一是可以取消$X_1,X_2,\cdots$同分布的假定，仅假定$X_1,X_2,\cdots$相互独立，而加上一些相应的条件．作者们期望，这一结果对得出刀切von-Mises统计量的收敛速度也许会有所裨益．

在这个简报里，我们简略地叙述一下[5]中证明的思路和方法．为此，取$\lambda_n=n-[n^{\frac{2}{3}}]$，简记$h_{jk}=h(X_j,X_k)$，$g_j=g(X_j)$，并令

$$S_n=\frac{1}{\sqrt{n}\,\sigma_g}\sum_{j=1}^{n}g_j,\quad S_n'=\frac{1}{\sqrt{n}\,\sigma_g}\sum_{j=1}^{\lambda n}g_j,\quad S_n''=S_n-S_n',$$

$$Y_{jk}=h_{jk}-g_j-g_k,\quad g_j=E(h_{jk}|X_j),$$
$$\hat{h}_{jk}=h_{jk}I_{(|h_{jk}|\leq\sqrt{n})},\quad h_{jk}^*=\hat{h}_{jk}-E\hat{h}_{jk},$$

$$y^*_{jk}=h^*_{jk}-g^*_j-g^*_k,\quad g^*_j=E(h^*_{jk}|x_j),$$

$$\Delta_n=\frac{\sqrt{n}}{2\sigma_g}\binom{n}{2}^{-1}\sum_{1\le j<k\le n}y_{jk},\qquad \Delta^*_n=\frac{\sqrt{n}}{2\sigma_g}\binom{n}{2}^{-1}\sum_{1\le j\wedge k\le n}y^*_{jk},$$

$$\Delta^*_{n1}=\frac{\sqrt{n}}{2\sigma_g}\binom{n}{2}^{-1}\sum_{1\le j<k\le n}y^*_{jk},\qquad \Delta^*_{n2}=\Delta^*_n-\Delta^*_{n1}$$

[5]中证明了，对$|x|\geqslant 1$,

$$P\left(\left|\frac{\sqrt{n}}{2\sigma_g}U_n-(S_n+\Delta^*_n)\right|\geqq|x|/\sqrt{n}\right)\leqq cn^{-1/2}(1+|x|)^{-3},$$

$$P\left(|(S_n+\Delta^*_n)-(S_n+\Delta^*_{n1})|\geqq|x|/\sqrt{n}\right)\leqq cn^{-1/2}(1+|x|)^{-3},$$

因而把定理1的证明归结为证明

$$|P(S_n+\Delta^*_{n1}\leqq x)-\Phi(x)|\leqq cn^{-1/2}(1+|x|)^{-3},\tag{6}$$

为此，我们证明了，存在与n无关的常数$\eta>0$，$\mu>0$，当$|t|\leqq\sqrt{n}\eta$时，有

1°　$$|E(S''_n)^je^{itS''_n}|\leqq cn^{-\delta(j)/3}(\varepsilon_{1j}+|t|^j)\exp(-\mu n^{-1/3}t^2),\quad j=0,1,2,3.$$

其中，
$$\delta(j)=\begin{cases}0, & 若\ j=0,\\ 1, & 若\ j=1\ 或\ 2,\\ 2, & 若\ j=3.\end{cases}\qquad \varepsilon_{1j}=\begin{cases}0, & j=1,\\ 1, & j\neq 1.\end{cases}$$

2°　$$|E(\Delta^{*j}_{n1}e^{itS'_n})|\leqq Cn^{-1/2}(\varepsilon_{1j}+|t|^{j+1})e^{-\mu t^2},\quad j=1,2,3.$$

$$|Ee^{itS'_n}(e^{it\Delta^*_{n1}}-1)|\leqslant Cn^{-1/2}(t^2+|t|^7)e^{-\mu t^2}+Cn^{-2}t^4.$$

3°　$$|E(S'_n)e^{jtS'_n}(e^{it\Delta^*_{n1}}-1)|\leqslant Cn^{-1/2}(|t|+|t|^{j+3}e^{-\mu t^2}+Cn^{-1}t^2,\quad j=0,1,2$$

$$|E(S'_n)^3e^{itS'_n}(e^{it\Delta^*_{n1}}-1)|\leqq C/\sqrt{n}(|t|+|t|^{11})e^{-\mu t^2}+Cn^{-3/2}$$
$$\cdot(t^2+t^4)+Cn^{-2}(|t|^5+|t|^7).$$

4°　$$\int_{|t|\leqq\sqrt{n}\eta}|t|^{-1}\cdot|E(S_n+\Delta^*_{n1})^3e^{its_n}(e^{it\Delta^*_{n1}}-1)|dt\leqq C/\sqrt{n}.$$

5°　$$\int_{|t|\leqq\sqrt{n}\eta}|t|^{-1}\cdot|i^3E(S_n+\Delta^*_{n1})^3e^{its_n}-\psi_n(t)|dt\leqq C/\sqrt{n},$$

其中，　$$\psi_n(t)=i^3Es_n^3e^{its_n}+i^3[3Es_n^2\Delta^*_{n1}+3Es_n\Delta^{*2}_{n1}+E\Delta^{*3}_{n1}]e^{-\frac{t^2}{2}}$$

6°　设$G_n(x)$为非降函数，$H_n(x)$在$(-\infty,\infty)$上有有界变差，其F.S.变换分别记为g_n和h_n。假定$G_n(\pm\infty)=H_n(\pm\infty)$，$\int_{-\infty}^{\infty}|x|^3|d\{G_n(x)-H_n(x)\}|<\infty$，且

$$|H_n(x)-\Phi(x)|\leqslant Cn^{-\frac{1}{2}}(1+|x|)^{-3},$$

则对 $T_n=\sqrt{n}\eta$（$\eta>0$ 与 n 及 x 无关），有

$$|G_n(x)-H_n(x)|\leqslant C(1+|x|)^{-3}\Big\{\int_{|t|\leq T_n}|t|^{-1}|g_n(t)-h_n(t)|dt$$

$$+\int_{|t|\leq T_n}|t|^{-1}|\delta_3^{(n)}(t)|dt+C/\sqrt{n}\Big\},$$ 对充分大的 n 和所有 x，

此处，$\delta_3^{(n)}(t)=\int_{-\infty}^{\infty}e^{itx}d\{x^3(G_n(x)-H_n(x))\}$.

利用上述引理及一些已知结果，即可证明（6）式．详细证明可参看[5]．

参考文献

[1] Callaert, H. and Janssen, P., Ann, *Statist*., 6(1978), 417—421.

[2] 赵林城，科学探索，1(1981)，89—94.

[3] Petrov, V.V., Sums of Independent Random Variables, Springer-Verlag, 1975.

[4] 赵林城，U-统计量的非一致性界限，已投《数学年刊》.

[5] 赵林城，陈希孺，U-统计量的分布的非一致性收敛速度，已投《中国科学》.