应用概率统计 第三卷
第四期 1987年11月

Chinese Journal of Applied Probability and Statistics Vol. 3. No. 4. Nov. 1987

随机加权法的渐近展开*

涂冬生 郑忠国
(中国科学院系统科学研究所) (北京大学)

摘 要

本文讨论了样本均值规范化误差 $(\overline{X}-\mu)/\sigma$ 之分布函数 $F_n(x)$ 的模拟问题. 在本文中采用了一种与 Bootstrap 方法不同的方法——随机加权法. 记 F_n^* 为 $\Sigma(X_i-\overline{X})V_i/\sigma^*(\Sigma(X_i-\overline{X})V_i)$ 之分布, 其中 $(V_1, \cdots, V_n)$ 遵从 Dirichlet 分布, σ^{*2} 表示 $X_1, \cdots, X_n$ 固定之下 $\Sigma(X_i-\overline{X})V_i$ 之方差. 本文的主要结论是当 $E|X_i|^3<\infty$ 时, $\sqrt{n}\sup_x|F_n^*(x)-F_n(x)|\to 0$, a.e..

§1. 前 言

Efron 在1979年提出了 Bootstrap 方法[1], 近几年来这个方法受到很多理论和实际工作者的广泛注意. 最近, 郑忠国用随机加权的思想提出了一类比较简便的方法[6], 称为随机加权法. 在[6]中, 利用满足 Dirichlet 分布 $D(1, 1, \cdots, 1)$ 的随机变量 $(V_1, \cdots, V_n)$ 作权, 处理了样本均值误差. 设 $X_1, \cdots, X_n$ 为来自总体 F 的 iid 样本, $EX_i=\mu$, $\overline{X}=\frac{1}{n}\Sigma X_i$. 为了讨论 $\overline{X}-\mu$ 的统计特性, 我们取随机变量 $(V_1, \cdots, V_n)$, 其分布为 Dirichlet 分布 $D(1, 1, \cdots, 1)$, 即它满足 $V_1+\cdots+V_n=1$, 并且 $(V_1, \cdots, V_{n-1})$ 的密度为

$$f(v_1, \cdots, v_{n-1})=\Gamma(n), \quad (v_1, \cdots, v_{n-1})\in S_{n-1}, \tag{1.1}$$

其中 $S_{n-1}=\left\{(v_1, \cdots, v_{n-1}):\ v_i\geqslant 0,\ \sum_{i=1}^{n-1}v_i\leqslant 1\right\}$. 我们用 $D_n=\Sigma(X_i-\overline{X})V_i$ 的统计特性去模拟 $T_n=\overline{X}-\mu$ 的统计特性. [6]中指出, 在 $x_1, x_2, \cdots$ 固定之下, $\sqrt{n}\,D_n$ 的条件分布弱收敛到 $N(0, \sigma^2)$, 其中 $\sigma^2=\mathrm{var}(X_i)$. 这说明在均值误差的情形, 随机加权法在大样本意义下是可用的.

本文中, 我们主要讨论均值误差的随机加权法的渐近展开问题. 我们的目的是要寻找 $\sqrt{n}(\overline{X}-\mu)/\sigma$ 的分布的估计. 众所周知, $\Phi(x)$(标准正态分布)是很好的近似. 但是用 Φ 作为近似, 不可能得到

$$\sqrt{n}\ \{P\{\sqrt{n}(\overline{X}-\mu)/\sigma<x\}-\Phi(x)\}=o(1).$$

我们希望用 $D_n/\overline{D}_n$ 的条件分布来逼近 $\sqrt{n}(\overline{X}-\mu)/\sigma$ 之分布, 其中 $\overline{D}_n^2=\mathrm{var}^*D_n$ 表示 $x_1, \cdots, x_n$ 固定之下 D_n 之方差. 经过研究发现 $D_n/\overline{D}_n$ 之分布的 Edgeworth 展开式与 $\sqrt{n}(\overline{X}-\mu)/\sigma$ 的展式不同, 因而不能得到诸如

本文1986年5月19日收到, 1986年8月25日收到修改稿.
* 本文得到国家科学基金的部分资助.

$$\sqrt{n}\sup_y\left|P\{\sqrt{n}(\bar{X}-\mu)/\sigma\leqslant y\}-P^*\left\{\frac{D_n}{\bar{D}_n}\leqslant y\right\}\right|=o(1)\quad \text{a.e.} \tag{1.2}$$

那样的逼近式. 注意上式中 $P^*(E^*$, var^* 等)表示 $x_1, \cdots, x_n$ 已知条件下的概率(期望、方差), 以后不再单独说明. 为寻找(1.2)的逼近式, 我们转而寻找$(V_1, V_2, \cdots, V_n)$的适当分布. 设$(V_1, V_2, \cdots, V_n)$的联合分布为 $D_n(4, 4, \cdots, 4)$, 即 $V_1, \cdots, V_n$ 满足 $V_1+\cdots+V_n=1$, 并且 $V_1, \cdots, V_{n-1}$ 的联合分布密度为

$$f(x_1, \cdots, x_{n-1})=\frac{\Gamma(4n)}{(\Gamma(4))^n}x_1^3\cdots x_{n-1}^3(1-x_1-x_2\cdots-x_{n-1})^3,\ (x_1, \cdots, x_{n-1})\in S_{n-1}, \tag{1.3}$$

其中 S_{n-1} 与(1.1)中的意义相同. 当 $D_n=\sum x_iV_i-\bar{x}$ 中的$(V_1, \cdots, V_n)$之分布由(1.3)给出时, 其对应的 $D_n/\bar{D}_n$ 之分布可达到(1.2)之逼近式. 下面我们将介绍本文的主要结果.

§2. 主要结果

最近白志东、赵林城在[5]中对于相互独立的随机变量之和的分布函数给出了渐近展式. 下面我们首先介绍他们的结果.

设 $X_1, \cdots, X_n$ 相互独立, $EX_j=0$, $EX_j^2=\sigma_j^2$, 记 $\frac{1}{B_n}\sum X_j$ 的分布为 $F_n(x)$, 其中

$$B_n^2=\sum_{j=1}^n\sigma_j^2>0.$$

记 X_j 的特征函数为 $v_j(t)$, 其 ν 阶半不变量为 $\gamma_{\nu j}$, 记

$$\lambda_{\nu,n}=n^{(\nu-2)/2}B_n^{-\nu}\sum_{j=1}^n\gamma_{\nu j},\quad \nu=2, \cdots, k, \tag{2.1}$$

$$\begin{aligned}Q_{\nu,n}(x)&=\sum{}'(-1)^{\nu+2s}\Phi^{(\nu+2s)}(x)\prod_{m=1}^{\nu}\frac{1}{k_m!}\left(\frac{\lambda_{m+2,n}}{(m+2)!}\right)^{k_m}\\&=-\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\sum{}'H_{\nu+2s-1}(x)\prod_{m=1}^{\nu}\frac{1}{k_m!}\left(\frac{\lambda_{m+2,n}}{(m+2)!}\right)^{k_m},\\&\qquad\nu=1, \cdots, k-2\end{aligned} \tag{2.2}$$

其中 $\Phi(x)$ 为标准正态分布函数,

$$H_m(x)=(-1)^me^{x^2/2}\frac{d^m}{dx^m}e^{-\frac{x^2}{2}}$$

为 m 次 Hermite 多项式, $\sum'$ 表示对满足下述条件的非负整数 $k_1, \cdots, k_\nu$ 求和

$$\sum_{m=1}^{\nu}mk_m=\nu,\qquad \sum_{m=1}^{\nu}k_m=s. \tag{2.3}$$

用 $I(A)$ 记集合 A 的示性函数, 且对任一实数 x, 记

$$Y_{nj}=X_jI(|X_j|\leqslant B_n), \tag{2.4}$$

$$Z_{nj}^{(x)}=X_jI(|X_j|\leqslant B_n(1+|x|)),\quad j=1, \cdots, n, \tag{2.5}$$

$$W_{nj}^{(x)}=X_jI(|X_j|>B_n(1+|x|)),\quad j=1, \cdots, n, \tag{2.6}$$

引理 1 若 $X_1, \cdots, X_n$ 为上述相互独立的随机变量序列, 若对每一个 X_j, 满足 $E|X_j|^k<\infty$, 其中 $k\geqslant3$ 为整数, 则

$$\left| F_n(x)-\Phi(x)-\sum_{\nu=1}^{k-2} Q_{\nu,n}(x)n^{-\nu/2} \right|$$
$$\leqslant C(k)\left\{ (1+|x|)^{-k}B_n^{-k}\sum_{j=1}^{n} E|W_{nj}^{(x)}|^k+(1+|x|)^{-k-1}B_n^{-k-1} \right.$$
$$\left. \times\sum_{j=1}^{n} E|Z_{nj}^{(x)}|^{k+1}+(1+|x|)^{-k-1}n^{k(k+1)/2}\left(\sup_{|t|\geqslant\delta_n}\frac{1}{n}\sum_{j=1}^{n}|v_j(t)|+\frac{1}{2n}\right)^n \right\}, \tag{2.7}$$

其中
$$\delta_n=\frac{1}{12}B_n^2\left(\sum_{j=1}^{n}E|Y_{nj}|^3\right)^{-1},$$
$C(k)$为仅与k有关的常数.

证明见[5],定理1.

现在我们转而讨论随机加权法中分布$P^*\left\{\frac{D_n}{\overline{D}_n}\leqslant x\right\}$之渐近展开问题, 其中
$$D_n=\sum_{i=1}^{n}(x_i-\bar{x})V_i \tag{2.8}$$
而$(V_1, \cdots, V_n)$具有Dirichlet分布$D(4, 4, \cdots, 4)$. 由[4]中关于Dirichlet分布之理论知若$Z_1, \cdots, Z_n\sim \text{iid} f(x)$, $f(x)=2^4x^3e^{-2x}/3!$, $x>0$, 则$Z_i\Big/\sum_{j=1}^{n}Z_j$, $i=1, \cdots, n$具有Dirichlet分布$D(4, 4, \cdots, 4)$. 因此统计量D_n((2.8)式)可以改写成
$$D_n=\sum_{i=1}^{n}(x_i-\bar{x})Z_i\Big/\sum_{l=1}^{n}Z_l \tag{2.9}$$
上式中小写的x_i表示随机变量X_i取的值x_i, $i=1, \cdots, n$, 在计算D_n的分布时, 它们是常数. 经计算,
$$\overline{D}_n^2=\operatorname{var}^*\left(\sum_i x_iV_i\right)=\frac{1}{(4n+1)n}\sum_i(x_i-\bar{x})^2. \tag{2.10}$$
故
$$\begin{aligned} P^*\{D_n/\overline{D}_n\leqslant y\}&=P^*\left\{\sum_{i=1}^{n}(x_i-\bar{x})Z_i\Big/\sum_{l=1}^{n}Z_l\leqslant y\overline{D}_n\right\}\\ &=P^*\{\sum\beta_{in}(y)(Z_i-2)\leqslant\rho_n(y)\}, \end{aligned} \tag{2.11}$$
其中
$$\beta_{in}(y)=\alpha_{in}(y)\Big/\left(\sum_j\alpha_{jn}^2(y)\right)^{\frac{1}{2}}$$
$$\alpha_{in}(y)=x_i-\bar{x}-y\overline{D}_n,$$
$$\rho_n(y)=-2\sum_{i=1}^{n}\beta_{in}(y).$$
若记$\rho_{jn}=(Z_j-2)\beta_{jn}$, $j=1, \cdots, n$, 则当$x_1, x_2, \cdots, x_n$给定时, 它是一串相互独立的随机变量. ρ_{jn}的条件密度为
$$f_{jn}(u)=\begin{cases} 2^4\left(\dfrac{u}{\beta_{jn}}+2\right)^3e^{-2\left(\frac{u}{\beta_{jn}}+2\right)}, & \dfrac{u}{\beta_{jn}}>-2\\ 0, & \text{其它}, \end{cases} \tag{2.12}$$
其特征函数为
$$v_{jn}(t)=e^{-2it\beta_{jn}}\left(1-i\frac{t\beta_{jn}}{2}\right)^{-4}, \quad j=1, 2, \cdots, n, \tag{2.13}$$
其数字特征为
$$E^*\rho_{jn}=0, \ \operatorname{var}^*\rho_{jn}=\beta_{jn}^2, \quad j=1, \cdots, n. \tag{2.14}$$
ρ_{in}的各阶半不变量为
$$\gamma_{1j}=0, \ \gamma_{\nu j}=\beta_{jn}^{\nu}\frac{4(\nu-1)!}{2^\nu}, \quad j=1, 2, \cdots, n, \ \nu\geqslant 2. \tag{2.15}$$

定理 1 设 D_n 由(2.8)给出，若 $E|X|^2<\infty$，则对于 $k\geqslant 3$，

$$\left|P^*\left\{\frac{D_n}{\bar{D}_n}\leqslant y\right\}-\Phi(\rho_n(y))-\sum_{\nu=1}^{k-2}\bar{Q}_{\nu,n}(\rho_n(y))\right|\leqslant\frac{C(k)\varepsilon_{nk}}{(1+|\rho_n(y)|)^k}\left(\sum_{j=1}^{n}|\beta_{jn}(y)|^k\right) \tag{2.16}$$

其中

$$\bar{Q}_{\nu,n}(y)=-\frac{1}{\sqrt{2\pi}}e^{-y^2/2}\sum{}' H_{\nu+2s-1}(y)\prod_{m=1}^{\nu}\frac{1}{k_m!}\left(\frac{4}{(m+2)2^{m+2}}\sum_{j=1}^{n}\beta_{jn}^{m+2}\right)^{k_m}, \tag{2.17}$$

上式中求和号 $\sum'$ 的意义与(2.2)中的相同. 式中 ε_{nk} 满足 $\varepsilon_{nk}\to 0$ a.e, 并且收敛是对 $y\in(-\infty, +\infty)$ 一致的.

证明 利用(2.1)以及引理 1 可知

$$\begin{aligned}&\left|P^*\left\{\frac{D_n}{\bar{D}_n}\leqslant y\right\}-\Phi(\rho_n(y))-\sum_{\nu=1}^{k-2}Q_{\nu,n}(\rho_n(y))n^{-\frac{\nu}{2}}\right|\\&\leqslant\frac{C(k)}{(1+|\rho_n(y)|)^k}\left\{B_n^{-k}\sum_{j=1}^{n}E^*|W_{nj}^{(\rho_n(y))}|^k+B_n^{-k-1}\sum_{j=1}^{n}E^*|Z_{nj}^{(\rho_n(y))}|^{k+1}\right.\\&\quad\left.+n^{k(k+1)/2}\left(\sup_{|t|\geqslant\delta_n}\frac{1}{n}\sum_{j=1}^{n}|\nu_j(t)|+\frac{1}{2n}\right)^n\right\},\end{aligned} \tag{2.18}$$

经计算可知 $B_n=1$, 利用(2.15), (2.1)可得

$$\lambda_{\nu,n}=n^{(\nu-2)/2}\sum_{j=1}^{n}\beta_{jn}^{\nu}\cdot\frac{4(\nu-1)!}{2^\nu},\quad \nu=3,\ 4,\ \cdots.$$

将上式代入(2.2)式，可得

$$Q_{\nu,n}(x)=\bar{Q}_{\nu,n}(x)\cdot n^{\nu/2}$$

上式说明(2.16)式和(2.18)式的左端是一样的.

现在来研究不等式(2.18)左边诸项的渐近性质. 首先由 β_{in} 之定义知

$$\beta_{in}(y)=\frac{X_i-\bar{X}-y\bar{D}_n}{\left(1+\frac{y^2}{4n+1}\right)^{\frac{1}{2}}(\sum(X_i-\bar{X})^2)^{\frac{1}{2}}},$$

故

$$\begin{aligned}\max_{i=1,\cdots,n}\sup_{y}|\beta_{in}(y)|&\leqslant\frac{\max\limits_{i=1,\cdots,n}|X_i|}{(\sum\limits_i(X_i-\bar{X})^2)^{\frac{1}{2}}}+\frac{|\bar{X}|}{(\sum(X_i-\bar{X})^2)^{\frac{1}{2}}}+\sup_y\left(\frac{\frac{y^2}{(4n+1)n}}{1+\frac{1}{4n+1}y^2}\right)^{\frac{1}{2}}\\&\leqslant\frac{\max\limits_{i=1,\cdots,n}|X_i|/\sqrt{n}}{\left(\frac{1}{n}\sum(X_i-\bar{X})^2\right)^{\frac{1}{2}}}+\frac{1}{\sqrt{n}}\frac{|\bar{X}|}{\left(\frac{1}{n}\sum(X_i-\bar{X})^2\right)^{\frac{1}{2}}}+\frac{1}{\sqrt{n}}.\end{aligned} \tag{2.19}$$

由于 X_i, $i=1,\cdots,n$ 为 iid 序列并且 $E|X|^2<\infty$, 由标准的极限定理(见[3], p. 12)可知

$$\lim_{n\to\infty}\frac{\max\limits_{i=1,\cdots,n}|X_i|}{\sqrt{n}}=0,\quad \text{a.e.}, \tag{2.20}$$

由(2.19), (2.20)可知

$$\lim_{n\to\infty}\max_{i=1,\cdots,n}\sup_y|\beta_{in}(y)|=0,\quad \text{a.e.}, \tag{2.21}$$

现在研究(2.18)右边花括弧内第一个和项的渐近性质.

$$\sum_{j=1}^{n} E^* |W_{nj}^{(\rho_n(y))}|^k = \sum_{j=1}^{n} |\beta_{jn}|^k E^* \left\{ |Z_j-2|^k I\left(|Z_j-2| > \frac{1+|\rho_n(y)|}{|\beta_{jn}|} \right) \right\}$$

$$\leqslant \sum_{j=1}^{n} |\beta_{jn}|^k E^* \left\{ Z_1-2|^k I\left(|Z_1-2| > \frac{1}{\max\limits_{j=1,\cdots,n} \sup\limits_{y} |\beta_{jn}(y)|} \right) \right\}$$

$$\triangleq \varepsilon_{nk} \left(\sum_{j=1}^{n} |\beta_{jn}|^k \right) \tag{2.22}$$

显然此处 ε_{nk} 只与 n, k 以及 X_1, X_2, $\cdots$,有关,并且

$$\varepsilon_{nk} \to 0, \quad \text{a.e.}. \tag{2.23}$$

现在再看(2.18)右边花括弧内第二个和项的渐近性质.

$$\sum_{j=1}^{n} E^* |Z_{nj}^{(\rho_n(y))}|^{k+1} \leqslant \sum_{j=1}^{n} |\beta_{jn}|^{k+1} E^* \{ |Z_j-2|^{k+1} I(|\beta_{jn}(Z_j-2)| \leqslant 1+|\rho_n(y)|) \}$$

$$\leqslant \max_{j=1,\cdots,n} \sup_{y} |\beta_{jn}| \sum_{j=1}^{n} |\beta_{jn}|^k \cdot E^* |Z_1-2|^{k+1}$$

$$= \varepsilon_{nk} \left(\sum_{j=1}^{n} |\beta_{jn}|^k \right) \tag{2.24}$$

其中 ε_{nk} 满足(2.23),并且只与 n, k 以及 X_1, X_2, $\cdots$ 有关.

现在考察(2.18)右边花括弧内第三项的性质. 由 β_{jn} 定义立即可得

$$\sum_{j=1}^{n} \beta_{jn}^2 = 1. \tag{2.25}$$

利用上式可知

$$\sum_{j=1}^{n} E^* |Y_{nj}|^3 = \sum_{j=1}^{n} |\beta_{jn}|^3 E(|Z_j-2|^3 I(|\beta_{jn}(Z_j-2)| \leqslant 1))$$

$$\leqslant \max_{j=1,\cdots,n} \sup_{y} |\beta_{jn}| \sum |\beta_{jn}|^2 \to 0, \quad \text{a.e.}$$

故

$$\delta_n = \frac{1}{12} \left(\sum_{j=1}^{n} E^* |Y_{nj}|^3 \right)^{-1} \to \infty, \quad \text{a.e.},$$

上式的收敛是对 $y \in (-\infty, +\infty)$ 一致地成立的. 由(2.13)知

$$|v_j(t)| = \frac{1}{\left|1 - i\frac{t}{2}\beta_{jn}\right|^4} = \left(\frac{1}{1+\frac{t^2\beta_{jn}^2}{4}} \right)^2,$$

则当 n 充分大时

$$\sup_{|t| \geqslant \delta_n} \frac{1}{n} \sum_{j=1}^{n} |v_j(t)| \leqslant \frac{1}{n} \sum \frac{1}{1+\delta_n^{\frac{1}{2}} \beta_{jn}^2}$$

$$\leqslant 1 - \frac{1}{n} \sum \beta_{jn}^2 \delta_n^{\frac{1}{2}} + \frac{1}{n} \sum |\beta_{jn}|^4 \delta_n. \tag{2.26}$$

现将 $n^{k(k+1)/2} \left(\sup\limits_{|t| \geqslant \delta n} \frac{1}{n} \sum\limits_{j=1}^{n} |v_j(t)| + \frac{1}{2n} \right)^n$ 写成 $\varepsilon_{nk} \left(\sum\limits_{j=1}^{n} |\beta_{in}|^k \right)$ 的形式, 我们将证明 $\varepsilon_{n,k}$ 为某一 a.e. 趋于零的随机变量. 取

$$\varepsilon_{nk} = \left(\sum_{j=1}^{n} |\beta_{jn}|^k \right)^{-1} n^{k(k+1)/2} \left(\sup_{|t| \geqslant \delta_n} \frac{1}{n} \sum_{j=1}^{n} |v_j(t)| + \frac{1}{2n} \right)^n. \tag{2.27}$$

由

$$\left(\frac{\sum |\beta_{jn}|^k}{n} \right)^{\frac{1}{k}} \geqslant \left(\frac{\sum |\beta_{jn}|^2}{n} \right)^{\frac{1}{2}} = \left(\frac{1}{n} \right)^{\frac{1}{2}}$$

可知

$$(\sum |\beta_{jn}|^k)^{-1} \leqslant n^{\frac{k}{2}-1}.$$

将上式代入(2.27)，再利用(2.26)，可得

$$\varepsilon_{nk}\leqslant n^{\frac{k(k+1)}{2}+\frac{k}{2}-1}\left(\sup_{|t|\geqslant\delta_n}\frac{1}{n}\sum_{j=1}^{n}|v_j(t)|+\frac{1}{2n}\right)^n$$

$$\leqslant n^{\frac{k(k+1)}{2}+\frac{k}{2}-1}\left(1-\frac{1}{n}\delta_n^{\frac{1}{2}}+\frac{1}{n}+\frac{1}{n}\sum|\beta_{jn}|^4\delta_n\right)^n$$

现在分析上式右端圆括弧内诸项之渐近特性，首先当 n 充分大时

$$\sum|\beta_{jn}|^4\delta_n=\frac{\sum|\beta_{jn}|^4}{12\sum_{j=1}^{n}|\beta_{jn}|^3E(|Z_j-2|^3I(|\beta_{jn}(Z_j-2)|\leqslant1))}$$

$$\leqslant\max_{1\leqslant j\leqslant n}|\beta_{jn}|\cdot\frac{1}{12E(|Z_1-2|^3I(|\beta_{jn}(Z_1-2)|\leqslant1))}\to0,\quad \text{a.e.}.$$

同时，利用通常的概率论工具可以证明

$$\lim_{n\to\infty}\sup\sup_{y}\frac{1}{\sqrt{n}}\delta_n<\infty,$$

$$\lim_{n\to\infty}\inf\inf_{y}\frac{1}{\sqrt{n}}\delta_n>0.$$

由此可知存在 $c>0$，使对几乎所有的样本序列，当 n 充分大时，

$$\left(1-\frac{1}{n}\delta_n^{\frac{1}{2}}+\frac{1}{n}+\frac{1}{n}\sum|\beta_{jn}|^4\delta_n\right)^n\leqslant e^{-n^{\frac{1}{4}}c}.$$

由上式之成立，可知 $\varepsilon_{nk}\to0$, a.e.. 这样在(2.18)右端花括弧内的三项都具有 $\varepsilon_{nk}\left(\sum_{j=1}^{n}|\beta_{jn}|^k\right)$ 的形式，从而定理得证.

在定理 1 中若取 $k=3$，便得到不等式

$$\left|P^*\left\{\frac{D_n}{\overline{D}_n}\leqslant y\right\}-\Phi(\rho_n(y))+\varphi(\rho_n(y))(\rho_n^2(y)-1)\left(\frac{1}{6}\sum_{j=1}^{n}\beta_{jn}^3\right)\right|$$

$$\leqslant\frac{C(3)\varepsilon_{n3}}{(1+|\rho_n(y)|)^3}\left(\sum_{j=1}^{n}|\beta_{jn}(y)|^3\right),\tag{2.28}$$

式中 $\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$.

引理 2 存在常数 c，使得下面两式成立：

$$\sup_{y}|\Phi(\rho_n(y))-\Phi(y)|\leqslant\frac{c}{n},\tag{2.29}$$

$$\sup_{y}|\varphi(\rho_n(y))(\rho_n^2(y)-1)-\varphi(y)(y^2-1)|\leqslant\frac{c}{n},\tag{2.30}$$

证明 由于两式之证明是类似的，我们只给出(2.29)的证明. 经计算知

$$\rho_n(y)=\frac{2\sqrt{n}\,y}{\sqrt{4n+1+y^2}}.$$

由于标准正态密度函数之对称性，我们只需对于 $y\geqslant0$ 证明(2.29)即可. 当 $y\leqslant\sqrt{4n-1}$ 时，$\frac{y}{\sqrt{2}}\leqslant\rho_n(y)\leqslant y$. 利用中值定理可知

$$y-\rho_n(y)\leqslant y(1+y^2)/(8n).$$

故 $$\sup_{0\leqslant y\leqslant\sqrt{4n-1}}|\Phi(y)-\Phi(\rho_n(y))|\leqslant\frac{1}{n}\sup_{y}\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{4}}y(1+y^2)=\frac{c}{n},$$

其中 c 为某一常数. 当 $y>\sqrt{4n-1}$ 时, $\min\{y^2, \rho^2(y)\}\geqslant n$, $n\geqslant 2$, 故对于 $y>\sqrt{4n-1}$,

$$|\Phi(\rho_n(y))-\Phi(y)|=\frac{1}{\sqrt{2\pi}}\int_{\rho_n(y)}^{y}e^{-\frac{1}{2}x^2}dx\leqslant\frac{1}{n}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}x^2e^{-\frac{1}{2}x^2}dx=\frac{1}{n}.$$

综合所得的不等式, 可知(2.29)成立.

利用强大数定律不难证得, 当 $E|X|^3<\infty$ 时

$$\limsup_{n\to\infty}\sqrt{n}\sup_y\sum_{j=1}^{n}|(\beta_{jn}(y)|^3<\infty\quad \text{a.e.},\tag{2.31}$$

$$\lim_{n\to\infty}\sqrt{n}\sup_y\left|\varphi(y)(y^2-1)\left(\frac{1}{6}\frac{\mu_3}{\sqrt{n}\,\sigma^3}-\frac{1}{6}\sum_{j=1}^{n}\beta_{jn}^3(y)\right)\right|=0,\quad \text{a.e.}.\tag{2.32}$$

由(2.28), (2.31), (2.32)以及引理 2 可知, 当 $E|X|^3<\infty$ 时, 对几乎所有的 $x_1, x_2, \cdots$,

$$\lim_{n\to\infty}\sup_y\sqrt{n}\left|P^*\left\{\frac{D_n}{\overline{D}_n}\leqslant y\right\}-\Phi(y)+\varphi(y)(y^2-1)\frac{\mu_3}{6\sqrt{n}\,\sigma^3}\right|=0.\tag{2.33}$$

另一方面由[2]知, 若 X 的分布不是格子点分布并且 $E|X|^3<\infty$, 则

$$\lim_{n\to\infty}\sup_y\sqrt{n}\left|P\{\sqrt{n}(\overline{X}-\mu)/\sigma\leqslant y\}-\Phi(y)+\varphi(y)(y^2-1)\frac{\mu_3}{6\sqrt{n}\,\sigma^3}\right|=0.\tag{2.34}$$

将(2.33), (2.34)综合, 可得

定理 2 设 $\{X_i\}$ 为 iid 随机变量, X_i 的分布为非格子点分布, 并且 $E|X_1|^3<\infty$, 则对几乎所有的样本序列 $x_1, x_2, \cdots$,

$$\lim_n\sqrt{n}\sup_y|P\{\sqrt{n}(\overline{X}-\mu)/\sigma\leqslant y\}-P^*\{D_n/\overline{D}_n\leqslant y\}|=0.\tag{2.35}$$

在定理 2 中, 要求 X 的分布具有三阶矩. 当这个条件不满足时, 由于 $P\{\sqrt{n}(\overline{X}-\mu)/\sigma\leqslant y\}$ 不具有渐近展开, 因此不可能得到(2.35)那样的结果. 现在讨论当 $EX^2<\infty$ 时的渐近结果.

引理 3 设 $X_1, \cdots, X_n$ 相互独立, $EX_j=0$, $EX_j^2=\sigma_j^2$. 记

$$B_n^2=\sum\sigma_j^2>0,\ L_{k,n}=B_n^{-k}\sum_{j=1}^{n}E|X_j|^k<\infty,$$

其中 $k\geqslant 3$ 为整数. 则

$$|F_n(y)-\Phi(y)|\leqslant C(k)(1+|y|)^{-k}(L_{3,n}+L_{k,n}),\tag{2.36}$$

其中 $C(k)$是只与 k 有关的常数.

引自[5]之定理 5.

将引理 3 之结论应用于随机变量序列 $\beta_{jn}(y)(Z_j-2)$, $j=1, \cdots, n$(见(2.11)), 便得

$$|P^*\{D_n/\overline{D}_n\leqslant y\}-\Phi(\rho_n(y))|\leqslant C(k)\frac{L_{3,n}}{(1+|\rho_n(y)|)^3}\tag{2.37}$$

其中
$$L_{3,n}=\sum_{j=1}^{n}\beta_{jn}(y)^3,$$

而 $C(k)$是只与 k 有关的常数. 由(2.21), 不难证得

$$\lim_{n\to\infty}\sup_y|L_{3,n}|=0,\quad \text{a.e.},$$

结合(2.37), 可知对几乎所有的 $x_1, x_2, \cdots$,

$$\lim_{n\to\infty}\sup_y|P^*\{D_n/\overline{D}_n\leqslant y\}-\Phi(\rho_n(y))|=0.$$

再利用引理 2, 可得

定理3 设 $X_1, \cdots, X_n \sim \text{iid } F$, $EX_1^2<\infty$,则对几乎所有的样本 $x_1, x_2, \cdots$,

$$\sup_y |P^*\{D_n/\bar{D}_n \leqslant y\} - P\{\sqrt{n}(\bar{X}-\mu)/\sigma \leqslant y\}| \to 0. \tag{2.38}$$

定理2和定理3为统计量 D_n 的应用提供了理论根据. 为了进一步推广应用,在此我们提供 D_n 的一个计算方法. 计算 D_n 的关键是统计量 $V_1, \cdots, V_n$ 的模拟. 取 $Z_1, Z_2, \cdots Z_{4n-1}$ 为 iid $U(0, 1)$ 随机序列,即 Z_i 为$(0, 1)$上均匀分布的随机变量序列,取 $Z_{(1)}, \cdots, Z_{(4n-1)}$ 为它们的次序统计量. 取 $V_1=Z_{(4)}$, $V_2=Z_{(8)}-Z_{(4)}$, $\cdots$, $V_{n-1}=Z_{(4(n-1))}-Z_{(4(n-2))}$, $V_n=1-Z_{(4(n-1))}$, 则$(V_1, \cdots, V_n)\sim D(4, \cdots, 4)$. 由此看出$(V_1, \cdots, V_n)$的模拟是非常方便的.

参 考 文 献

[1] Efron, B., Bootstrap method: arother look at the Jackknife. *Ann. Statist.* **7** (1979), **1—26.**
[2] Petrov, V. V., Sums of indopendent random variables *Springer verlag.* 1975.
[3] Serfling, R., Approximation theorems of mathematical statistics, *John Wiley*, 1980.
[4] Wilks, S., Mathematical statistics. *John Wiley*, 1962-
[5] 白志东、赵林城,独立随机变量之和的分布函数的渐近展开,中国科学(A 辑)8, (1985).
[6] 郑忠国,随机加权法. 应用数学学报, Vol. 10, No. 2 (1987).

THE EDGEWORTH EXPANSION FOR THE RANDOM WEIGHTING METHOD

TU DONGSHENG (Institute of Systems Science, Academia Sinica)

ZHENG ZHONGGUO (Beijing University)

The random weighting method, which differs from bootstrap, is a new approch to estimate of the distribution of pivotal statistics. In this paper, we develop an Edgeworth expansion for the random weighting distribution of sample mean. Let $F_n(x)$ be distribution function of $(\bar{X}-\mu)/\sigma$ and $F_n^*(x)$ distribution function of $\sum(X_i-\bar{X})V_i/\sigma^*(\sum(X_i-\bar{X})V_i)$, where the distribution of $(v_1, v_2, \cdots, v_n)$ is a Dirichlet $D(4, 4, \cdots, 4)$ distribution and σ^{*2} the variance of $(\sum(X_i-\bar{X})v_i)$ given $X_1, X_2, \cdots, X_n$. Using the expansion we have: If $E|X|^3<\infty$, then

$$\sqrt{n}\sup_x |F_n^*(x) - F_n(x)| \to 0 \quad \text{(a.e.)}$$