系 统 科 学 与 数 学
J. Sys. Sci. & Math. Scis.
29(5) (2009, 5), 693–705

序集抽样中 M 估计分布的随机加权逼近*

吴 耀 华　　刘 驰 宇

(中国科学技术大学统计与金融系， 合肥 230026)

摘要　序集抽样是一种适用于准确测量花费太高而排序费用可以忽略不记时的一种抽样方法. 讨论了序集抽样下的对于一般分布族 M 估计的相合性和渐近正态性并且通过随机加权的方法来估计 M 估计的分布.

关键词　序集抽样， M 估计，随机加权，渐近正态性.

MR(2000) 主题分类号　62E17, 62G06

1　引　言

寻找有效的抽样方法一直是统计工作者关心的问题，尤其当准确测量花费太高或者太花时间的时候. 上世纪 50 年代初， McIntyre 在 [1] 提出了序集抽样 (RSS) 的方法． RSS 的基础，是假定在一个无穷总体中，对每一个体进行准确测量的费用很高或时间很长，而对他们或其相关的量进行某种排序的费用可以忽略不计的情形.

关于 RSS 的性质前人已作了不少研究工作. 关于总体均值和方差的矩估计以及它们相对于简单抽样的相对准确性可以见文 [1–5]. [1–5] 研究了把经验分布作为总体分布的估计并讨论了在 RSS 下的 Kolmogrov-Smirnov 检验． [6] 讨论了在 RSS 下的 Mann-Whitney-Wilcoxon 检验． [7] 考虑了相应的符号检验． [8] 讨论了 RSS 下的 U 统计量. 关于参数过程，包括 RSS 的最优方案设计可以看 [9–11]. 在 [12] 中有关于 RSS 抽样各个性质的讨论． [13] 中讨论了序集抽样下关于对称分布族的 M 估计的相合性和渐近正态性.

本文中我们将问题扩大到一般分布族，证明了序集抽样下相应的 M 估计的性质，得到在同样的样本量下，序集抽样比普通抽样具有更高的效率. 并且证明了 M 估计分布的随机加权逼近的有效性，模拟结果表明序集抽样下进行随机加权能较好逼近 M 估计的分布.

现在我们来描述 RSS 抽样过程：假设需要从关心的总体 X(分布为 F) 中抽取一个样本量为 n 的样本. 首先我们从总体中随机抽出 n 组，每组包含 k 个元素，在每组内按照某种机制对 k 个元素进行排序 (此时并不需要对每个元素进行精确的测量)，我们只精确测量排在事先指定位置上 X 值，而其他位置上的元素则抛弃不用. 假设我们总共在 $n_r(r = 1, 2, \cdots, k$, 且满足 $\sum_{r=1}^{k} n_r = n)$ 组中测量的是排在第 r 个位置上的元素，记这 n_r 个测量值为 $X_{[r]i}(i = 1, 2, \cdots, n_r)$.

* 国家自然科学基金 (10871188) 和中国科学院知识创新工程 (KJCX3-SYW-S02) 资助课题.
收稿日期：2006-11-07, 收到修改稿日期：2007-09-24.

这样，我们得到 n 个独立观测值：

$$\begin{array}{c} X_{[1]1},\ X_{[1]2},\ \cdots,\ X_{[1]n_1}; \\ X_{[2]1},\ X_{[2]2},\ \cdots,\ X_{[2]n_2}; \\ \cdots\ \cdots\ \cdots\cdots \\ X_{[k]1},\ X_{[k]2},\ \cdots,\ X_{[k]n_k}. \end{array}$$

这 n 个观测值即为 RSS 样本. 其中第 $r(r=1,2,\cdots,k)$ 行的 n_r 个观测值服从相同的分布 $F_{[r]}$. 当 $F_{[r]}=F_{(r)}$(第 r 个次序统计量的分布), 这种排序机制称为是完美的 (Perfect). 当 $n_1=n_2=\cdots=n_k$ 时称为均衡 RSS 抽样；当

$$F(x)=\frac{1}{k}\sum_{r=1}^{k}F_{[r]}$$

时，称为相合 RSS 抽样. 完美的抽样是一个相合的抽样，此外还有一些抽样方法也是相合的，具体可见文 [12].

我们现在考虑均衡 RSS 抽样中位置参数的 M 估计. 设总体 X 满足

$$E\rho(X-\mu_0)=\min_{\mu}E\rho(X-\mu), \tag{1}$$

其中 ρ 是 R^1 上一个非单调的凸函数. 考虑 μ_0 的 M 估计 $\widehat{\mu}_n$, $\widehat{\mu}_n$ 为满足

$$\sum_{r=1}^{k}\sum_{i=1}^{m}\rho(X_{[r]i}-\widehat{\mu}_n)=\inf_{\mu\in R}\sum_{r=1}^{k}\sum_{i=1}^{m}\rho(X_{[r]i}-\mu) \tag{2}$$

的一个解.

为讨论 $\widehat{\mu}_n$ 的性质，我们作如下假定

1) ρ 为凸函数，Ψ_- 和 Ψ_+ 分别为 ρ 的左右导数, 函数 Ψ 满足 $\Psi_-(u)\le\Psi(u)\le\Psi_+(u)$, 对任意 $u\in R$;

2) 存在正常数 c 和 h_0, 使得对任给的 $h\in(0,h_0)$ 和所有的 u 均有

$$\Psi(u+h)-\Psi(u)\le c;$$

3) $G(u)=E\Psi(X-\mu_0+u)$, 在 $u=0$ 附近存在有限， $G(0)=0$, 且 $G(u)$ 在零点的一个邻域 $(-\Delta,\Delta)$ 内有有界导数，且 $\lambda\equiv G'(0)>0$;

4) $0<E\Psi^2(X-\mu_0)=\sigma^2<\infty$, 且

$$\lim_{\mu\to\mu_0}E[\Psi(X-\mu)-\Psi(X-\mu_0)]^2=0.$$

为了估计的 $\widehat{\mu}$ 分布，我们取一列随机变量 $\{\omega_i\}$, 满足如下假定

5) $\omega_1,\omega_2,\cdots,\omega_m$ 为 iid, $P(\omega_1>0)=1, E\omega_1=1, E\omega_1^2=\tau\ge1$, 且序列 $\{\omega_i\}$ 与 $\{X_{[r]i}\}$ 独立.

令 μ^* 为加权的 M 估计，定义为

$$\sum_{r=1}^{k}\sum_{i=1}^{m}\omega_i\rho(X_{[r]i}-\mu^*)=\inf_{\mu\in R}\sum_{r=1}^{k}\sum_{i=1}^{m}\omega_i\rho(X_{[r]i}-\mu), \tag{3}$$

我们将用给定 $X_{[1]1}, X_{[1]2}, \cdots, X_{[k]m}$ 下 $\sqrt{n}(\mu^* - \widehat{\mu}_n)$ 的条件分布来估计 $\sqrt{n}(\widehat{\mu}_n - \mu_0)$ 的分布. 在这篇文章中, 我们用符号 $\mathcal{L}^*, P^*, E^*, \mathrm{Var}^*$ 来表示在给定 $X_{[1]1}, X_{[1]2}, \cdots, X_{[k]m}$ 下的条件概率运算.

定理 1.1 如果均衡 RSS 抽样是相合的, k 固定. 在假设 1)– 5) 下, μ^* 由 (3) 定义, 则沿着几乎所有的样本序列, 当给定 $X_{[1]1}, X_{[1]2}, \cdots, X_{[k]m}$ 时, 我们有

$$\sqrt{n}(\mu^* - \mu_0) = \frac{1}{\lambda\sqrt{n}} \sum_{r=1}^{k} \sum_{i=1}^{m} \omega_i \Psi(X_{[r]i} - \mu_0) + o_p(1), \quad \text{当 } n \to \infty, \tag{4}$$

特别地, 当在假设 5) 中取 $\tau = 1$ 时, 有

$$\sqrt{n}(\widehat{\mu}_n - \mu_0) = \frac{1}{\lambda\sqrt{n}} \sum_{r=1}^{k} \sum_{i=1}^{m} \Psi(X_{[r]i} - \mu_0) + o_p(1) \xrightarrow{\mathcal{L}} N\Big(0, \frac{\sigma_{RSS}^2}{\lambda^2}\Big), \quad \text{当 } n \to \infty, \tag{5}$$

其中

$$\sigma_{RSS}^2 = \Big[\sigma^2 - \frac{1}{k} \sum_{r=1}^{k} \Big(\int \Psi(x - \mu_0) \mathrm{d}F_{[r]}(x) \Big)^2 \Big]. \tag{6}$$

定理 1.2 如果均衡 RSS 抽样是相合的, k 固定. 当假设 1)– 5) 成立, 在假定 5) 中取 $\tau = 2$, 且 $E|\Psi(X)|^{2+\delta} < \infty$. 令 $\widehat{\mu}_n$ 和 μ^* 分别如 (2) 和 (3) 定义, 则沿着几乎所有样本序列

$$\mathcal{L}^*(\sqrt{n}(\mu^* - \widehat{\mu}_n)) \to N\Big(0, \frac{\sigma_{RSS}^2}{\lambda^2}\Big), \quad \text{当 } n \to \infty. \tag{7}$$

本文的结构是这样的, 第 2 节给出了定理的证明, 其中需要的引理及其证明放在第 3 节中给出, 第 4 节给出了关于 RSS 抽样 M 估计的相对效, 并在第五节进行了数值模拟.

2　定理的证明

为证明叙述方便, 定义

$$e_{[r]i} = X_{[r]i} - \mu_0, \quad r = 1, 2, \cdots, k, \ i = 1, 2, \cdots, m. \tag{8}$$

而 $\mu(n)$ 满足

$$\sum_{r=1}^{k} \sum_{i=1}^{m} \rho\Big(e_{[r]i} - \frac{\mu(n)}{\sqrt{n}}\Big) = \min_{\mu} \sum_{r=1}^{k} \sum_{i=1}^{m} \rho\Big(e_{[r]i} - \frac{\mu}{\sqrt{n}}\Big), \tag{9}$$

则我们有

$$\mu(n) = \sqrt{n}(\widehat{\mu}_n - \mu_0). \tag{10}$$

定理 1.1 的证明　不失一般性, 可假设在模型中 $\mu_0 = 0$. 定义

$$\overline{\mu}^*(n) = \frac{1}{\lambda\sqrt{n}} \sum_{i=1}^{m} \sum_{r=1}^{k} \omega_i \Psi(e_{[r]i}).$$

首先证明对几乎所有的样本序列, 给定 $X_{[1]1}, X_{[1]2}, \cdots, X_{[k]m}$, 当 $n \to \infty$ 时, 对任意 $b > 0$ 有

$$|\overline{\mu}^*(n)| = o_p(n^b), \tag{11}$$

记

$$\overline{\mu}(n)=\frac{1}{\lambda\sqrt{n}}\sum_{i=1}^{m}\sum_{r=1}^{k}\Psi(e_{[r]i}),$$

因为

$$E\Big|\sum_{r=1}^{k}\Psi(e_{[r]i})\Big|^2\le C_rE\sum_{r=1}^{k}\Psi^2(e_{[r]i})\le M, \tag{12}$$

则

$$\sum_{n=1}^{\infty}\frac{E\Big|\sum\limits_{r=1}^{k}\Psi^2(e_{[r]i})\Big|^2}{n^{b+\frac{1}{2}}}\le\sum_{n=1}^{\infty}\frac{M}{n^{b+\frac{1}{2}}}<\infty. \tag{13}$$

在引理 3.2 中令 $a_n=n^{b+\frac{1}{2}},\ p=2,\ X_k=\sum\limits_{r=1}^{\infty}\Psi(e_{[r]k})$ 则有

$$\frac{1}{n^{b+\frac{1}{2}}}\sum_{i=1}^{m}\sum_{r=1}^{k}\Psi(e_{[r]i})=o(1),\quad \text{a.s.} \tag{14}$$

因此对任给的 $\varepsilon>0$,

$$\begin{aligned}&P^*\Big\{\Big|\frac{1}{n^{b+\frac{1}{2}}}\sum_{i=1}^{m}\sum_{r=1}^{k}\omega_i\Psi(e_{[r]i})\Big|>\varepsilon\Big\}\\&\le P^*\Big\{\Big|\frac{1}{n^{b+\frac{1}{2}}}\sum_{i=1}^{m}\sum_{r=1}^{k}(\omega_i-1)\Psi(e_{[r]i})\Big|>\frac{\varepsilon}{2}\Big\}\\&\le\frac{4}{\varepsilon^2n^{2b+1}}\sum_{i=1}^{m}\sum_{r=1}^{k}\tau\Psi^2(e_{[r]i}),\end{aligned} \tag{15}$$

在引理 3.2 中，令 $a_n=n^{2b+1},\ p=1,\ X_k=\sum\limits_{r=1}^{\infty}\Psi^2(e_{[r]k})$, 则有

$$\Big|\frac{1}{n^{2b+1}}\sum_{i=1}^{m}\sum_{r=1}^{k}\Psi^2(e_{[r]i})\Big|=o(1),\quad \text{a.s.} \tag{16}$$

由 (15), (16) 知，对几乎所有的样本序列，给定 $X_{[1]1},X_{[1]2},\cdots,X_{[k]m}$, 当 $n\to\infty$ 时

$$\frac{1}{n^{b+\frac{1}{2}}}\sum_{i=1}^{m}\sum_{r=1}^{k}\omega_i\Psi(e_{[r]i})=o_p(1), \tag{17}$$

由此即得 (11). 即任给 $\varepsilon>0$, 对几乎所有的样本序列，当 n 充分大时，

$$P^*\{|\overline{\mu}^*(n)|>n^b\}<\frac{\varepsilon}{2}. \tag{18}$$

对任意给定的 $\delta>0$, 取 $0<b<\frac{1}{20}$, 由引理 3.4 知，当 k 固定时，对几乎所有的样本序列，有

$$I(|\overline{\mu}^*(n)|\le n^b)\sup_{|\gamma-\overline{\mu}^*(n)|=\delta}\left|\sum_{r=1}^{k}\sum_{i=1}^{m}\omega_i\left\{\rho\Big(e_{[r]i}-\frac{\gamma}{\sqrt{n}}\Big)-\rho\Big(e_{[r]i}-\frac{\overline{\mu}(n)}{\sqrt{n}}\Big)\right\}-\frac{1}{2}\lambda(\gamma-\overline{\mu}(n))^2\right|=o_p(1),\tag{19}$$

由 (18), (19) 知，当 n 充分大时，

$$P^*\left\{\inf_{|\gamma-\overline{\mu}(n)|=\delta}\sum_{r=1}^{k}\sum_{i=1}^{m}\omega_i\rho\Big(X_{[r]i}-\frac{\gamma}{\sqrt{n}}\Big)\ge\sum_{r=1}^{k}\sum_{i=1}^{m}\omega_i\rho\Big(X_{[r]i}-\frac{\overline{\mu}(n)}{\sqrt{n}}\Big)+\frac{1}{4}\lambda\delta^2\right\}\ge1-\varepsilon.\tag{20}$$

由 (20) 及 ρ 的凸性，当 n 充分大时，

$$P^*\{|\mu^*-\overline{\mu}(n)|<\delta\}\ge1-\varepsilon,\quad\text{a.s.}\tag{21}$$

因为 ε 和 δ 都是任意的，故

$$\mu^*(n)-\overline{\mu}(n)\to0,\quad\text{a.s.}\quad\text{当 }n\to\infty.\tag{22}$$

由此即得 (4), 定理证毕.

定理 1.2 的证明　不失一般性，我们假定 $0<\delta<1$. 令

$$Z_n=\frac{1}{\sqrt{n}}\sum_{i=1}^{m}\sum_{r=1}^{k}(\omega_i-1)\,\Psi(e_{[r]i}),\tag{23}$$

$$(\mu^*(n)-\widehat{\mu}_n)=\frac{1}{\lambda}(Z_n+\zeta_n),\tag{24}$$

由定理 1.1, 我们有

$$E^*(|\zeta_n|\wedge1)\to0,\quad\text{a.s.}\quad\text{当 }n\to\infty,\tag{25}$$

其中 $a\wedge b=\min(a,b)$.

令

$$\eta_i=(\omega_i-1)\sum_{r=1}^{k}\frac{1}{\sqrt{n}}\Psi(e_{[r]i}),$$

则有

$$Z_n=\sum_{i=1}^{m}\eta_i,\quad E^*\eta_i=0,$$

对给定的 k, 有

$$\begin{aligned}E\Big(\sum_{r=1}^{k}\Psi(e_{[r]i})\Big)^2&=\mathrm{Var}\Big(\sum_{r=1}^{k}\Psi(e_{[r]i})\Big)+\Big(E\sum_{r=1}^{k}\Psi(e_{[r]i})\Big)^2\\&=\sum_{r=1}^{k}\mathrm{Var}(\Psi(e_{[r]i}))\\&=k\sigma_{RSS}^2\le k\sigma^2<\infty,\end{aligned}\tag{26}$$

由强大数律

$$\mathrm{Var}^*(Z_n)=\frac{1}{n}\sum_{i=1}^{m}\left(\sum_{r=1}^{k}\Psi(e_{[r]i})\right)^2\to\sigma^2_{RSS}\quad\text{a.s.},\tag{27}$$

则我们希望能够证明

$$\mathcal{L}^*(Z_n)\to N(0,\sigma^2_{RSS}),\quad\text{a.s.}\quad\text{当 } n\to\infty.\tag{28}$$

因此要证

$$(\mathrm{Var}^*(Z_n))^{-1}\sum_{i=1}^{m}E^*\eta_i^2I(|\eta_i|\geq\varepsilon[\mathrm{Var}^*(Z_n)]^{\frac{1}{2}})\to 0\quad\text{a.s.}\tag{29}$$

对任给的 $\varepsilon>0$, 令

$$L_n(\varepsilon)\equiv\sum_{r=1}^{k}\sum_{i=1}^{m}\frac{1}{n}\Psi^2(e_{[r]i})E^*\left((\omega_i-1)^2I(\Big|(\omega_i-1)\sum_{r=1}^{k}\frac{1}{\sqrt{n}}\Psi(e_{[r]i})\Big|\geq\varepsilon)\right),\tag{30}$$

令

$$T_m=\max_{1\leq i\leq m}\left|\sum_{r=1}^{k}\frac{1}{\sqrt{n}}\Psi(e_{[r]i})\right|,\tag{31}$$

则

$$T_m\leq\sum_{r=1}^{k}\frac{1}{\sqrt{n}}\max_{1\leq i\leq m}|\Psi(e_{[r]i})|\to 0,\quad\text{a.s.}\quad\text{当 } m\to\infty.\tag{32}$$

由 (31), (32) 知

$$L_n(\varepsilon)\leq\sum_{r=1}^{k}\sum_{i=1}^{m}\frac{1}{n}\Psi^2(e_{[r]i})E^*(\omega_i-1)^2I(T_m|(\omega_i-1)|\geq\varepsilon)\to 0,\quad\text{a.s.}\tag{33}$$

由中心极限定理, (28) 得证. 定理 1.2 证毕.

3 引 理

本节中将给出上节定理证明中所需要的引理及证明. 考虑区间 $D=\{\mu\in R:|\mu\leq n^b\}$ 其中 $0<b<\frac{1}{20}$ 为一常数. 现将 D 分割成一列互不相交的小区间 $D_1,D_2,\cdots,D_N$, 使得每个小区间的长度不大于 $\nu=n^{-2b}$. 由此可选取分割, 使得

$$N\leq cn^{3b},$$

对 $a\in R$ 和 $\mu\in D$, 定义

$$R^*_{ri}(a,\mu)=\int_0^{-\frac{\mu}{\sqrt{n}}}\left\{a[\Psi(e_{[r]i}+u)-\Psi(e_{[r]i})]-G(u)\right\}\mathrm{d}u,\tag{34}$$

$$A_{ri}(\varepsilon,\gamma)=\left\{\sup_{|\alpha_i-\beta_i|\leq\gamma}\left|\sum_{r=1}^{k}\sum_{i=1}^{m}(R_{ri}(1,\beta_i)-R_{ri}(1,\alpha_i))\right|\geq\varepsilon\right\},\tag{35}$$

同时记 $X=(X_{[1]1},X_{[1]2},\cdots,X_{[k]m})$.

引理 3.1　设 $\xi_1,\xi_2,\cdots,\xi_n$ 为相互独立的随机变量，使得 $|\xi_i|\le b<\infty$ 且 $E\xi_i=0,\ i=1,2,\cdots,n$. 记 $\nu=E(\xi_1^2)+\cdots+E(\xi_n^2)$, 则对任给的 $\varepsilon>0$, 有

$$P\Big\{\sum_{i=1}^{n}\xi_i\le\varepsilon\Big\}\le\exp\Big(-\frac{\varepsilon^2}{2\nu+2b\varepsilon}\Big). \tag{36}$$

证　见文 [14].

引理 3.2　设 X_n 是独立 r.v. 序列，$EX_n=0$, 正数序列 $a_n\uparrow\infty$, 且对某个 $p\ge1$

$$\sum_{n=1}^{\infty}\frac{E|X_n|^p}{a_n^p}<\infty, \tag{37}$$

那么

$$\frac{1}{a_n}\sum_{k=1}^{n}X_k\to0\quad\text{a.s.}$$

证　见文 [15].

引理 3.3　在定理 1.1 的条件下，任给 $\alpha_i,\ \beta_i\in D$, 且 $|\alpha_i-\beta_i|\le2\gamma$, 记

$$\zeta_{ri}=\bigg(\int_{\frac{\alpha_i}{\sqrt{n}}}^{\frac{\beta_i}{\sqrt{n}}}(\Psi(e_{[r]i}+u)-\Psi(e_{[r]i}))\,\mathrm{d}u\bigg)^2,\quad i=1,2,\cdots,m,\quad r=1,2,\cdots,k, \tag{38}$$

而对任给的 $\varepsilon>0$, 存在与 $\{\alpha_i\},\{\beta_i\}$ 无关的常数 c_i 和 n_1, 当 $n>n_1$ 时，

$$P\Big\{\sum_{r=1}^{k}\sum_{i=1}^{m}\zeta_{ri}\ge\varepsilon n^{3b-\frac12}\Big\}\le2\exp(-c_1n^{3b+\frac12}). \tag{39}$$

证　由假设 2), 3) 知，对充分大的 n 有

$$\begin{aligned}
&\Big|\zeta_{ri}-E\zeta_{ri}\Big|\le c\Big|\frac{1}{\sqrt{n}}(\alpha_i-\beta_i)\Big|^2\le\frac{c}{\sqrt{n}},\\
&\Big|\sum_{r=1}^{k}\sum_{i=1}^{m}E\zeta_{ri}\Big|\le\sum_{r=1}^{k}\sum_{i=1}^{m}E\bigg(\int_{\frac{\alpha_i}{\sqrt{n}}}^{\frac{\beta_i}{\sqrt{n}}}(\Psi(e_{[r]i}+u)-\Psi(e_{[r]i}))\,\mathrm{d}u\bigg)^2\\
&\qquad\le\sum_{r=1}^{k}\sum_{i=1}^{m}\Big|\frac{1}{\sqrt{n}}(\alpha_i-\beta_i)\Big|\cdot\bigg|\int_{\frac{\alpha_i}{\sqrt{n}}}^{\frac{\beta_i}{\sqrt{n}}}E(\Psi(e_{[r]i}+u)-\Psi(e_{[r]i}))^2\,\mathrm{d}u\bigg|\\
&\qquad\le\frac{c}{\sqrt{n}}\sum_{r=1}^{k}\sum_{i=1}^{m}\bigg|\int_{\frac{\alpha_i}{\sqrt{n}}}^{\frac{\beta_i}{\sqrt{n}}}|u|\,\mathrm{d}u\bigg|\\
&\qquad\le\frac{c}{\sqrt{n}}\sum_{r=1}^{k}\sum_{i=1}^{m}\Big|\frac{\alpha_i^2}{n}+\frac{\beta_i^2}{n}\Big|\le cn^{2b-\frac12},
\end{aligned} \tag{40}$$

$$\sum_{r=1}^{k}\sum_{i=1}^{m}\mathrm{Var}(\zeta_{ri}) \le \sum_{r=1}^{k}\sum_{i=1}^{m}E(\zeta_{ri}^2) = \sum_{r=1}^{k}\sum_{i=1}^{m}E\Big(\int_{\frac{\alpha_i}{\sqrt{n}}}^{\frac{\beta_i}{\sqrt{n}}}(\Psi(e_{[r]i}+u)-\Psi(e_{[r]i}))\,\mathrm{d}u\Big)^4$$
$$\le \sum_{r=1}^{k}\sum_{i=1}^{m}\Big|\frac{1}{\sqrt{n}}(\alpha_i-\beta_i)\Big|\cdot\Big|\int_{\frac{\alpha_i}{\sqrt{n}}}^{\frac{\beta_i}{\sqrt{n}}}E(\Psi(e_{[r]i}+u)-\Psi(e_{[r]i}))^{\frac{4}{3}}\mathrm{d}u\Big|$$
$$\le \frac{c}{\sqrt{n}}\sum_{r=1}^{k}\sum_{i=1}^{m}\Big|\int_{\frac{\alpha_i}{\sqrt{n}}}^{\frac{\beta_i}{\sqrt{n}}}|u|\,\mathrm{d}u\Big|^3$$
$$\le \frac{c}{\sqrt{n}}\sum_{r=1}^{k}\sum_{i=1}^{m}\Big|\frac{\alpha_i^2}{n}+\frac{\beta_i^2}{n}\Big|^3 \le cn^{6b-\frac{5}{2}}, \tag{41}$$

其中 c 为一与 $\{\alpha_i\},\{\beta_i\}$ 无关的常数. 由引理 3.1 知, 对任给的 $\varepsilon>0$, 存在与 $\{\alpha_i\},\{\beta_i\}$ 无关的常数 $c_1>0$ 和 n_1, 当 $n\ge n_1$ 时

$$P\Big\{\sum_{r=1}^{k}\sum_{i=1}^{m}\zeta_{ri}\ge\varepsilon n^{3b-\frac{1}{2}}\Big\} \le P\Big\{\Big|\sum_{r=1}^{k}\sum_{i=1}^{m}(\zeta_{ri}-E\zeta_{ri})\Big|>\frac{\varepsilon}{2}n^{3b-\frac{1}{2}}\Big\}$$
$$\le 2\exp\Big(-\frac{\varepsilon^2 n^{6b-1}}{8cn^{6b-\frac{5}{2}}+4cn^{3b-\frac{3}{2}}}\Big)$$
$$\le 2\exp(-c_1 n^{3b+\frac{1}{2}}). \tag{42}$$

引理 3.4 在定理 1.1 的条件下, 给定 $\gamma>0$, 对几乎所有的样本序列 $X_{[1]1},X_{[1]2},\cdots,X_{[k]m}$ 有

$$\sup_{|\alpha-\beta|\le\gamma}\Big|\sum_{r=1}^{k}\sum_{i=1}^{m}(R_{ri}(\omega_i,\beta)-R_{ri}(\omega_i,\alpha))\Big| = o_p(1),\quad m\to\infty. \tag{43}$$

证 不妨假定 α,β 同号. 若不同号, 则由 Ψ 的单调性知

$$\Big|\int_{\frac{\alpha}{\sqrt{n}}}^{\frac{\beta}{\sqrt{n}}}\omega_i\{\Psi(e_{[r]i}+u)-\Psi(e_{[r]i})\}\,\mathrm{d}u\Big|$$
$$\le \int_0^{\frac{\beta}{\sqrt{n}}}\omega_i\{\Psi(e_{[r]i}+u)-\Psi(e_{[r]i})\}\,\mathrm{d}u+\int_{\frac{\alpha}{\sqrt{n}}}^{0}\omega_i\{\Psi(e_{[r]i}+u)-\Psi(e_{[r]i})\}\,\mathrm{d}u,$$

其余证明如下. 由 Schwarz 不等式,

$$\sum_{r=1}^{k}\sum_{i=1}^{m}\Big|\int_{\frac{\alpha}{\sqrt{n}}}^{\frac{\beta}{\sqrt{n}}}|G(u)|\,\mathrm{d}u\Big| \le c\sum_{r=1}^{k}\sum_{i=1}^{m}\Big|\int_{\frac{\alpha}{\sqrt{n}}}^{\frac{\beta}{\sqrt{n}}}|u|\,\mathrm{d}u\Big|$$
$$\le \frac{c}{n}\sum_{r=1}^{k}\sum_{i=1}^{m}|\alpha-\beta|*|\alpha+\beta|$$
$$\le cn^b*n^{-2b}=o(1), \tag{44}$$

故对任给的 $\varepsilon>0$, 存在 $n_2\geq n_1$, 当 $n\geq n_2$ 且 $X\notin A_n(\frac{\varepsilon}{6},1)$ 时，有

$$\begin{aligned}
&P^*\left\{\sup_{\alpha,\beta\in D}\left|\sum_{r=1}^{k}\sum_{i=1}^{m}\{R_{ri}^*(\omega_i,\beta)-R_{ri}^*(\omega_i,\alpha)\}\right|\geq\frac{\varepsilon}{3}\right\}\\
&\leq P^*\left\{\sup_{\alpha,\beta\in D}\left|\sum_{r=1}^{k}\sum_{i=1}^{m}\{(R_{ri}^*(\omega_i,\beta)-R_{ri}^*(\omega_i,\alpha))-(R_{ri}^*(1,\beta)-R_{ri}^*(1,\alpha))\}\right|\geq\frac{\varepsilon}{6}\right\}\\
&\leq c\sum_{r=1}^{k}\sum_{i=1}^{m}\mathrm{Var}^*\big\{R_{ri}^*(\omega_i,\beta)-R_{ri}^*(\omega_i,\alpha)\big\}\\
&\leq c_2\sum_{r=1}^{k}\sum_{i=1}^{m}\left[\int_{\frac{\alpha}{\sqrt{n}}}^{\frac{\beta}{\sqrt{n}}}\{\Psi(e_{[r]i}+u)-\Psi(e_{[r]i})\}\mathrm{d}u\right]^2\\
&=c_2\sum_{r=1}^{k}\sum_{i=1}^{m}\zeta_{ir},
\end{aligned}\tag{45}$$

同理可得，对任给的 $\varepsilon>0,\gamma>0$, 当 $\alpha,\beta\in D$ 且 $|\alpha-\beta|\leq 2\gamma$, $X\notin A_n(\frac{\varepsilon}{6},\,2\gamma)$ 时，存在 $n_3\geq n_2$, 使得当 $n\geq n_3$ 时，

$$\begin{aligned}
&P^*\left\{\left|\sum_{r=1}^{k}\sum_{i=1}^{m}\left[R_{ri}^*(\omega_i,\beta)-R_{ri}^*(\omega_i,\alpha)\right]\right|\geq\frac{\varepsilon}{3}\right\}\\
&\leq c_3\sum_{r=1}^{k}\sum_{i=1}^{m}\left(\int_{\frac{\alpha}{\sqrt{n}}}^{\frac{\beta}{\sqrt{n}}}\{\Psi(e_{[r]i}+u)-\Psi(e_{[r]i})\}\,\mathrm{d}u\right)^2\\
&=c_3\sum_{r=1}^{k}\sum_{i=1}^{m}\zeta_{ir}(\alpha,\beta),
\end{aligned}\tag{46}$$

任意取定 $\beta_l\in D_l, l=1,2,\cdots,N$. 由上所证，对任给的 $\varepsilon>0,\gamma>0$, 当 $n\geq n_3$ 时，

$$\begin{aligned}
P_{ri}^*(\varepsilon)&=P^*\left\{\sup_{|\alpha-\beta|\leq\gamma}\left|\sum_{r=1}^{k}\sum_{i=1}^{m}(R_{ri}^*(\omega_i,\beta)-R_{ri}^*(\omega_i,\alpha))\right|\geq\varepsilon\right\}\\
&\leq 2N\sum_{l=1}^{N}P^*\left\{\sup_{\beta\in D_l}\left|\sum_{r=1}^{k}\sum_{i=1}^{m}\{R_{ri}^*(\omega_i,\beta)-R_{ri}^*(\omega_i,\beta_l)\}\right|\geq\frac{\varepsilon}{3}\right\}\\
&\quad+\sum_{|\beta_j-\beta_k|\leq 2\gamma}P^*\left\{\left|\sum_{r=1}^{k}\sum_{i=1}^{m}\{R_{ri}^*(\omega_i,\beta_j)-R_{ri}^*(\omega_i,\beta_k)\}\right|\geq\frac{\varepsilon}{3}\right\}\\
&\leq 2Nc_2\sum_{l=1}^{N}\sum_{r=1}^{k}\sum_{i=1}^{m}\zeta_{irl}+c_3\sum_{|\beta_j-\beta_k|\leq 2\gamma}\sum_{r=1}^{k}\sum_{i=1}^{m}\zeta_{ri}(\beta_j,\beta_k),
\end{aligned}\tag{47}$$

任给 $\varepsilon>0$, 则由 (47) 和引理 3.3 与 b 的取法知，当 $n\geq n_3$ 时，

$$\begin{aligned}
&P\Big\{P^*_{ri}(\varepsilon)\geq\epsilon_1, X\notin A_{ri}\Big(\frac{\varepsilon}{6},\ 1\Big)\bigcup A_{ri}\Big(\frac{\varepsilon}{6},\ 2\gamma\Big)\Big\}\\
&\leq\sum_{l=1}^{N}P\Big\{\sum_{r=1}^{k}\sum_{i=1}^{m}\zeta_{ril}\geq\frac{\varepsilon_1}{4c_2N^2}\Big\}+\sum_{|\beta_j-\beta_k|\leq2\gamma}P\Big\{\sum_{r=1}^{k}\sum_{i=1}^{m}\zeta_{ri}(\beta_j,\beta_k)\geq\frac{\varepsilon_1}{2c_3N^*}\Big\}\\
&\leq\sum_{l=1}^{N}P\Big\{\sum_{r=1}^{k}\sum_{i=1}^{m}\zeta_{ril}\geq n^{3b-\frac12}\Big\}+\sum_{|\beta_j-\beta_k|\leq2\gamma}P\Big\{\sum_{r=1}^{k}\sum_{i=1}^{m}\zeta_{ri}(\beta_j,\beta_k)\geq n^{3b-\frac12}\Big\}\\
&\leq 4N^2\exp\{-c_1n^{3b+\frac12}\},
\end{aligned}\tag{48}$$

由此立得，对任给 $\varepsilon_1>0$,

$$P\Big\{P^*_{ri}(\varepsilon)\geq\varepsilon_1,\ X\notin A_{ri}\Big(\frac{\varepsilon}{6},\ 1\Big)\bigcup A_{ri}\Big(\frac{\varepsilon}{6},\ 2\gamma\Big),\quad\text{i.o.}\Big\}=0.\tag{49}$$

用类似方法可以得出，对任给的 $\varepsilon>0$,

$$P\Big\{X\notin A_{ri}\Big(\frac{\varepsilon}{6},\ 1\Big)\bigcup A_{ri}\Big(\frac{\varepsilon}{6},\ 2\gamma\Big),\ \text{i.o.}\Big\}=0.\tag{50}$$

由 (49), (50) 式得，对几乎所有样本序列，当 $X_{[1]1},X_{[1]2},\cdots,X_{[k]m}$ 给定时，

$$\sup_{|\alpha-\beta|\leq\gamma}\Big|\sum_{r=1}^{k}\sum_{i=1}^{m}(R^*_{ri}(\omega_i,\beta)-R^*_{ri}(\omega_i,\alpha))\Big|=o_p(1),\quad m\to\infty,\tag{51}$$

而

$$\begin{aligned}
R_{ri}(\alpha,\beta)&=a\int_0^{\frac{\beta}{\sqrt{n}}}\{\Psi(e_{[r]i}+u)-\Psi(e_{[r]i})\}\,\mathrm{d}u-\frac12\lambda\Big(\frac{\beta}{\sqrt{n}}\Big)^2\\
&=R^*_{ri}(\alpha-\beta)+\int_0^{\frac{\beta}{\sqrt{n}}}(G(u)-\lambda u)\,\mathrm{d}u.
\end{aligned}\tag{52}$$

由 Schwarz 不等式，对任意的 $\alpha,\beta\in D$ 和 $0<b<\frac{\delta}{5}$, 有

$$\begin{aligned}
&\sum_{r=1}^{k}\sum_{i=1}^{m}\Big|\int_{\frac{\alpha}{\sqrt{n}}}^{\frac{\beta}{\sqrt{n}}}(G(u)-\lambda u)\mathrm{d}u\Big|\\
&\leq\sum_{r=1}^{k}\sum_{i=1}^{m}\Big|\int_{\frac{\alpha}{\sqrt{n}}}^{\frac{\beta}{\sqrt{n}}}u^{1+\delta}\mathrm{d}u\Big|\leq c\sum_{r=1}^{k}\sum_{i=1}^{m}\Big(\Big|\frac{\beta}{\sqrt{n}}\Big|^{2+\delta}+\Big|\frac{\alpha}{\sqrt{n}}\Big|^{2+\delta}\Big)\\
&\leq cn^{-\delta(\frac12-\frac{2+\delta}{\delta}b)}=o(1).
\end{aligned}\tag{53}$$

结合 (51), (53) 即可得结论.

4　M 估计的相对效率

这节中我们考虑序集抽样的 M 估计相对与简单抽样的 M 估计的效率. $\widehat{\mu}_n$ 如 (2) 定义. 简单抽样下 μ 的 M 估计 $\widetilde{\mu}$, $\widetilde{\mu}$ 为满足

$$\sum_{i=1}^{n}\rho(X_i-\widetilde{\mu})=\inf_{\mu\in R}\sum_{i=1}^{n}\rho(X_i-\mu) \tag{54}$$

的一个解. 则在假定 1)– 3) 下，有

$$\sqrt{n}(\widetilde{\mu}-\mu_0)\xrightarrow{\mathcal{L}}N\Big(0,\frac{\sigma^2}{\lambda^2}\Big). \tag{55}$$

具体可见文 [16]. 则渐近相对效率

$$\begin{aligned}\mathrm{ARE}(\widetilde{\mu},\widehat{\mu}_n)&=\frac{\sigma^2}{\sigma^2_{RSS}}\\&=\frac{\sigma^2}{\sigma^2-\frac{1}{k}\sum\limits_{r=1}^{k}\big(\int\Psi(x-\mu_0)\mathrm{d}F_{[r]}(x)\big)^2}.\end{aligned} \tag{56}$$

易知 ARE 大于 1.

5　模拟的结果

我们对 X 取指数分布的情况进行模拟. 表 1 是在 $\rho(x)=x^2$, $\rho(x)=\frac{1}{2}x^2I(|x|\le 1)+(|x|-\frac{1}{2})I(|x|>1)$, 以及 $\rho(x)=\frac{1}{2}xI(x>0)+xI(x\le 0)$ 下，对 k 分别取 2 到 8 时的渐近相对效率. 从中我们也可以看出， $\mathrm{ARE}>1$, 序集抽样比简单抽样更有效，并且 k 越大效果越明显.

为了估计 $\sqrt{n}(\widehat{\mu}_n-\mu_0)$ 的分布，我们在给定样本的情况下作了关于随机加权的模拟. 这里 ω 取分布为 Gamma(1,1), 仅对 $\rho(x)=\frac{1}{2}xI(x>0)+xI(x\le 0)$ 及 $k=3$, $m=2000$ 取 2000 个样本作模拟. 从图 1 和图 2 上可以看出在给定样本下随机加权的分布与正态分布还是很接近的，在分布图上可以看出， $\sqrt{n}(\widehat{\mu}_n-\mu_0)$ 的分布曲线 (Mestimate) 与随机加权曲线 (Weighted) 要比正态分布曲线 (Normal) 更为接近. 图 1 中，横坐标为标准正态分位数，纵坐标为 $\sqrt{n}(\mu^*-\widehat{\mu}_n)$ 的分位数. 此外，我们还分别计算了这三个分布的主要几个分位点，从表 2 可以看出 $\sqrt{n}(\widehat{\mu}_n-\mu_0)$ 的主要分位点与 $\sqrt{n}(\mu^*-\widehat{\mu}_n)$ 更为接近.

表 1　x 取 exp(0.5)-2 时的渐近相对效率

$\rho(x)$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$
x^2	1.361	1.632	1.984	1.777	2.462	2.822	2.512
$\begin{cases}\frac{1}{2}x^2, & \lvert x\rvert\le 1\\ \lvert x\rvert-\frac{1}{2}, & \lvert x\rvert>1\end{cases}$	1.515	1.766	2.501	2.838	2.967	3.397	4.044
$\begin{cases}-x, & x\le 0\\ \frac{1}{2}x, & x>0\end{cases}$	1.323	1.533	1.634	1.872	2.166	2.290	2.288

表 2 分位点的比较

α	0.01	0.025	0.05	0.1	0.9	0.95	0.975	0.99
N	-3.044	-2.565	-2.153	-1.677	1.677	2.153	2.565	3.044
$\widehat{\mu}_n$	-1.899	-1.586	-1.308	-1.054	1.014	1.250	1.531	1.834
$\widehat{\mu}_n^w$	-1.821	-1.627	-1.477	-1.303	0.984	1.225	1.570	1.649

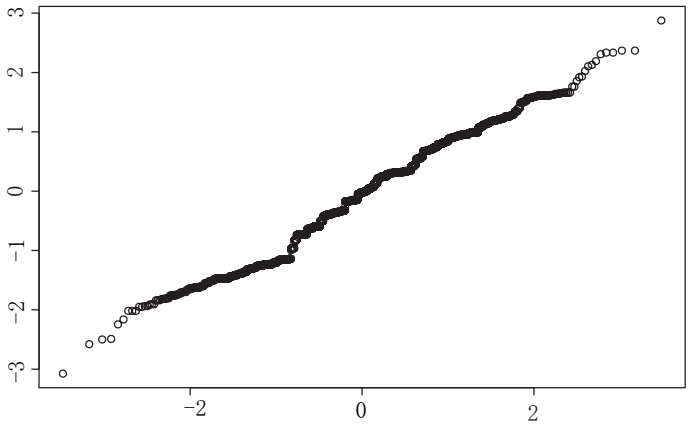

图 1

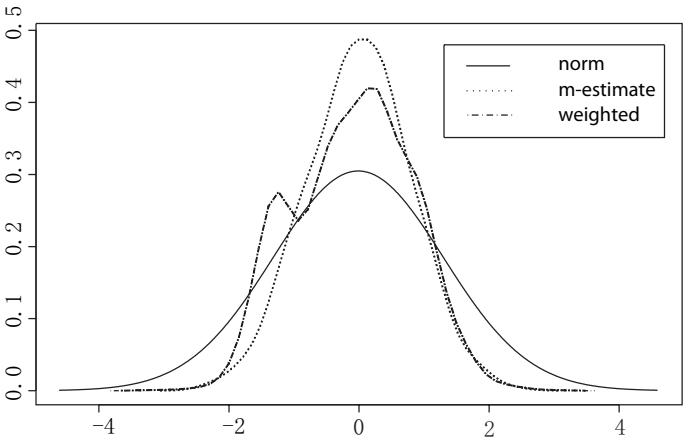

图 2 分布函数图

参 考 文 献

[1] McIntyre G A. A method for unbiased selective sampling using ranked sets. *Australian Journal of Agricultural Research*, 1952, **3**: 385–390.

[2] Takahasi K and Wakimoto K. On unbiased estimates of the population mean based on the sample stratified by means of ordering. *Annals of the Institute of Statistical Mathematics*, 1968, **20**: 1–31.

[3] Dell T R and Clutter J L. Ranked set sampling theory with order statistics background. *Biometrika*, 1972, **28**: 545–555.
[4] Stokes S L. Estimation of variance using judgement ordered ranked set samples. *Biometrics*, 1980, **36**: 35–42.
[5] Stokes S L and Sager T W. Characterization of a ranked set sample with application to estimating distribution functions. *Journal of the American Statistical Association*, 1988, **83**: 374–381.
[6] Bohn L L and Wolfe D A. Non-parametric two sample procedures for ranked set sampling data. *J. Amer. Statist. Assoc.*, 1992, **87**: 552–561.
[7] Hettmansperger T P. The ranked set sample sign test. *Journal of Nonparametric Statistics*, 1995, **4**: 263–270.
[8] Presnell B and Bohn L L. U-statistics and imperfect ranking in ranked set samping. *Journal of Nonparametric Statistics*, 1999, **10**: 111–126.
[9] Chen Z and Bai Z D. The optimal ranked-set sampling scheme for parametric families. *sankhya Ser.* A, 2000, **62**: 178–192.
[10] Shen W H. Use of ranked set sampling for test of a normal mean. *Calcutta Statistical Association Bulletin*, 1994, **44**: 183–193.
[11] Stokes S L. Parametric ranked set sampling. *Annals of the Institute of Statistical Mathematics*, 1995, **47**: 465–482.
[12] Chen Z, Bai Z D and Sinha B K. Ranked Set Sampling: Theory and Applications. Springer, New York, 2003.
[13] Zhao X and Chen Z. On the ranked-set sampling M-estimates for symmetric location families. *Annals of the Institute of Statistical Mathematics*, 2002, **54**: 626–640.
[14] Bennett G. Probability inequalities for sums of independent random variables. *J. Amer. Statist. Assoc.*, 1962, **57**: 33–45.
[15] 林正炎，陆传荣，苏中根．概率极限理论基础．北京：高等教育出版社， 1999.
[16] 陈希孺，赵林城．线性模型中的 M 方法．上海科学技术出版社， 1996.
[17] Rao C R and Zhao L C. Approximation to the distribution of M-estimates in linear models by randomly weighted bootstrap. *The Indian Journal of Statistics*, 1992, **54**: 323–331.
[18] 吴耀华，赵林城．线性模型中随机加权自助法的大样本研究．中国科学， 1999, **29**: 616–624.
[19] Zhao L C. Rate of a.s. convergence of the estimation of error variance in linear models. *Chin. Ann. of Math. B*, 1983, **4**: 95–103.

APPROXIMATION TO THE DISTRIBUTION OF M-ESTIMATES IN RANKED-SET SAMPLING BY RANDOMLY WEIGHTED BOOTSTRAP

WU Yaohua LIU Chiyu

(*University of Science and Technology of China, Hefei, Anhui* 230026)

Abstract Ranked-Set Sampling(RSS) is a sampling method when a set of sampling units drawn from the population can be ranked by certain means rather cheaply without the actual measurement of the variable of interest which is costly and/or time consuming. This paper is concerned with the consistency and asymptotic normality on the RSS M-estimates and approximation to its distribution by randomly weighted bootstrap.

Key words Ranked-Set sampling, M-estimates, randomly weighted bootstrap, asymptotic normality.