应用概率统计　第二十卷
第二期　2004 年 5 月

Chinese Journal of Applied Probability
and Statistics Vol. 20 No. 2 May 2004

k-U 统计量的渐进正态性

张　日　权
(中科院数学与系统科学研究院，北京， 100080；　雁北师范学院数学系，大同， 037000)

摘　要

称 $U_n^{m,k}=\sum\limits_{1\le i_1<\cdots<i_m\le k}g(X_{i_1},\cdots,X_{i_m})+\sum\limits_{i_m-k+1\le i_1<\cdots<i_{m-1}<i_m}\sum\limits_{k+1\le i_m\le n}g(X_{i_1},\cdots,X_{i_m})$ 为 k-U 统计量，其中 $g(x_1,\cdots,x_m)$ 是一对称函数， k 为小于等于 n 的自然数， k 可能依赖于 n. 这一表达形式是一类统计量，在 $k=n$ 时， $\left(\begin{array}{c}n\\m\end{array}\right)^{-1}U_n^{m,n}$ 就是 U- 统计量．本文证明了 $U_n^{m,k}$ 的渐进正态性．

关 键 词: k-U 统计量， U- 统计量，渐进正态性，正态分布.
学 科 分 类 号: O212.1.

§1.　引　　言

众所周知， $U_n=\left(\begin{array}{c}n\\m\end{array}\right)^{-1}\sum\limits_{1\le i_1<\cdots<i_m\le n}g(X_{i_1},\cdots,X_{i_m})$ 称 U- 统计量，其中 $X_1,\cdots,X_n$ 是从分布为 F 的母体中抽取的随机样本， $g(x_1,\cdots,x_m)$ 是对称函数．1948 年 Hoeffding 证明了 U- 统计量的渐近正态性．本文引入如下比 U- 统计量更广泛的一类统计量．我们称

$$U_n^{m,k}=\sum_{1\le i_1<\cdots<i_{m-1}<i_m\le k}g(X_{i_1},\cdots,X_{i_m})+\sum_{i_m-k+1\le i_1<\cdots<i_{m-1}<i_m}\sum_{k+1\le i_m\le n}g(X_{i_1},\cdots,X_{i_m}) \tag{1.1}$$

为 k-U 统计量．k 为小于等于 n 的自然数，k 可能依赖于 n. U- 统计量是它的特殊情况 $\left(\left(\begin{array}{c}n\\m\end{array}\right)^{-1}U_n^{m,n}=U_n\right)$. 本文证明了这类统计量的渐近正态性．

§2.　k-U 统计量的数字特征

设 k-U 统计量 $U_n^{m,k}$ 共有 d 项．显然

$$d=\left(\begin{array}{c}k\\m\end{array}\right)+(n-k)\left(\begin{array}{c}k-1\\m-1\end{array}\right). \tag{2.1}$$

假设 $\mathsf{E}_F[g(X_1,\cdots,X_m)]=\theta(F)$, 可知 $\mathsf{E}_F[U_n^{m,k}]=d\theta(F)$. 令

$$g_c(x_1,\cdots,x_c)=\mathsf{E}_F[g(x_1,\cdots,x_c,X_{c+1},\cdots,X_m)],\qquad c=0,1,\cdots,m,$$

其中 g_0 理解为 $\mathsf{E}_F[g(X_1,\cdots,X_m)]$, $g_m=g(x_1,\cdots,x_m)$. 显然， $\mathsf{E}_F[g_c(X_1,\cdots,X_c)]=\theta(F)$.

假设 $\mathsf{E}_F[g^2(X_1,\cdots,X_m)]\le M<\infty$. 令 $Y=g(X_1,\cdots,X_m)$, 则

$$g_c^2(X_1,\cdots,X_c)=[\mathsf{E}_F(Y|X_1,\cdots,X_c)]^2\le \mathsf{E}_F(Y^2|X_1,\cdots,X_c).$$

该式两边取期望，得 $\mathsf{E}_F[g^2(X_1,\cdots,X_c)]\le M$. 这样一来可以定义 $\sigma_c^2=\mathsf{Var}_F[g_c(X_1,\cdots,X_c)]$, $c=0,1,\cdots,m$.

本文 2002 年 10 月 29 日收到， 2003 年 5 月 4 日收到修改稿.

下面讨论 $\mathrm{Var}_F(U_n^{m,k})$. 不失一般性假定 $\theta(F)=0$, 从而 $\mathrm{Var}_F(U_n^{m,k})=\mathrm{E}_F(U_n^{m,k})^2$. 令

$$T_1=\sum_{1\le i_1<\cdots<i_m\le k} g(X_{i_1},\cdots,X_{i_m}),$$
$$T_i=\sum_{i\le i_1<\cdots<i_{m-1}\le k+i-2} g(X_{i_1},\cdots,X_{i_{m-1}},X_{k+i-1}),\qquad i=2,\cdots,n-k+1.$$

则 $U_n^{m,k}=\sum\limits_{i=1}^{n-k+1}T_i$. 由于 $T_2,\cdots,T_{n-k+1}$ 是平稳序列，所以,

$$\begin{aligned}&\mathrm{Var}_F(U_n^{k,m})\\ =\ &\sum_{i=1}^{n-k+1}\mathrm{Var}_F(T_i)+2\sum_{i=2}^{\min(k,n-k+1)}\mathrm{Cov}_F(T_1,T_i)+2\sum_{i=3}^{\min(k+1,n-k+1)}(n-k-i+2)\mathrm{Cov}_F(T_2,T_i).\end{aligned}\tag{2.2}$$

下面分别讨论 (2.2) 式各和号内的部分.

(i) 计算 $\mathrm{Var}_F(T_1)$. 由 U- 统计量的方差公式可知

$$\mathrm{Var}_F(T_1)=\sum_{c=1}^{m}\binom{k}{m}\binom{m}{c}\binom{k-m}{m-c}\sigma_c^2;\tag{2.3}$$

(ii) 计算 $\mathrm{Var}_F(T_i)$, $(i=2,\cdots,n-k+1)$. 用类似计算 $\mathrm{Var}_F(T_1)$ 的方法，可知

$$\mathrm{Var}_F(T_i)=\sum_{c=1}^{m}\binom{k-1}{m-1}\binom{m-1}{c-1}\binom{k-m}{m-c}\sigma_c^2,\qquad (i=2,\cdots,n-k+1);\tag{2.4}$$

(iii) 计算 $\mathrm{Cov}_F(T_1,T_i)$, $(i=2,\cdots,\min(k,n-k+1))$. 由于 $\mathrm{Cov}_F(T_1,T_i)=\sum_2\mathrm{E}_F[g(X_{i_1},\cdots,X_{i_m})\cdot g(X_{j_1},\cdots,X_{j_{m-1}},X_{k+i-1})]$, 其中 $\sum_2$ 求和范围是 $1\le i_1<\cdots<i_m\le k$, $i\le j_1<\cdots<j_{m-1}\le k+i-2$. 设足标 $(i_1,\cdots,i_m)$ 和 $(j_1,\cdots,j_{m-1},k+i-1)$ 中有 c 个公共足标，则

$$\mathrm{E}_F[g(X_{i_1},\cdots,X_{i_m})g(X_{j_1},\cdots,X_{j_{m-1}},X_{k+i-1})]=\begin{cases}0 & c=0;\\ \sigma_c^2 & c=1,\cdots,\min(m-1,k-i+1).\end{cases}$$

在和 $\sum_2$ 中，足标恰有 c 个公共的那种项，共有

$$\sum_{l=\max(0,m-c-i+1)}^{\min(k-c-i+1,m-c-1)}\binom{k-i+1}{c}\binom{k-i+1-c}{l}\binom{i-2}{m-1-c-l}\binom{k-c-l}{m-c}\text{ 个}.$$

其计算方法如下. 在数组 $(i,\cdots,k)$ 中取 c 个数，有 $\binom{k-i+1}{c}$ 种取法. 接着在其剩下的 $k-i+1-c$ 个数中取 l 个数，有 $\binom{k-i+1-c}{l}$ 种取法，再在数组 $(k+1,\cdots,k+i-2)$ 中取 $m-1-c-l$ 个数，有 $\binom{i-2}{m-1-c-l}$ 种取法. 接下来在数组 $(1,\cdots,k)$ 去掉前面已取走的 $c+l$ 个数后剩下的 $k-(c+l)$ 个数中取 $m-c$ 个数, 有 $\binom{k-c-l}{m-c}$ 种取法. 这样就完成了在数组 $(1,\cdots,k)$ 中取 m 个数, 在数组 $(i,\cdots,k+i-2)$ 中取 $m-1$ 个数, 取出的两组数中恰有 c 个相同的数. 其中 l 取值在 $\max(0,m-c-i+1)$ 和 $\min(k-c-i+1,m-c-1)$ 之间. 所以

$$\begin{aligned}&\mathrm{Cov}_F(T_1,T_i)\\ =\ &\sum_{c=1}^{\min(m-1,k-i+1)}\sum_{l=m-c-i+1}^{m-c-1}\binom{k-i+1}{c}\binom{k-i+1-c}{l}\binom{i-2}{m-1-c-l}\binom{k-c-l}{m-c}\sigma_c^2,\\ &(i=2,\cdots,\min(k,n-k+1));\end{aligned}\tag{2.5}$$

(iiii) 计算 $\mathrm{Cov}_F(T_2,T_i)$, $(i=3,\cdots,\min(k+1,n-k+1))$. 由于

$$\mathrm{Cov}_F(T_2,T_i)=\sum{}_3 \mathrm{E}_F[g(X_{i_1},\cdots,X_{i_{m-1}},X_{k+1})\cdot g(X_{j_1},\cdots,X_{j_{m-1}},X_{k+i-1})],$$

其中 $\sum_3$ 求和范围是 $2\le i_1<\cdots<i_{m-1}\le k$, $i\le j_1<\cdots<j_{m-1}\le k+i-2$. 用类似于计算 $\mathrm{Cov}_F(T_1,T_i)$ 的方法，可得：

$$\begin{aligned}
&\mathrm{Cov}_F(T_2,T_i)\\
=&\sum_{c=1}^{\min(m-1,k-i+2)}\sum_{l=\max(0,m-c-i+2)}^{\min(k-i-c+1,m-c-1)}\binom{k-i+1}{c}\binom{k-i+1-c}{l}\binom{i-2}{m-1-c-l}\binom{k-1-c-l}{m-1-c}\sigma_c^2\\
&+\sum_{c=1}^{\min(m-1,k-i+2)}\sum_{l=\max(0,m-c-i+2)}^{\min(k-i-c+2,m-c-1)}\binom{k-i+1}{c-1}\binom{k-i+2-c}{l}\binom{i-2}{m-1-c-l}\binom{k-c-l}{m-c}\sigma_c^2\\
&\qquad(i=3,\cdots,\min(k+1,n-k+1)).
\end{aligned}\tag{2.6}$$

令

$$\Delta_1^c=\binom{k}{m}\binom{m}{c}\binom{k-m}{m-c},\qquad \Delta_2^c=\binom{k-1}{m-1}\binom{m-1}{c-1}\binom{k-m}{m-c},$$

$$\Delta_3^{c,l,i}=\binom{k-i+1}{c}\binom{k-i+1-c}{l}\binom{i-2}{m-1-c-l}\binom{k-c-l}{m-c},$$

$$\Delta_4^{c,l,i}=\binom{k-i+1}{c}\binom{k-i+1-c}{l}\binom{i-3}{m-1-c-l}\binom{k-1-c-l}{m-1-c},$$

$$\Delta_5^{c,l,i}=\binom{k-i+1}{c-1}\binom{k-i+2-c}{l}\binom{i-3}{m-1-c-l}\binom{k-c-l}{m-c}.$$

将 (2.2) 、 (2.3) 、 (2.4) 、 (2.5) 和 (2.6) 结合起来得

$$\begin{aligned}
\mathrm{Var}_F(U_n^{k,m}) =& \sum_{c=1}^{m}\Delta_1^c\sigma_c^2+(n-k)\sum_{c=1}^{m}\Delta_2^c\sigma_c^2+2\sum_{i=2}^{\min(k,n-k+1)}\sum_{c=1}^{\min(m-1,k-i+1)}\sum_{l=\max(0,m-c-i+1)}^{\min(k-c-i+1,m-c-1)}\Delta_3^{c,l,i}\sigma_c^2\\
&+2\sum_{i=3}^{\min(k+1,n-k+1)}\sum_{c=1}^{\min(m-1,k-i+2)}(n-k-i+2)\\
&\cdot\Big(\sum_{l=\max(0,m-c-i+2)}^{\min(k-c-i+1,m-c-1)}\Delta_4^{c,l,i}+\sum_{l=\max(0,m-c-i+2)}^{\min(k-c-i+2,m-c-1)}\Delta_5^{c,l,i}\Big)\sigma_c^2.
\end{aligned}\tag{2.7}$$

故 $k\le n/2$ 时,

$$\begin{aligned}
\mathrm{Var}_F(U_n^{k,m}) =& \sum_{c=1}^{m}\Delta_1^c\sigma_c^2+(n-k)\sum_{c=1}^{m}\Delta_2^c\sigma_c^2+2\sum_{i=2}^{k}\sum_{c=1}^{\min(m-1,k-i+1)}\sum_{l=\max(0,m-c-i+1)}^{\min(k-c-i+1,m-c-1)}\Delta_3^{c,l,i}\sigma_c^2\\
&+2\sum_{i=3}^{k+1}\sum_{c=1}^{\min(m-1,k-i+2)}(n-k-i+2)\\
&\cdot\Big(\sum_{l=\max(0,m-c-i+2)}^{\min(k-c-i+1,m-c-1)}\Delta_4^{c,l,i}+\sum_{l=\max(0,m-c-i+2)}^{\min(k-c-i+2,m-c-1)}\Delta_5^{c,l,i}\Big)\sigma_c^2;
\end{aligned}\tag{2.8}$$

$k>n/2$ 时,

$$\begin{aligned}
\mathrm{Var}_F(U_n^{k,m}) =& \sum_{c=1}^{m}\Delta_1^c\sigma_c^2+(n-k)\sum_{c=1}^{m}\Delta_2^c\sigma_c^2+2\sum_{i=2}^{n-k+1}\sum_{c=1}^{\min(m-1,k-i+1)}\sum_{l=\max(0,m-c-i+1)}^{\min(k-c-i+1,m-c-1)}\Delta_3^{c,l,i}\sigma_c^2\\
&+2\sum_{i=3}^{n-k+1}\sum_{c=1}^{\min(m-1,k-i+2)}(n-k-i+2)\\
&\cdot\Big(\sum_{l=\max(0,m-c-i+2)}^{\min(k-c-i+1,m-c-1)}\Delta_4^{c,l,i}+\sum_{l=\max(0,m-c-i+2)}^{\min(k-c-i+2,m-c-1)}\Delta_5^{c,l,i}\Big)\sigma_c^2.
\end{aligned}\tag{2.9}$$

§3. 极限分布

3.1 在 k 为某个固定的常数时 $U_n^{k,m}$ 的极限分布

由于 k 为固定的常数，当 n 充分大时，$k\le n/2$ 成立．故 $U_n^{k,m}$ 的方差为 (2.8) 式．令

$$\delta_1^2=\sum_{c=1}^{m}\Delta_2^c\sigma_c^2+2\sum_{i=3}^{k+1}\sum_{c=1}^{\min(m-1,k-i+2)}\Big(\sum_{l=\max(0,m-c-i+2)}^{\min(k-c-i+1,m-c-1)}\Delta_4^{c,l,i}+\sum_{l=\max(0,m-c-i+2)}^{\min(k-c-i+2,m-c-1)}\Delta_5^{c,l,i}\Big)\sigma_c^2. \tag{3.1}$$

通过计算可得：$\mathrm{Var}_F(U_n^{k,m})=n\delta_1^2+o(n)$.

下面讨论 $n^{-1/2}U_n^{k,m}$ 的分布．由于 $n^{-1/2}U_n^{k,m}=n^{-1/2}T_1+n^{-1/2}\sum\limits_{i=2}^{n-k+1}T_i$, 而

$$\mathrm{Var}\,(n^{-1/2}T_1)=n^{-1}\sum_{c=1}^{m}\Delta_1^c\sigma_c^2\to 0,$$

所以，$n^{-1/2}T_1\xrightarrow{P}0$; 又 $T_2,\cdots,T_{n-k+1}$ 是 $k-1$- 相依平稳序列，且当 $n\to\infty$ 时，

$$\begin{aligned}\mathrm{Var}\Big(n^{-1/2}\sum_{i=2}^{n-k+1}T_i\Big)&=n^{-1}(n-k)\sum_{c=1}^{m}\Delta_2^c\sigma_c^2+2n^{-1}\sum_{i=3}^{k+1}\sum_{c=1}^{\min(m-1,k-i+2)}(n-k-i+2)\\&\quad\cdot\Big(\sum_{l=\max(0,m-c-i+2)}^{\min(k-c-i+1,m-c-1)}\Delta_4^{c,l,i}+\sum_{l=\max(0,m-c-i+2)}^{\min(k-c-i+2,m-c-1)}\Delta_5^{c,l,i}\Big)\sigma_c^2\to\delta_1^2,\end{aligned}$$

由 m- 相依平稳序列的中心极限定理得：$n^{-1/2}\sum\limits_{i=2}^{n-k+1}T_i\xrightarrow{L}N(0,\delta_1^2)$. 应用 Slutsky 定理有：$n^{-1/2}U_n^{k,m}\xrightarrow{L}N(0,\delta_1^2)$. 总结以上得如下定理：

定理 3.1 设 $X_1,\cdots,X_n$ 是从分布为 F 的母体中抽取的随机样本，$g(x_1,\cdots,x_m)$ 是对称函数，且 $\mathrm{E}_Fg(X_1,\cdots,X_m)=\theta(F)$, $\mathrm{E}_F[g^2(X_1,\cdots,X_m)]\le M<\infty$, 那么在 k 为常数时，由 (1.1) 构造的统计量 $U_n^{k,m}$ 有极限分布

$$n^{-1/2}(U_n^{k,m}-d\theta(F))\xrightarrow{L}N(0,\delta_1^2). \tag{3.2}$$

其中 δ_1^2 见 (3.1) 式．由 (2.1) 式知，$d=n\binom{k-1}{m-1}+c$, c 是一个常数，所以 (3.2) 式亦可写成

$$\binom{k-1}{m-1}\sqrt{n}\Big(\frac{U_n^{k,m}}{d}-\theta(F)\Big)\xrightarrow{L}N(0,\delta_1^2).$$

3.2 在 $k=O(n^{1-\varepsilon})$, $0<\varepsilon<1$ 时 $U_n^{k,m}$ 的极限分布

由于 $k=o(n)$, 当 n 充分大时，$k\le n/2$ 成立，故 $U_n^{k,m}$ 的方差为 (2.8) 式．令

$$I_1=\sum_{c=1}^{m}\Delta_1^c\sigma_c^2,\qquad I_2=(n-k)\sum_{c=1}^{m}\Delta_2^c\sigma_c^2,\qquad I_3=2\sum_{i=2}^{k}\sum_{c=1}^{\min(m-1,k-i+1)}\sum_{l=\max(0,m-c-i+1)}^{\min(k-c-i+1,m-c-1)}\Delta_3^{c,l,i}\sigma_c^2,$$

$$I_4=2\sum_{i=3}^{k+1}\sum_{c=1}^{\min(m-1,k-i+2)}(n-k-i+2)\Big(\sum_{l=\max(0,m-c-i+2)}^{\min(k-c-i+1,m-c-1)}\Delta_4^{c,l,i}+\sum_{l=\max(0,m-c-i+2)}^{\min(k-c-i+2,m-c-1)}\Delta_5^{c,l,i}\Big)\sigma_c^2,$$

则 $\mathrm{Var}_F(U_n^{k,m})=I_1+I_2+I_3+I_4$. 下面分别讨论 I_1, I_2, I_3, I_4.

1) 由于 $\Delta_1^1=O(k^{2m-1})$, $\Delta_1^j=O(k^{2m-j})$, $j>1$, 所以 $I_1=\Delta_1^1\sigma_1^2+o(k^{2m-1})$;

2) 由于 $\Delta_2^1=O(k^{2m-2})$, $\Delta_2^j=O(k^{2m-1-j})$, $j>1$, 所以 $I_2=n\Delta_2^1\sigma_1^2+o(nk^{2m-2})$;

3) 由于 $\Delta_3^{c,l,i}=O(k^{l+m})$, 由此可知 l 取得越大，$\Delta_3^{c,l,i}$ 的阶数越高，而 l 最大是在 $c=1$ 时，这时 $l=m-2$, 所以 $I_3=2\sum\limits_{i=2}^{k}\Delta_3^{1,m-2,i}\sigma_1^2+o(k^{2m-1})$;

4) 类似于 I_3 的分析可知 $I_4 = 2n\sum_{i=3}^{k+1}(\Delta_4^{1,m-2,i}+\Delta_5^{1,m-2,i})\sigma_1^2 + o(nk^{2m-2})$.

故

$$\mathrm{Var}_F(U_n^{k,m}) = n\Delta_2^1\sigma_1^2 + 2\sum_{i=2}^{k}\Delta_3^{1,m-2,i}\sigma_1^2 + 2n\sum_{i=3}^{k+1}(\Delta_4^{1,m-2,i}+\Delta_5^{1,m-2,i})\sigma_1^2 + o(nk^{2m-2}). \tag{3.3}$$

令

$$\delta_2^2 = \Big(\frac{1}{(m-1)!}\Big)^2 + 2\Big[\Big(\frac{1}{(m-2)!}\Big)^2\frac{1}{m} + \frac{1}{(m-1)(m-1)!(m-2)!}\Big]. \tag{3.4}$$

则 $\mathrm{Var}_F(U_n^{k,m}) = nk^{2m-2}\delta_2^2\sigma_1^2 + o(nk^{2m-2})$.

下面分段讨论 $k^{1-m}n^{-1/2}U_n^{k,m}$ 的极限分布．令

$$k^{1-m}n^{-1/2}U_n^{k,m} = k^{1-m}n^{-1/2}\Big(T_1 + \sum_{i=0}^{q-1}A_i + \sum_{i=0}^{q-1}B_i + C_q\Big), \tag{3.5}$$

其中：$A_i = \sum_{j=i(t+k-1)+2}^{i(t+k-1)+t+1}T_j$, $B_i = \sum_{j=i(t+k-1)+t+2}^{i(t+k-1)+t+k}T_j$, $C_q = \sum_{j=q(t+k-1)+1}^{n-k+1}T_j$, $(i=0,1,\cdots,q-1)$. 这里要求，$k/t\to 0$, $t/n\to 0$, $tq/n\to 1$, $kq/n\to 0$, 由 $k=O(n^{1-\varepsilon})$ 知，这样的 t, q 是存在的 (比如取 $t=O(n^{1-\varepsilon/2})$, $q=O(n^{\varepsilon/2})$). 以下分别讨论 (3.5) 式的各部分．

1) 由于 $I_1 = \Delta_1^1\sigma_1^2 + o(k^{2m-1})$, 所以 $\mathrm{Var}(k^{1-m}n^{-1/2}T_1) = k^{2-2m}n^{-1}I_1 = o(1)$, 从而

$$k^{1-m}n^{-1/2}T_1 \xrightarrow{P} 0. \tag{3.6}$$

2) 由于 $\{B_i\}$ 是 i.i.d., 故

$$\begin{aligned}
\mathrm{Var}\Big(k^{1-m}n^{-1/2}\sum_{i=0}^{q-1}B_i\Big) &= k^{2-2m}n^{-1}q\mathrm{Var}(B_0) = k^{2-2m}n^{-1}q\mathrm{E}\Big(\sum_{i=2}^{k}T_i\Big)^2 \\
&= k^{2-2m}n^{-1}q\Big((k-1)\sum_{c=1}^{m}\Delta_2^c\sigma_c^2 + 2\sum_{i=3}^{k+1}\sum_{c=1}^{\min(m-1,k-i+2)} \\
&\quad\cdot(k-i+1)\Big(\sum_{l=\max(0,m-c-i+2)}^{\min(k-c-i+1,m-c-1)}\Delta_4^{c,l,i} + \sum_{l=\max(0,m-c-i+2)}^{\min(k-c-i+2,m-c-1)}\Delta_5^{c,l,i}\Big)\sigma_c^2\Big) \\
&= k^{2-2m}n^{-1}q\cdot O(k^{2m-1}) = o(1),
\end{aligned}$$

故

$$k^{1-m}n^{-1/2}\sum_{i=0}^{q-1}B_i \xrightarrow{P} 0. \tag{3.7}$$

3) 由于

$$\begin{aligned}
\mathrm{Var}(k^{1-m}n^{-1/2}C_q) &= k^{2-2m}n^{-1}\mathrm{E}\Big(\sum_{i=2}^{n-1+k-q(t+k-1)}T_i\Big)^2 \\
&= k^{2-2m}n^{-1}(n-k+1-q(t+k-1))\cdot O(k^{2m-2}) = o(1),
\end{aligned}$$

所以

$$k^{1-m}n^{-1/2}C_q \xrightarrow{P} 0. \tag{3.8}$$

4) 由于 $\{A_i\}$ 是 i.i.d., 且

$$\begin{aligned}
&\mathrm{Var}(A_0) = \mathrm{E}\Big(\sum_{i=2}^{t+1}T_i\Big)^2 \\
=&\ t\sum_{c=1}^{m}\Delta_2^c\sigma_c^2 + 2\sum_{i=3}^{k+1}\sum_{c=1}^{\min(m-1,k-i+2)}(t-i+2)\Big(\sum_{l=\max(0,m-c-i+2)}^{\min(k-c-i+1,m-c-1)}\Delta_4^{c,l,i} + \sum_{l=\max(0,m-c-i+2)}^{\min(k-c-i+2,m-c-1)}\Delta_5^{c,l,i}\Big)\sigma_c^2 \\
=&\ t\Delta_2^1\sigma_1^2 + 2\sum_{i=3}^{k+1}(t-i)(\Delta_4^{1,m-2,i}+\Delta_5^{1,m-2,i})\sigma_1^2 + o(tk^{2m-2}) = tk^{2m-2}\delta_2^2\sigma_1^2 + o(tk^{2m-2}),
\end{aligned}$$

所以由独立同分布中心极限定理知： $k^{1-m}(qt\delta_2^2\sigma_1^2)^{-1/2}\sum_{i=0}^{q-1}A_i \xrightarrow{L} N(0,1)$, 即

$$k^{1-m}n^{-1/2}\sum_{i=0}^{q-1}A_i \xrightarrow{L} N(0,\delta_2^2\sigma_1^2). \tag{3.9}$$

将 (3.5)-(3.9) 与 Slutsky 定理结合起来得： $k^{1-m}n^{-1/2}U_n^{k,m} \xrightarrow{L} N(0,\delta_2^2\sigma_1^2)$.

总结以上得如下定理：

定理 3.2 设 $X_1,\cdots,X_n$ 为从分布为 F 的母体中抽取的随机样本， $g(x_1,\cdots,x_m)$ 是对称函数，且 $\mathsf{E}_F g(X_1,\cdots,X_m)=\theta(F)$, $\mathsf{E}_F[g^2(X_1,\cdots,X_m)]\le M<\infty$, 那么在 $k=o(n)$ 时，由 (1.1) 构造的统计量 $U_n^{k,m}$ 有极限分布

$$k^{1-m}n^{-1/2}(U_n^{k,m}-d\theta(F)) \xrightarrow{L} N(0,\delta_2^2\sigma_1^2), \tag{3.10}$$

其中， δ_2^2 见 (3.4) 式．又 $d=(k^m+O(k^{m-1}))/m!+(n-k)(k^{m-1}+O(k^{m-2}))/(m-1)!$, 故 (3.10) 亦可写成

$$\frac{\sqrt{n}}{(m-1)!}\Big(\frac{U_n^{k,m}}{d}-\theta(F)\Big) \xrightarrow{L} N(0,\delta_2^2\sigma_1^2).$$

3.3 在 $k=O(n)$ 时 $U_n^{k,m}$ 的极限分布

由于在 $k\le n/2$ 时与 $k>n/2$ 时，统计量 $U_n^{k,m}$ 的方差表达形式不同，所以分两种情况讨论其极限分布．

3.3.1 在 $k=O(n)$ 且 $k\le n/2$ 时 $U_n^{k,m}$ 的极限分布

$k=O(n)$ 且 $k\le n/2$ 等价于 $k=cn+u$, 其中 $0<c\le 1/2$, u 是个常数．由于 u 是一有限的常数，它对 $U_n^{k,m}$ 的极限分布没有影响．不失一般性，仅讨论 $k=cn$ 或者 $n=bk$, 其中 $b=1/c$, 这一情况．这时 $U_n^{k,m}$ 的方差为 (2.8) 式．类似于 (3.4) 式的分析，有：

$$\begin{aligned}\mathsf{Var}_F(U_n^{k,m}) &= \Delta_1^1\sigma_1^2+(n-k)\Delta_2^1\sigma_1^2+2\sum_{i=2}^{k}\Delta_3^{1,m-2,i}\sigma_1^2\\ &\quad+2\sum_{i=3}^{k+1}(n-k-i+2)(\Delta_4^{1,m-2,i}+\Delta_5^{1,m-2,i})\sigma_1^2+o(k^{2m-1})\\ &= k^{2m-1}\delta_3^2\sigma_1^2+o(k^{2m-1}),\end{aligned}$$

其中，

$$\begin{aligned}\delta_3^2 &= b\Big(\frac{1}{(m-1)!}\Big)^2+\frac{2}{m(m-1)!(m-2)!}-2\Big(\frac{1}{(m-2)!}\Big)^2\frac{1}{m(m+1)}\\ &\quad+2(b-1)\frac{1}{m}\Big(\frac{1}{(m-2)!}\Big)^2+\frac{2(b-1)}{(m-1)(m-1)!(m-2)!}-\frac{2}{(m-1)m(m-1)!(m-2)!}.\end{aligned} \tag{3.11}$$

下面讨论 $k^{-(2m-1)/2}U_n^{k,m}$ 的极限分布．令 $Y_n=\sum_{i=1}^{n}g_1(X_i)$, 由于 $\{g_1(X_i)\}$ 是 i.i.d., 故

$$n^{-1/2}Y_n \xrightarrow{L} N(0,\sigma_1^2); \tag{3.12}$$

又由于

$$\begin{aligned}&\mathsf{Var}\,(k^{-(2m-1)/2}U_n^{k,m}-\lambda n^{-1/2}Y_n)\\ =\ &\mathsf{Var}\,(k^{-(2m-1)/2}U_n^{k,m})-2\lambda k^{-(2m-1)/2}n^{-1/2}\mathsf{E}(U_n^{k,m}\cdot Y_n)+\lambda^2\sigma_1^2,\end{aligned} \tag{3.13}$$

其中 λ 为待定常数．而 $\mathsf{Var}\,(k^{-(2m-1)/2}U_n^{k,m})=\delta_3^2\sigma_1^2+o(1)$,

$$\begin{aligned}\mathsf{E}(U_n^{k,m}Y_n) &= \mathsf{E}\Big(T_1\sum_{i=1}^{n}g_1(X_i)\Big)+(n-k)\mathsf{E}\Big(T_2\sum_{i=1}^{n}g_1(X_i)\Big)\\ &= k\binom{k-1}{m-1}\sigma_1^2+(n-k)\Big((k-1)\binom{k-2}{m-2}+\binom{k-1}{m-1}\Big)\sigma_1^2\\ &= \Big[\frac{1}{(m-1)!}+(b-1)\Big(\frac{1}{(m-2)!}+\frac{1}{(m-1)!}\Big)\Big]k^m\sigma_1^2+o(k^m),\end{aligned}$$

故

$$\begin{aligned}&\mathrm{Var}\,(k^{-(2m-1)/2}U_n^{k,m}-\lambda n^{-1/2}Y_n)\\=\;&\delta_3^2\sigma_1^2-2b^{-1/2}\lambda\Big[\frac{1}{(m-1)!}+(b-1)\Big(\frac{1}{(m-2)!}+\frac{1}{(m-1)!}\Big)\Big]\sigma_1^2+\lambda^2\sigma_1^2+o(1).\end{aligned}$$

取 λ 为如下方程的正根，

$$\lambda^2-2b^{-1/2}\Big[\frac{1}{(m-1)!}+(b-1)\Big(\frac{1}{(m-2)!}+\frac{1}{(m-1)!}\Big)\Big]\lambda+\delta_3^2=0. \tag{3.14}$$

所以 $k^{-(2m-1)/2}U_n^{k,m}-\lambda n^{-1/2}Y_n\xrightarrow{P}0$. 由 Slutsky 定理得 $k^{-(2m-1)/2}U_n^{k,m}\xrightarrow{L}N(0,\lambda^2\sigma_1^2)$.

总结以上得如下定理：

定理 3.3 设 $X_1,\cdots,X_n$ 是从分布为 F 的母体中抽取的随机样本， $g(x_1,\cdots,x_m)$ 是对称函数，且 $\mathrm{E}_Fg(X_1,\cdots,X_m)=\theta(F)$, $\mathrm{E}_F[g^2(X_1,\cdots,X_m)]\le M<\infty$, 那么在 $k=cn$ 或者 $n=bk$ 时，其中 $b=1/c$, $0<c\le 1/2$, 由 (1.1) 构造的统计量 $U_n^{k,m}$ 有极限分布

$$k^{-(2m-1)/2}(U_n^{k,m}-d\theta(F))\xrightarrow{L}N(0,\lambda^2\sigma_1^2), \tag{3.15}$$

其中 λ 见 (3.14) 式. 由于 $d=n^m(c^m/m!+(1-c)c^{m-1}/(m-1)!)+O(n^{m-1})$, 所以 (3.15) 亦可写成

$$n^{1/2}(c^{1/2}/m!+(1-c)c^{-1/2}/(m-1)!)\Big(\frac{U_n^{k,m}}{d}-\theta(F)\Big)\xrightarrow{L}N(0,\lambda^2\sigma_1^2).$$

3.3.2 在 $k=O(n)$ 且 $k>n/2$ 时 $U_n^{k,m}$ 的极限分布

$k=O(n)$ 且 $k>n/2$ 等价于 $k=en+u$, 其中 $1/2<e\le 1$, u 是一常数. 不失一般性，仅讨论 $k=en$ 或者 $n=ak$, 其中 $a=1/e$ 这一情况. 这时 $U_n^{k,m}$ 的方差为 (2.9) 式. 类似于 (3.4) 式的分析，有

$$\begin{aligned}\mathrm{Var}_F(U_n^{k,m})\;&=\;\Delta_1^1\sigma_1^2+(n-k)\Delta_2^1\sigma_1^2+2\sum_{i=2}^{n-k+1}\Delta_3^{1,m-2,i}\sigma_1^2\\&\quad+2\sum_{i=3}^{n-k+1}(n-k-i+2)(\Delta_4^{1,m-2,i}+\Delta_5^{1,m-2,i})\sigma_1^2+o(k^{2m-1})\\&=\;k^{2m-1}\delta_4^2\sigma_1^2+o(k^{2m-1}),\end{aligned}$$

其中

$$\begin{aligned}\delta_4^2\;&=\;a\Big(\frac{1}{(m-1)!}\Big)^2+\frac{2(a-1)}{(m-1)!(m-2)!}v_{m-1}^0+2\Big(\frac{a-1}{(m-2)!}\Big)^2v_{m-1}^0\\&\quad-2(a-1)\Big(\frac{1}{(m-2)!}\Big)^2v_{m-1}^1+\frac{2(a-1)^2}{(m-1)!(m-2)!}v_{m-2}^0-\frac{2(a-1)}{(m-1)!(m-2)!}v_{m-2}^1,\end{aligned} \tag{3.16}$$

$$v_t^l=\sum_{i=0}^{t}(-1)^i\frac{1}{i+l+1}\binom{t}{i}(a-1)^{i+l},$$

$$v_t^0=\frac{(2-a)^{t+1}-1}{(t+1)(1-a)},\qquad v_t^1=\frac{(2-t)^{t+1}((t+2)(a-1)+1)-1}{(t+1)(t+2)(1-a)}.$$

下面讨论 $k^{-(2m-1)/2}U_n^{k,m}$ 的极限分布. 由于

$$\begin{aligned}&\mathrm{Var}\,(k^{-(2m-1)/2}U_n^{k,m}-\gamma n^{-1/2}Y_n)\\=\;&\mathrm{Var}\,(k^{-(2m-1)/2}U_n^{k,m})-2\gamma k^{-(2m-1)/2}n^{-1/2}\mathrm{E}(U_n^{k,m}\cdot Y_n)+\gamma^2\sigma_1^2\\=\;&\delta_4^2\sigma_1^2-2a^{-1/2}\Big[\frac{1}{(m-1)!}+(a-1)\Big(\frac{1}{(m-2)!}+\frac{1}{(m-1)!}\Big)\Big]\gamma\sigma_1^2+\gamma^2\sigma_1^2+o(1),\end{aligned}$$

其中 γ 为待定常数. 取 γ 为如下方程的正根,

$$\gamma^2-2a^{-1/2}\Big[\frac{1}{(m-1)!}+(a-1)\Big(\frac{1}{(m-2)!}+\frac{1}{(m-1)!}\Big)\Big]\gamma+\delta_4^2=0. \tag{3.17}$$

故 $k^{-(2m-1)/2}U_n^{k,m}-\gamma n^{-1/2}Y_n\xrightarrow{P}0$. 由 (3.12) 式和 Slutsky 定理得 $k^{-(2m-1)/2}U_n^{k,m}\xrightarrow{L}N(0,\gamma^2\sigma_1^2)$.

总结以上得如下定理:

定理 3.4 设 $X_1,\cdots,X_n$ 为从分布为 F 的母体中抽取的随机样本, $g(x_1,\cdots,x_m)$ 是对称函数, 且 $\mathsf{E}_Fg(X_1,\cdots,X_m)=\theta(F)$, $\mathsf{E}_F[g^2(X_1,\cdots,X_m)]\le M<\infty$, 那么在 $k=en$ 或者 $n=ak$, 其中 $a=1/e$, $1/2<e\le 1$ 时, 由 (1.1) 构造的统计量 $U_n^{k,m}$ 有极限分布

$$k^{-(2m-1)/2}(U_n^{k,m}-d\theta(F))\xrightarrow{L}N(0,\gamma^2\sigma_1^2), \tag{3.18}$$

其中 γ 见 (3.17) 式. (3.18) 亦可写成

$$n^{1/2}(e^{1/2}/m!+(1-e)e^{-1/2}/(m-1)!)\Big(\frac{U_n^{k,m}}{d}-\theta(F)\Big)\xrightarrow{L}N(0,\gamma^2\sigma_1^2).$$

在 $e=1$, 即 $k=n$ 时, $\binom{n}{m}^{-1}U_n^{m,n}$ 就是 U- 统计量 U_n. 这时, $\delta_4^2=(1/(m-1)!)^2$, $\gamma=1/(m-1)!$, $d=\binom{n}{m}$, 从而 $n^{1/2}(U_n-\theta(F))\xrightarrow{L}N(0,m^2\sigma_1^2)$, 这就是 1948 年 Hoeffding 所证明的 U- 统计量的渐进正态性.

参 考 文 献

[1] Ferguson, T.S., *A Course in Large Sample Theory*, Chapman & Hall, 1993.

[2] Hoeffding, W., A class of statistics with asymptotic normal distribution, *Ann. Math. Statist.*, **19**(1948), 293–325.

[3] Sen, P.K. and Singer, J.M., *Large Sample Methods in Statistics: An Introduction with Applications*, Chapman & Hall, 1993.

Asymptotic Normality of k-U Statistic

ZHANG RIQUAN

(*Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, 100080*)

(*Department of Mathematics, Yanbei Normal Colloge, Datong, 037000*)

$U_n^{m,k}=\sum\limits_{1\le i_1<\cdots<i_m\le k}g(X_{i_1},\cdots,X_{i_m})+\sum\limits_{i_m-k+1\le i_1<\cdots<i_{m-1}<i_m}\sum\limits_{k+1\le i_m\le n}g(X_{i_1},\cdots,X_{i_m})$ is called the k-U statistic, where $g(x_1,\cdots,x_m)$ is a symmetric function, k is a positive integer, and equal to or less than n and probably dependent on n. This statistic is a class of statistic and when $k=n$, $\binom{n}{m}^{-1}U_n^{m,n}$ is U-statistic. The asymptotic normality of the k-U statistic will be studied in this paper.