应用概率统计 第九卷
第四期 1993年11月

Chinese Journal of Applied Probability
and Statistics Vol. 9 No. 4, Nov., 1993

一种 Wilcoxon 型统计量的中心极限定理

朱 力 行
(中国科学院应用数学研究所,北京,100080)

摘 要

本文给出了一种类似于两样本 Wilcoxen 统计量的秩和统计量的中心极限定理,特别地当变量服从球对称分布 P 时,其极限分布与分布 P 无关. 此统计量可对分布的球对称性进行初步的检验.

§1. 引 言

球对称分布,在统计理论和应用上起着一定的作用. 在一元情形,人们提出了各种非参数型的统计量进行对称性检验. 如 Wilcoxon 统计量. 然而,目前对多元分布的球对称性的检验尚比较困难.本文考虑一种 Wilcoxon 型统计量,对球对称性进行必要性检验.

一个分布为球对称分布,其充分和必要条件是任何方向上的一维边际分布都是同分布.由此对 d 维分布,随机地选取 d 个正交方向,$\boldsymbol{a}_1, \cdots, \boldsymbol{a}_d$(选取的方法将在下面说明)通过已经得到的观察值 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n$,得到投影观察值 $\boldsymbol{a}_i^\tau \boldsymbol{x}_j$, $j=1, \cdots, n$; $i=1, \cdots, d$,然后构造下述随机变量和:

$$g_n(\boldsymbol{a}_1, \boldsymbol{a}_i)=\frac{1}{n(n-1)} \sum_{k=1}^{n} \sum_{j\neq k}^{n} I(\boldsymbol{a}_i^\tau \boldsymbol{x}_k \leqslant \boldsymbol{a}_1^\tau \underset{\sim}{x}_j), \tag{1.1}$$

$i=2, \cdots, d$,其中 $I(A)$ 表示集合 A 的示性函数,$g_n(\boldsymbol{a}_1, \boldsymbol{a}_i)$ 实际上相当于做了一个随机坐标旋转后,样本各分量之间的秩和统计量. 由于各分量之间具有相依性,因此它与两样本 Wilcoxon 统计量并不相同.但是,类似于 Wilcoxon 统计量. $g_n(\boldsymbol{a}_1, \boldsymbol{a}_i)$ 也可以用于检验 $\boldsymbol{a}_1^\tau \underset{\sim}{x}$ 与 $\boldsymbol{a}_i^\tau \boldsymbol{x}$ 的分布是否相同. 因此 $\max\limits_{2\leqslant i\leqslant d} g_n(\boldsymbol{a}_1, \boldsymbol{a}_i)$ 可以用于检验 $\boldsymbol{a}_1^\tau \boldsymbol{x}, \cdots, \boldsymbol{a}_i^\tau \boldsymbol{x}$ 的分布是否相同,也就是说,对 $\boldsymbol{x}$ 的分布的球对称性给了一个必要性的检验.

上面提到的随机选取 d 个正交方向,主要是为增加检验的可信程度.我们利用 Anderson 等(1985)[2]提出的产生随机正交矩阵的方法选取这 d 个方向,如下所述.

$$G_{ij}=\begin{array}{c} \\ i \\ \\ j \\ \\ \end{array}\begin{array}{c} \begin{array}{ccccc} i & & & j & \end{array} \\ \begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \cos\theta_{ij} & \boldsymbol{0} & -\sin\theta_{ij} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \sin\theta_{ij} & \boldsymbol{0} & \cos\theta_{ij} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I} \end{pmatrix} \end{array}$$

本文1990年9月15日收到,1991年10月14日收到修改稿.

是 $d\times d$ 矩阵，其中 $-\pi/2\leqslant\theta_{ij}\leqslant\pi/2$. 则每一个 $d\times d$ 正交矩阵 G 可以表示成

$$G=(G_{12},\cdots G_{1d})(G_{23}\cdots G_{2d})\cdots(G_{(d-1)d})D_\varepsilon,$$

这里 $D_\varepsilon=\mathrm{diag}(\varepsilon_1,\cdots,\varepsilon_d)$, $\varepsilon_i=\pm1$, $i=1,\cdots,d$.

设 G 均匀分布于所有 $d\times d$ 正交矩阵所成的集合 $\mathbf{O}(d)$ 中，如 Adderson 等指出：

(i) 所有的 θ_{ij}, $1\leqslant i<j\leqslant d$, ε_i, $i=1,\cdots,d$ 相互独立；

(ii) $\theta_{ij}(i<j)$ 具有密度函数

$$p_{ij}(\theta)=\frac{\pi(\cos\theta)^{j-i}}{B\left(1/2,\ \frac{j-i+1}{2}\right)}\qquad -\frac{\pi}{2}<\theta<\frac{\pi}{2};$$

其中 $B(\cdot,\cdot)$ 为 β-函数；

(iii) $P(\varepsilon_i=\pm1)=1/2$, $i=1,\cdots,d$.

因此由 Monte Carlo 方法可以去产生 ε_i 和 θ_{ij}，进而产生一个随机正交矩阵 G，我们把 G 的每一个列向量作为我们所需要的正交方向.

对于 $\max\limits_{2<r<d} g_n(\boldsymbol{a}_1,\boldsymbol{a}_i)$ 的极限随机变量(将由定理 3 给出)的分布，我们可以在 Biometrika 表(Pearson 和 Hartley 编)[3]中查到 1% 和 5% 置信水平的临界值.

在本文中，我们给出当分布 P 为球对称时，随机向量和 $\{g_n(\boldsymbol{a}_1,\boldsymbol{a}_i):\ i=2,\cdots,d\}$ 的中心极限定理和 $\max\limits_{2\leqslant i\leqslant d} g_n(\boldsymbol{a}_1,\boldsymbol{a}_i)$ 的极限分布，极限分布与 P 无关.

如上所说，我们在一定意义下，给出了一个推广的两样本 Wilcoxon 统计量的极限性质，当 P 不是球对称时，也可以得到中心极限定理，但这时极限正态分布可能与 P 有关，这种情况的初步讨论放在第三节.

§2. 主 要 结 果

先给出球对称分布的几个性质，以便我们对主要结果的讨论. 以下设 $d\geqslant3$.

引理 1 设 $\boldsymbol{x}=(x_1,\cdots,x_d)^T$，则 $\boldsymbol{x}$ 服从球对称分布当且仅当对每一个 R^d 中的单位向量

$$\boldsymbol{a}^T\boldsymbol{x}\overset{d}{=}x_1, \tag{2.1}$$

这里记号"$\overset{d}{=}$"表示依分布相等.

(参见[3]定理 2.5).

引理 2 若原点不是 P 的原子，即 $P(\boldsymbol{x}=\boldsymbol{0})=0$，则 P 的任一个 $l(<d)$ 维边际分布具有形如 $f(x_1^2+\cdots+x_l^2)$ 的密度函数. 进一步在给定 $x_{l+1},\cdots,x_d$ 时，$x_1,\cdots,x_l$ 的条件分布密度函数具有 $f(x_1^2+\cdots+x_d^2)/f_1(x_{l+1}^2+\cdots+x_d^2)$ 的形式.

(参见[3]定理 2.2.5).

记 $G(t)=P(x_1\leqslant t)$.

引理 3 若原点不是 P 的原子，则

$$E\bar{G}(\boldsymbol{a}_1^T\boldsymbol{x})\tilde{G}(\boldsymbol{a}_2^T\boldsymbol{x})=\frac{1}{4}. \tag{2.2}$$

这里 $\bar{G}(\cdot)=G(\cdot)$ 或者 $1-G(\cdot)$，$\tilde{G}(\cdot)=G(\cdot)$ 或者 $1-G(\cdot)$.

证明 由球对称性和引理 2，$(\boldsymbol{a}_1^T\boldsymbol{x},\ \boldsymbol{a}_i^T\boldsymbol{x})$ 服从密度函数形如 $f(y_1^2+y_2^2)$ 的球对称分布，且 $1-G(\boldsymbol{a}_1^T\boldsymbol{x})=G(-\boldsymbol{a}_1^T\boldsymbol{x})$. 由此可证，

$$E(1-G(\boldsymbol{a}_1^T\boldsymbol{x}))G(\boldsymbol{a}_2^T\boldsymbol{x})=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(1-G(\boldsymbol{a}_1^T\boldsymbol{x}))G(\boldsymbol{a}_2^T\boldsymbol{x})f((\boldsymbol{a}_1^T\boldsymbol{x})^2+(\boldsymbol{a}_2^T\boldsymbol{x})^2)d\boldsymbol{a}_1^T\boldsymbol{x}d\boldsymbol{a}_2^T\boldsymbol{x}$$

$$=EG(\boldsymbol{a}_1^T\boldsymbol{x})G(\boldsymbol{a}_2^T\boldsymbol{x}). \tag{2.3}$$

$$EG(\boldsymbol{a}_1^T\boldsymbol{x})(1-G(\boldsymbol{a}_2^T\boldsymbol{x}))=EG(\boldsymbol{a}_1^T\boldsymbol{x})G(\boldsymbol{a}_2^T\boldsymbol{x}). \tag{2.4}$$

所以

$$1=E(1-G(\boldsymbol{a}_1^T\boldsymbol{x})+G(\boldsymbol{a}_1^T\boldsymbol{x}))(1-G(\boldsymbol{a}_2^T\boldsymbol{x})+G(\boldsymbol{a}_2^T\boldsymbol{x}))=4EG(\boldsymbol{a}_1^T\boldsymbol{x})G(\boldsymbol{a}_2^T\boldsymbol{x}). \tag{2.5}$$

则证得(2.2)式.

我们知道$(\boldsymbol{a}_1^T\boldsymbol{x},\ \boldsymbol{a}_2^T\boldsymbol{x})$服从球对称分布,现记 $G(t|\boldsymbol{a}_2^T\boldsymbol{x})$为在 $\boldsymbol{a}_2^T\boldsymbol{x}$ 给定时 $\boldsymbol{a}_1^T\boldsymbol{x}$ 的条件分布函数.

引理4

$$EG(\boldsymbol{a}_2^T\boldsymbol{x}_1|\boldsymbol{a}_2^T\boldsymbol{x}_2)=EG(\boldsymbol{a}_2^T\boldsymbol{x}_1|-\boldsymbol{a}_2^T\boldsymbol{x}_2)=1/2, \tag{2.6}$$

$$EG(\boldsymbol{a}_2^T\boldsymbol{x}_1|\boldsymbol{a}_2^T\boldsymbol{x}_2)G(\boldsymbol{a}_2^T\boldsymbol{x}_2|\boldsymbol{a}_2^T\boldsymbol{x}_1)=1/4 \tag{2.7}$$

证明 由球对称性和引理2,

$$1-G(\boldsymbol{a}_2^T\boldsymbol{x}_1|\boldsymbol{a}_2^T\boldsymbol{x}_2)=G(-\boldsymbol{a}_2^T\boldsymbol{x}_1|\boldsymbol{a}_2^T\boldsymbol{x}_2),$$

$$EG(\boldsymbol{a}_2^T\boldsymbol{x}_1|\boldsymbol{a}_2^T\boldsymbol{x}_2)=EG(-\boldsymbol{a}_2^T\boldsymbol{x}_1|\boldsymbol{a}_2^T\boldsymbol{x}_2).$$

由于 $G(t|\boldsymbol{a}_2^T\boldsymbol{x}_2)=G(t|-\boldsymbol{a}_2^T\boldsymbol{x}_2)$. 则(2.6)得证. (2.7)类似地证明.

记

$$h_{nk}(\boldsymbol{a}_1,\ \boldsymbol{a}_i)=\frac{1}{n-1}\sum_{j\neq k}^{n}\left\{I(\boldsymbol{a}_i^T\boldsymbol{x}_k<\boldsymbol{a}_1^T\boldsymbol{x}_j)-\frac{1}{2}\right\},\quad i=2,\ \cdots,\ d$$

$$\tilde{h}_{nk}(\boldsymbol{a}_1,\ \boldsymbol{a}_i)=h_{nk}(\boldsymbol{a}_1,\ \boldsymbol{a}_i)-E(h_{nk}(\boldsymbol{a}_1,\ \boldsymbol{a}_i)|\mathscr{F}_n),\quad i=2,\ \cdots,d.$$

$$=\frac{1}{n-1}\sum_{j\neq k}^{n}\{I(\boldsymbol{a}_i^T\boldsymbol{x}_k<\boldsymbol{a}_1^T\boldsymbol{x}_j)-G(\boldsymbol{a}_1^T\boldsymbol{x}_j|\boldsymbol{a}_1^T\boldsymbol{x}_k)\}$$

其中 $\mathscr{F}_n$ 为由 $\boldsymbol{a}_1^T\boldsymbol{x}_1,\ \cdots\boldsymbol{a}_1^T\boldsymbol{x}_n$ 张成的 σ-域.

由于 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n$ 独立同分布,则知当 $\boldsymbol{a}_1^T\boldsymbol{x}_1,\cdots,\boldsymbol{a}_1^T\boldsymbol{x}_n$ 给定时$(\boldsymbol{a}_T^i\boldsymbol{x}_k:\ i=2,\cdots,d),\ \ k=1,2,\ \cdots,\ n$ 独立服从相同分布. $\{h_{rk}(\boldsymbol{a}_1,\ \boldsymbol{a}_i):\ i=2,\ \cdots,\ d\},\ k=1,\ 2,\ \cdots n$ 在 $\boldsymbol{a}_1^T\boldsymbol{x}_1,\ \cdots,\ \boldsymbol{a}_1^T\boldsymbol{x}_n$ 给定时,分别仅依赖于$\{\boldsymbol{a}_i^T\boldsymbol{x}_k,\ i=2,\ \cdots,\ d\},\ k=1,\ \cdots,\ n.$ 因此

$$\{h_{nk}(\boldsymbol{a},\ \boldsymbol{a}_i):i=2,\ \cdots,\ d\},\quad k=1,\ 2,\ \cdots,\ n.$$

$$\{\tilde{h}_{nk}(\boldsymbol{a}_1,\ \boldsymbol{a}_i):i=2,\ \cdots,\ d\},\quad k=1,\ \cdots,\ n.$$

$$\{\tilde{h}_{nk}(\boldsymbol{a}_1,\ \boldsymbol{a}_i)\tilde{h}_{nk}(\boldsymbol{a}_1,\ \boldsymbol{a}_j):2\leqslant i,\ j\leqslant d\},\quad k=1,\ \cdots,\ n.$$

都是相互独立的随机行向量序列.

引理5 $\dfrac{1}{n}\sum_{k=1}^{n}E\{\{\tilde{h}_{nk}(\boldsymbol{a}_1,\ \boldsymbol{a}_i):i=2,\ \cdots,\ d\}^T\{\tilde{h}_{nk}(\boldsymbol{a}_1,\ \boldsymbol{a}_i):i=2,\ \cdots,\ d\}|\mathscr{F}_n\}.$

依概率收敛于对角矩阵 $\dfrac{1}{12}I_{d-1}$.

证明 我们只要证明对每一对$(i,\ j),\ 2\leqslant i,\ j\leqslant d$ 当 $n\to\infty$ 时

$$E\left(\frac{1}{n}\sum_{k=1}^{n}E(\tilde{h}_{nk}(\boldsymbol{a}_1,\ \boldsymbol{a}_i)\tilde{h}_{nk}(\boldsymbol{a}_1,\ \boldsymbol{a}_j)|\mathscr{F}_n)-\frac{1}{12}\delta_{ij}\right)^2\triangleq I_n\to 0, \tag{2.8}$$

这里 $\delta_{ij}=1,\ i=j;\quad =0,\ i\neq j.$

记

$$X_{nk}=E(\tilde{h}_{nk}(\boldsymbol{a}_1,\ \boldsymbol{a}_i)\tilde{h}_{nk}(\boldsymbol{a}_1,\ \boldsymbol{a}_j)|\mathscr{F}_n).$$

$$EX_{nk}=E\tilde{h}_{nk}(\boldsymbol{a}_1,\ \boldsymbol{a}_i)\tilde{h}_{nk}(\boldsymbol{a}_1,\ \boldsymbol{a}_j)$$

这里

$$\tilde{h}_{nk}(\boldsymbol{a}_1,\ \boldsymbol{a}_i)=\frac{1}{n-1}\sum_{j\neq k}^{n}(I(\boldsymbol{a}_i^T\boldsymbol{x}_k<\boldsymbol{a}_1^T\boldsymbol{x}_j)-G(\boldsymbol{a}_1^T\boldsymbol{x}_j|\boldsymbol{a}_1^T\boldsymbol{x}_k))$$

则有

$$X_{nk}=\frac{1}{(n-1)^4}\sum_{j_1\neq k}^{n}\sum_{l_1\neq k}^{n}[E(I(\boldsymbol{a}_1^T\boldsymbol{x}_k<\boldsymbol{a}_1^T\boldsymbol{x}_{j_1})I(\boldsymbol{a}_1^T\boldsymbol{x}_k<\boldsymbol{a}_1^T\boldsymbol{x}_{l_1})|\mathscr{F}_n)$$
$$-G(\boldsymbol{a}_1^T\boldsymbol{x}_{j_1}|\boldsymbol{a}_1^T\boldsymbol{x}_k)G(\boldsymbol{a}_1^T\boldsymbol{x}_{l_1}|\boldsymbol{a}_1^T\boldsymbol{x}_k)] \tag{2.9}$$

现证明 $n\to\infty$ 时

$$E\Big(\frac{1}{n}\sum_{k=1}^{n}(X_{nk}-EX_{nk})\Big)^2\to 0. \tag{2.10}$$

为此只要证明对每一对 $m\neq k$, 在 $n\to\infty$ 时, 一致地有

$$E(X_{nk}-EX_{nk})(X_{nm}-EX_{nm})\to 0. \tag{2.11}$$

将 $X_{nk}\cdot X_{nm}$ 展开, 利用(2.9), 有

$$EX_{nk}X_{nm}=\frac{1}{(n-1)^4}\sum_{j_1\neq k}\sum_{j_2\neq m}\sum_{l_1\neq k}\sum_{l_2\neq m}[E\{E(I(\boldsymbol{a}_1^T\boldsymbol{x}_k<\boldsymbol{a}_1^T\boldsymbol{x}_{j_1})I(\boldsymbol{a}_j^T\boldsymbol{x}_k<\boldsymbol{a}_1^T\boldsymbol{x}_{l_1})|\mathscr{F}_n\}$$
$$\cdot E(I(\boldsymbol{a}_1^T\boldsymbol{x}_m<\boldsymbol{a}_1^T\boldsymbol{x}_{j_2})I(\boldsymbol{a}_j^T\boldsymbol{x}_m<\boldsymbol{a}_1^T\boldsymbol{x}_{l_2})|\mathscr{F}_n)\}.$$
$$-E\{G(\boldsymbol{a}_1^T\boldsymbol{x}_{j_1}|\boldsymbol{a}_1^T\boldsymbol{x}_k)G(\boldsymbol{a}_1^T\boldsymbol{x}_{l_1}|\boldsymbol{a}_1^T\boldsymbol{x}_k)G(\boldsymbol{a}_1^T\boldsymbol{x}_{j_2}|\boldsymbol{a}_1^T\boldsymbol{x}_m)G(\boldsymbol{a}_1^T\boldsymbol{x}_{l_2}|\boldsymbol{a}_1^T\boldsymbol{x}_m)\}]$$
$$\triangleq\frac{1}{(n-1)^4}\sum_{j_1\neq k}\sum_{j_2\neq m}\sum_{l_1\neq m}\sum_{l_2\neq m}[I_1(j_1,\ j_2,\ l_1,\ l_2)-I_2(j_1,\ j_2,\ l_1,\ l_2)]$$

当 $j_1\neq j_2\neq l_1\neq l_2\neq k\neq m$ 时, 由独立性, 可知

$I_1(j_1,\ j_2,\ l_1,\ l_2)$, $I_2(j_1,\ j_2,\ l_1,\ l_2)$与 $EX_{nk}EX_{nm}$ 的展开式对应项的值相同.

而这样的项的个数共有$(n-2)(n-3)(n-4)(n-5)$. 因此剩下的项的个数的阶为 $O(n^3)$; 并且 $I(j_1,\ j_2,\ l_1,\ l_2)$的绝对值总小于等于 1, 则知

$$|EX_{nk}X_{nm}-EX_{nk}EX_{nm}|=O\Big(\frac{1}{n}\Big). \tag{2.12}$$

下面, 我们仅需证明 $n\to\infty$ 时

$$\frac{1}{n}\sum_{k=1}^{n}E\tilde{h}_{nk}(\boldsymbol{a}_1,\ \boldsymbol{a}_i)\tilde{h}_{nk}(\boldsymbol{a}_1,\ \boldsymbol{a}_j)-\frac{1}{12}\delta_{ij}\to 0. \tag{2.13}$$

不失一般性, 我们仅计算 $k=1$ 的值.

$$E\tilde{h}_{n1}(\boldsymbol{a}_1,\boldsymbol{a}_i)\tilde{h}_{n1}(\boldsymbol{a}_1,\boldsymbol{a}_j)$$
$$=\frac{1}{(n-1)^2}\sum_{k=2}^{n}\sum_{l=2}^{n}\{E(I(\boldsymbol{a}_1^T\boldsymbol{x}_1<\boldsymbol{a}_1^T\boldsymbol{x}_k)I(\boldsymbol{a}_j^T\boldsymbol{x}_1<\boldsymbol{a}_1^T\boldsymbol{x}_l)+EG(\boldsymbol{a}_1^T\boldsymbol{x}_k|\boldsymbol{a}_1^T\boldsymbol{x}_1)G(\boldsymbol{a}_1^T\boldsymbol{x}_l|\boldsymbol{a}_1^T\boldsymbol{x}_1)$$
$$-EG(\boldsymbol{a}_1^T\boldsymbol{x}_k|\boldsymbol{a}_1^T\boldsymbol{x}_1)I(\boldsymbol{a}_j^T\boldsymbol{x}_1<\boldsymbol{a}_1^T\boldsymbol{v}_l)-EG(\boldsymbol{a}_1^T\boldsymbol{x}_l|\boldsymbol{a}_1^T\boldsymbol{x}_1)I(\boldsymbol{a}_j^T\boldsymbol{x}_1<\boldsymbol{a}_1^T\boldsymbol{x}_k)\}$$
$$\triangleq\frac{1}{(n-1)^2}\sum_{k=2}^{n}\sum_{l=2}^{n}\{J_1(k,\ l)+J_2(k,\ l)+J_3(k,\ l)+J_4(k,\ l)\}. \tag{2.14}$$

由引理 3 和引理 4 可得, 当 $i\neq j$ 时,

$$J_1(k,\ l)=\begin{cases}1/4, & k\neq l\\ 1-EG(\boldsymbol{a}_i^T\boldsymbol{x}_1\vee\boldsymbol{a}_j^T\boldsymbol{x}_1), & k=l.\end{cases}$$

$$J_3(k,\ l)=J_4(l,\ k)=\begin{cases}1/4, & k\neq l,\\ J_3(k,\ l), & k=l.\end{cases}$$

$$J_2(k,\ l)=\begin{cases}1/4, & k=l\\ 1/4, & k=l.\end{cases}$$

这里记号"$\vee$"表示取最大值. 则有

$$\frac{1}{(n-1)^2}\sum_{k=2}^{n}\sum_{l=2}^{n}(J_1(k,\ l)+J_2(k,\ l)-J_3(k,\ l)-J_4(k,\ l))$$

$$-\frac{1}{(n-1)}\left(1-EG(\boldsymbol{a}_i^T x_1 V \boldsymbol{a}_j^T x_1)-\frac{1}{3}\right)\to 0. \quad (n\to\infty) \tag{2.15}$$

当 $i=j$ 时，同样由引理 3 和引理 4 可得

$$J_1(k,\ l)=\begin{cases} E(1-G(\boldsymbol{a}_i^T x_1))^2=1/3, & k\neq l \\ EG(\boldsymbol{a}_1^T x_l)=\dfrac{1}{2}, & k=l, \end{cases}$$

$$J_3(k,\ l)=J_4(l,\ k)=\begin{cases} \dfrac{1}{4}, & k\neq l \\ J_3(k,\ l), & k=l, \end{cases}$$

$$J_2(k,\ l)=\begin{cases} 1/4, & k\neq l \\ 1/4, & k=l, \end{cases}$$

则当 $n\to\infty$ 时有

$$\begin{aligned} &\frac{1}{(n-1)^2}\sum_{k=2}^{n}\sum_{l=2}^{n}(J_1(k,\ l)+J_2(k,\ l)-J_3(k,\ l)-J_4(k,\ l)) \\ &\leqslant\frac{1}{(n-1)^2}\left\{(n-1)(n-2)\frac{1}{12}+(n-1)\right\}\to\frac{1}{12}. \end{aligned} \tag{2.16}$$

因此(2.10)式成立.

定理 1 $\quad \dfrac{1}{\sqrt{n}}\sum_{k=1}^{n}\{(\tilde{h}_{nk}(\boldsymbol{a}_1,\ \boldsymbol{a}_i):\ i=2,\ \cdots,\ d\}$

依分布收敛于多元正态分布 $N\left(0,\ \dfrac{1}{12}I_{d-1}\right)$.

证明 我们只要证明对每一个非零向量 $\gamma=(\gamma_2,\ \cdots,\ \gamma_d)\in R^{d-1}$ 使得

$$\sqrt{\frac{12}{\sum_{i=2}^{d}\gamma_i^2 n}}\sum_{k=1}^{n}\sum_{i=2}^{d}\gamma_i\tilde{h}_{nk}(\boldsymbol{a}_1,\ \boldsymbol{a}_i)$$

依分布收敛于正态分布 $N(0,\ 1)$. 不致混淆，记

$$Y_{nk}=\left\{\sum_{k=1}^{n}E\left(\left[\sum_{i=2}^{d}\gamma_i\tilde{h}_{nk}(\boldsymbol{a}_1,\boldsymbol{a}_i)\right]^2|\mathscr{F}_n\right)\right\}^{-\frac{1}{2}}\sum_{i=2}^{d}\gamma_i\tilde{h}_{nk}(\boldsymbol{a}_1,\boldsymbol{a}_i)$$

由引理 5，只要证明：$\sum_{k=1}^{n}Y_{nk}$ 依分布收敛于 $N(0,\ 1)$.

由通常的办法，考虑特征函数的逐点收敛性. 即当 $n\to\infty$ 时对每一个 $t\in R'$

$$E\left|\exp\left(it\sum_{k=1}^{n}Y_{nk}\right)\to\exp\left(-\frac{1}{2}t^2\right)\right. \tag{2.17}$$

记 $\sigma_{nk}^2=E(Y_{nk}^2|\mathscr{F}_n)$，我们有 $\sum_{k=1}^{n}\sigma_{nk}^2=1$.

选择随机变量$\{\eta_{nk}\}$，在 $\boldsymbol{a}_1^T x_1,\ \cdots,\ \boldsymbol{a}_1^T x_n$ 给定时，η_{nk} 独立且服从 $N(0\ \sigma_{nk}^2)$. 并且与 Y_{nl}，$1\leqslant l<k$ 相互独立，因此 $\sum_{k=1}^{n}\eta_{nk}$ 服从 $N(0,\ 1)$. 为证明(2.17)我们先作一些准备工作. 记

$$S_{nl}=\sum_{k=1}^{l-1}Y_{nk}+\sum_{k=l+1}^{n}\eta_{nk}.$$

则有

$$S_{nl}+Y_{nl}=S_{n(l+1)}+\eta_{n(l+1)}, \quad l=1,\ \cdots,\ n-1. \tag{2.18}$$

又记 $f(t)=\exp(it)$，$f^{(l)}(t)$为 $f(t)$的第 l 阶导函数，

$$R(x,\ y)=f(x+y)-f(x)-yf^{(1)}(x)-\frac{1}{2}y^2f^{(2)}(x).$$

$R(x,\ y)$可以写成 $\frac{1}{6}y^3f^{(3)}(x+\theta_1y)\quad|\theta_1|<1$. 同时，由 $f(\cdot)$的定义，可知 $\sup|f^{(l)}(x)|$ 为有界值，$l=1,\ 2,\ 3$. 又有对某个常数 $c>0$. 和任意的 $x,\ y$.

$$|R(x,\ y)|=\left|\frac{1}{2}y^2f^{(2)}(x+\theta_2y)-\frac{1}{2}y^2f^{(2)}(x)\right|<c|y|^3$$

由此我们知道，对某个 $c>0$. 和任意的 $\varepsilon>0$

$$|R(x,\ y)|<c\{|y|^3I\{|y|<\varepsilon\}+|y|^2I\{|y|<\varepsilon\}\}. \tag{2.19}$$

由条件独立性和 η_{nk} 的选取，我们有

$$E\{R(S_{nl},\ Y_{nl})|\mathscr{F}_n\}=E\left\{\left\{f(S_{nl}+Y_{nl})-f(S_{nl})-\frac{1}{2}Y_{nl}^2f^{(2)}(S_{nl})\right\}\middle|\mathscr{F}_n\right\} \tag{2.20}$$

和

$$E\{R(S_{nl},\ \eta_{nl})|\mathscr{F}_n\}=E\left\{\left\{f(S_{nl}+\eta_{nl})-f(S_{nl})-\frac{1}{2}Y_{nl}^2f^{(2)}(S_{rl})\right\}\middle|\mathscr{F}_n\right\} \tag{2.21}$$

现证明(2.17).

由 S_{nl} 的定义，有 $\sum_{k=1}^{n}Y_{nk}=S_{nn}+Y_{nn}$ $\sum_{k=1}^{n}\eta_{nk}=S_{n1}+\eta_{n1}$. 再由(2.18)

$$\left|E\exp\left(it\sum_{k=1}^{n}Y_{nk}\right)-\exp\left(-\frac{1}{2}t^2\right)\right|<\sum_{l=1}^{n}|E\exp(it(S_{nl}+Y_{nl})-E\exp(it(S_{nl}+\eta_{nl})|$$

$$=\sum_{l=1}^{n}|E\{E(R(S_{nl},\ Y_{nl})|\mathscr{F}_n)-E(R(S_{nl},\ \eta_{nl})|\mathscr{F}_n)\}|$$

$$<2c\sum_{l=1}^{n}E\{E\{|Y_{nl}|^3I(|Y_{nk}|<\varepsilon)|\mathscr{F}_n\}\}+2c\sum_{l=1}^{n}E\{E\{Y_{nl}^2I(|Y_{nl}|>\varepsilon)|\mathscr{F}_n\}\}.$$

$$<2c\varepsilon^3+2c\sum_{l=1}^{n}E\{E\{Y_{nl}^2I(|Y_{nl}|\geqslant\varepsilon)|\mathscr{F}_n\}\}. \tag{2.22}$$

由引理 5，我们有 $\max_{1<l<n}|Y_{nl}|$ 依概率收敛于零，同时就有 $\max_{1<l<n}I(\{|Y_{nl}|\geqslant\varepsilon\}$依概率收敛于零，而又有 $\sum_{l=1}^{n}E(Y_{nl}^2|\mathscr{F}_n)=1$ 为有界随机变量，因此我们利用控制收敛定理，就可以得到对每一个 $\varepsilon>0$，当 $n\to\infty$ 时

$$\sum_{l=1}^{n}E\{E(Y_{nl}^2I(|Y_{nl}|\geqslant\varepsilon)|\mathscr{F}_n)\}\to0 \tag{2.23}$$

由此，我们证得(2.17).

定理 2 $\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\{h_{nk}(x_1,\ x_i):\ i=2,\ \cdots,\ d\}$. 依分布收敛于多元正态分布 $N\left(0,\ \frac{1}{12}(I_{d-1}+V)\right)$. 这里 V 为 $(d-1)\times(d-1)$矩阵. 其中 V 中的每一个元素都是 1.

证明 在 $i=2,\ \cdots,\ d$ 时

$$\sum_{k=1}^{n}E(h_{nk}(x_1,x_i)|\mathscr{F}_n)=\frac{2}{n(n-1)}\sum_{k<j}\{G(x_1^Tx_i|x_1^Tx_k)+G(x_1^Tx_k|x_1^Tx_j)-1\}/2$$

$$\triangleq\frac{2}{n(n-1)}\sum_{k<j}h(x_j,\ x_k)$$

这是一个一样本 U 统计量.

利用引理 4，容易得到

$$Eh(x_1, x_2)=0$$
$$\mathrm{Var}(E(h(x_1, x_2)|x_1))=1/12. \tag{2.24}$$

则知$\left\{\frac{1}{\sqrt{n}}\sum_{k=1}^{n}E(h_{nk}(a_1, a_i)|\mathscr{F}_n): i=2, \cdots, d\right\}$依分布收敛于退化的多元正态分布$N\left(0, \frac{1}{12}V\right)$. 结合定理 1 就有$\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\{h_{nk}(a_1, a_i): i=2, \cdots, d\}$依分布收敛于两个正态随机向量和$X+Y=\{X_2+Y_1, \cdots, X_d+Y_1\}$. 其中$X$服从$N\left(0, \frac{1}{12}I_{d-1}\right)$. Y服从$N\left(0, \frac{1}{12}V\right)$，而同时，容易知道$\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\{\tilde{h}_{nk}(a_1, a_i): i=2, \cdots, d\}$与$\frac{1}{\sqrt{n}}\sum_{k=1}^{n}E(h_{nk}(a_1, a_i)|\mathscr{F}_n): i=2, \cdots, d\}$互不相关，则知$X$与$Y$互不相关，因此$X$与$Y$相互独立. 因此我们得到$X+Y$服从$N\left(0, \frac{1}{12}(I_{d-1}+V)\right)$.

定理 3 对每一个 λ，有

$$\lim_{n\to\infty}P\left\{\max_{2\le i\le d}\sqrt{\frac{12}{n}}\sum_{k=1}^{n}h_{nk}(a_1, a_i)<\lambda\right\}=\int_{-\infty}^{\infty}\Phi^{d-1}(\lambda-y)d\Phi(y). \tag{2.26}$$

其中$\Phi(y)$为标准正态分布函数.

证明 由定理 2，有$\max_{2\le i\le d}\sqrt{\frac{12}{n}}\sum_{k=1}^{n}h_{nk}(_1a, a_i)$依分布收敛于$\sqrt{12}\max_{2\le i\le d}\{X_i+Y_1\}=\max_{2\le i\le d}\{\sqrt{12}X_i\}+\sqrt{12}Y_1$. 而$\sqrt{12}Y_1$, $\sqrt{12}X_i$, $i=2, \cdots, d$. 相互独立并且服从相同分布$N(0, 1)$则容易知道.

$$P\{\max_{2\le i\le d}\{\sqrt{12}X_i+\sqrt{12}Y_1\}\le\lambda\}=\int_{-\infty}^{\infty}\Phi^{d-1}(\lambda-y)d\Phi(y) \tag{2.27}$$

(2.26)成立.

§3. 进一步讨论

对(1.1)中的统计量，我们可以看到它是一个秩和统计量. 当P为球对称分布时，我们得到(1.1)中的统计量的极限分布，与分布P无关，这是与传统的两样本 Wilcoxon 统计量的极限性质相符的. 而当P仅在d个固定的正交方向上具有相同的对称分布. 用上一节的方法，仍然可以得到统计量的渐近正态性，然而将得不到引理 4 中的良好性质，从而使得$\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\{h_{nk}(a_1, a_i): i=2, \cdots, d\}$依分布收敛于多元正态分布$N(0, \tilde{V})$，其中$\tilde{V}$中每一个元素为$E(G_1(-a_i^Tx)-G_i(a_1^Tx))(G_1(-a_j^Tx)-G_j(a_1^Tx))$，$G_i$记为$P$在$a_i$方向上的边际分布，特别地，当每一个$a_i$方向的边际分布为$\left[-\frac{1}{2}, \frac{1}{2}\right]$上的均匀分布时，有$E(G_1(-\underset{\sim}{a}_i^Tx)-G_i(\underset{\sim}{a}_i^Tx))(G_1(-a_j^Tx)-G_j(a_1^Tx))=(a_j^T+a_1^T)Exx^T(a_j+a_1)$. 我们看到这个量是与$x$的分布有关的. 因此用(1.1)的统计量是检验$P$的某$d-1$个正交方向上的边际分布是否与另一个与它们都正交的方向上对称边际分布相同，就不合理了. 由此可以看到这与我们从秩和的角度认为(1.1)的统计量可以检验对称边际分布的异同的直观感觉是有差异的. 另一方面，对于这个统计量检验的功效如何还是一个有趣的问题.

参 考 文 献

[1] Serfling, R. J. (1980): Approximation Theorems of Mathematical Statistics, John Wiley and Sons, New York.

[2] Anderson, T. W., Olkin, I. and Underhill, L. G. (1985), Generation of randon orthegonal matrics, Tech. Rep. No. 6, Econometric Workshop, Stanford University.

[3] Pearson, E. S. and Hartley, H. O. (1956), Biametrika Tables for Statosticians, Vol. I. the Syndics of the Cambridge University Press, Cambridges.

[4] Fang, K. T., Kotz, S. and Ng, K. W. (1990), Syinmetric Multivariate and Related Distributions, Chapman and Hall, London.

[5] Pollard, D. (1984), The Convergence of Stochastic Processes, Springer-Verlag, New York.

[6] Hsing, T. and Carroll, R. J. (in press), An Asymptotic Theory for Sliced Inverse Regressien, *Ann. Statist.*

THE CENTRAL LIMIT THEORY FOR A WILCOXON TYPE STATISTIC

ZHU LIXING

(*Institute of Applied Mathematrcs, Academia Sinica Beijing*, 100080)

In this paper, we obtain the central limit theorem for a rank sum statistic, which is similar to two Sample Wilcoxon statistic. Specially, if the random vectors are distributed with the spherically symmetric distribution P, the limit distribution is independent of P. We do a preliminary. necessarity, test for sphericity by using this statistic.