应用概率统计　第二十二卷
第四期　2006 年 11 月

Chinese Journal of Applied Probability
and Statistics Vol. 22 No. 4 Nov. 2006

区间数据任意阶原点矩的估计 *

邓文丽
(江西师范大学数学与信息科学学院，南昌，330027)

郑祖康
(复旦大学管理学院统计系，上海，200433)

摘　要

在生存分析和可靠性研究中，区间数据的存在常常使得传统的统计方法无法直接使用．本文从无偏转换的思想出发，对区间数据的任意阶原点矩进行了估计．当截断变量的分布密度函数已知时，得到了一批具有强相合性 (收敛速度可以达到 $n^{-1/2}(\log\log n)^{1/2}$) 和渐近正态性的估计量，并通过模拟计算对这种估计方法的可行性和有效性进行了验证．

关 键 词：区间数据，无偏转换，强相合性，渐近正态性.

学 科 分 类 号：O212.1.

§1.　背景介绍

在生存分析和可靠性研究中，常常因为客观条件的限制无法得到失效时间的准确观测值，只能观测到它所处的区间，统计学中一般称这类数据为区间截断数据 (Interval censored data), 简称区间数据．例如，在一些传染性疾病的感染时间研究中，实验对象被放入感染源后，染上传染病所经历的时间是无法准确观测到的，只能知道它处在某区间中．通过破坏性实验进行的产品寿命研究 (如灭火器), 产品的寿命一般是无法直接观测到的，只能知道该产品在某个时间点前后失效．这两个例子中得到的数据都是区间数据．区间数据有着广泛的应用背景，对它进行深入的研究是很有必要的．

为了统计处理上的方便，常常将区间数据分为以下两类 (用 Y 表示原本想要观测的随机变量). 在实验中，只观测到了 (V,δ), V 表示 "检查" 或 "观测" 的时间，$\delta=I_{(Y\le V)}$, 其中 $I_{(\cdot)}$ 是示性函数．这样的区间数据称为 *区间截断情况 1* (*"Case 1" Interval Censoring*) *或称当前状态数据* (*Current Status Data*). 在另一种情形中，只能知道 Y 相对于某个随机区间 (U,V) 的位置，实验只观测到 (U,V,δ_1,δ_2), $\delta_1=I_{(Y\le U)}$, $\delta_2=I_{(U<Y\le V)}$. 这样的区间数据称为 *区间截断情况 2* (*"Case 2" Interval Censoring*). 在区间截断情况 2 中，如果 $U=0$ (或者 $V=\infty$), 就退化成了区间截断情况 1. 本文将只对区间截断情况 2 进行研究．

当某随机变量的观测数据是区间数据时，大家比较关注的统计问题主要集中在两大类：一是如何通过这批区间数据去分析该随机变量的一些统计特征，如均值、方差以及分布函数；再就是如何通过这批区间数据去分析该随机变量和其他随机变量之间的关系，如回归分析．在以往的区间数据研究中，有大量的工作都是集中在这两方面的，它们大多数采取的是非参数极大似然的方法，Huang and Wellner (1997) 对此进行了比较详细的介绍．无论是区间数据的分布函数的估计，或者是一些特殊模型的参数 (或半参数) 回归问题，在用非参数极大似然方法进行

* 国家自然科学基金 (70171008) 资助项目，江西师范大学青年成长基金项目 (1069).

本文 2005 年 1 月 20 日收到，2005 年 7 月 22 日收到修改稿.

分析时，最后都可以归结为对似然方程求解的问题．从中所得到估计具有一般似然估计所具有的很好的大样本性质．但是似然方法有一个无法回避的弱点，那就是解方程的过程过于繁琐，一般来说都需要通过迭代计算才能将方程解出来，实际操作难度大，并且只能得到似然方程的渐近解．所以在本文中将尝试用较为简捷的 “无偏转换”(Unbiased transformation) 思想，来解决区间数据任意阶原点矩的估计问题．这种思想方法在右截断数据 (Right censored data) 的统计处理中曾发挥过很大的作用，得到了一些有较好统计性质的估计量．和极大似然方法相比，“无偏转换” 的做法还具有简单易操作的优势．

在简单线性模型的参数估计问题中，当响应变量是右截断数据时，最小二乘方法无法直接使用．为了解决这个问题， Burkley and James (1979), Koul, Susarla and Van Ryzin (1981) 设想：是否能将观测值进行修改，以代替被截断的响应变量，然后再用传统的最小二乘方法对参数进行估计．Zheng (1984) 提出了一类具有代表性的无偏转换方法，命名为 Class K 方法．这种方法的主要思想是：当随机变量 Y 被截断时，应该对它作一些补偿；当 Y 未被截断时，也要对它作一些调整．Zheng (1984) 证明了，当截断变量的分布函数已知时，在较一般的条件下，由 Class K 提供的回归系数的估计具有强相合性、渐近正态性以及 L^2 收敛的性质．在 Zheng (2003) 中，首次将 “无偏转换” 的思想用于区间数据的处理，解决了区间数据期望的估计问题以及当响应变量是区间数据时简单回归模型中回归系数的估计问题．

在这些成果的基础上, 本文将对区间数据的任意阶原点矩 $\mathsf{E}Y^r$ $(r>1)$ 进行估计．文中假定观测时间 (U,V) 的分布已知，这主要出于两方面的考虑．一方面，在实践中确实存在着 (U,V) 分布已知的情形, 例如, 为了对不易观测的随机变量 Y 进行分析, 自己设计试验过程, 将 (U,V) 设置为服从某特殊分布的二元随机向量；另一方面，如果 (U,V) 分布已知场合下的估计问题解决了， (U,V) 分布未知场合下的估计问题就会迎刃而解．文章的第二部分针对截断变量 (U,V) 的分布密度函数已知的情形，得到了一批具有强相合性 (收敛速度可以达到 $n^{-1/2}(\log\log n)^{1/2}$) 和渐近正态性的估计量．第三部分通过模拟计算对不同无偏转换方式的估计效果进行了对比．

为了表述上的方便，文中采用统一的记号表示：用 $F(\cdot)$ 表示 Y 的分布函数， $G(\cdot,\cdot)$ 表示 (U,V) 的联合分布函数， $g(\cdot,\cdot)$ 表示 (U,V) 的联合分布密度函数．

§2. 任意阶原点矩的估计

假定 Y 是一个非负随机变量，有未知的连续分布函数 $F(\cdot)$, 分布函数的具体形式未知，但满足 $\mathsf{E}Y^r<\infty$, r 是一个大于等于 1 的常数．要估计 $\mathsf{E}Y^r$.

(U,V) 是一个和 Y 独立的随机向量，有已知的正密度函数 $g(\cdot,\cdot)$, $0<U\le V<\infty$ 几乎处处成立，其分布函数记为 $G(\cdot,\cdot)$.

实验中只观测到了独立样本 $(U_i,V_i,\delta_{1i},\delta_{2i})$ $(i=1,2,\cdots,n)$, 其中 $\delta_{1i}=I_{(Y_i\le U_i)}$, $\delta_{2i}=I_{(U_i<Y_i\le V_i)}$.

令

$$Y_i^{(r)}=\phi_1(U_i,V_i)\delta_{1i}+\phi_2(U_i,V_i)\delta_{2i}+\phi_3(U_i,V_i)(1-\delta_{1i}-\delta_{2i}). \tag{2.1}$$

其中 $\phi_1(\cdot,\cdot)$, $\phi_2(\cdot,\cdot)$, $\phi_3(\cdot,\cdot)$ 是三个和 $F(\cdot)$ 无关的连续函数，但是可能和 G 有关．

有下面的定理：

定理 2.1 如果 $\phi_1(\cdot,\cdot)$, $\phi_2(\cdot,\cdot)$, $\phi_3(\cdot,\cdot)$ 满足下面的方程

$$\begin{cases}\displaystyle\int_{v=0}^{\infty}\int_{u=0}^{v}\phi_1(u,v)g(u,v)\mathrm{d}u\mathrm{d}v=0;\\ \displaystyle\int_{y}^{\infty}[\phi_2(y,v)-\phi_1(y,v)]g(y,v)\mathrm{d}v+\int_0^y[\phi_3(u,y)-\phi_2(u,y)]g(u,y)\mathrm{d}u=ry^{r-1}.\end{cases}\tag{2.2}$$

那么

$$\mathsf{E}Y_i^{(r)}=\mathsf{E}Y_i^r$$

对任意的 $F(\cdot)$ 成立.

证明:

$$\begin{aligned}
&\mathsf{E}[\phi_1(U_i,V_i)I_{(0<Y_i\le U_i)}+\phi_2(U_i,V_i)I_{(U_i<Y_i\le V_i)}+\phi_3(U_i,V_i)I_{(V_i<Y_i)}]\\
=&\iiint\limits_{0<y\le u\le v}\phi_1(u,v)g(u,v)\mathrm{d}u\mathrm{d}v\mathrm{d}F(y)+\iiint\limits_{0<u<y\le v<\infty}\phi_2(u,v)g(u,v)\mathrm{d}u\mathrm{d}v\mathrm{d}F(y)\\
&+\iiint\limits_{0<u\le v<y<\infty}\phi_3(u,v)g(u,v)\mathrm{d}u\mathrm{d}v\mathrm{d}F(y)\\
=&\int_0^{\infty}\Big[\int_{v=y}^{\infty}\int_{u=y}^{v}\phi_1(u,v)g(u,v)\mathrm{d}u\mathrm{d}v+\int_{v=y}^{\infty}\int_{u=0}^{y}\phi_2(u,v)g(u,v)\mathrm{d}u\mathrm{d}v\\
&+\int_{v=0}^{y}\int_{u=0}^{v}\phi_3(u,v)g(u,v)\mathrm{d}u\mathrm{d}v\Big]\mathrm{d}F(y).
\end{aligned}$$

令

$$\varphi=\phi_2-\phi_1,\qquad\theta=\phi_3-\phi_2.$$

根据 (2.2), 有

$$\begin{aligned}
&\int_{v=y}^{\infty}\int_{u=y}^{v}\phi_1(u,v)g(u,v)\mathrm{d}u\mathrm{d}v+\int_{v=y}^{\infty}\int_{u=0}^{y}\phi_2(u,v)g(u,v)\mathrm{d}u\mathrm{d}v\\
&+\int_{v=0}^{y}\int_{u=0}^{v}\phi_3(u,v)g(u,v)\mathrm{d}u\mathrm{d}v\\
=&\int_{v=y}^{\infty}\int_{u=y}^{v}\phi_1(u,v)g(u,v)\mathrm{d}u\mathrm{d}v+\int_{v=y}^{\infty}\int_{u=0}^{y}[\phi_1(u,v)+\varphi(u,v)]g(u,v)\mathrm{d}u\mathrm{d}v\\
&+\int_{v=0}^{y}\int_{u=0}^{v}[\phi_1(u,v)+\varphi(u,v)+\theta(u,v)]g(u,v)\mathrm{d}u\mathrm{d}v\\
=&\int_{v=0}^{\infty}\Big[\int_{u=0}^{v}\phi_1(u,v)g(u,v)\mathrm{d}u\Big]\mathrm{d}v+\Big(\int_{v=y}^{\infty}\int_{u=0}^{y}+\int_{v=0}^{y}\int_{u=0}^{v}\Big)\varphi(u,v)g(u,v)\mathrm{d}u\mathrm{d}v\\
&+y^r-\int_{v=0}^{y}\int_{t=v}^{\infty}\varphi(v,t)g(v,t)\mathrm{d}t\mathrm{d}v\\
=&\Big(\int_{v=y}^{\infty}\int_{u=0}^{y}+\int_{v=0}^{y}\int_{u=0}^{v}\Big)\varphi(u,v)g(u,v)\mathrm{d}u\mathrm{d}v+y^r\\
&-\Big(\int_{t=0}^{y}\int_{v=0}^{t}+\int_{t=y}^{\infty}\int_{v=0}^{y}\Big)\varphi(v,t)g(v,t)\mathrm{d}v\mathrm{d}t\\
=&\ y^r.
\end{aligned}$$

说明 1 (2.2) 的第一个式子主要用于将积分调整为常数．满足这样的条件的 $\phi_1(u,v)$ 有很多．例如：

(1) $\phi_1(u,v)=0$;

(2) $\phi_1(u,v)=u-\mathsf{E}U$, 当 $\mathsf{E}U<\infty$;

(3) $\phi_1(u,v)=v-\mathsf{E}V$, 当 $\mathsf{E}V<\infty$.

一般地说，要找到满足 (2.2) 第一个式子的 $\phi_1(u,v)$ 并不需要对 (U,V) 有太多假定，只需假定存在一个连续函数 $\rho(u,v)$, 使得

$$\int_0^\infty\int_0^v \rho(u,v)g(u,v)\mathrm{d}u\mathrm{d}v<\infty.$$

就可以保证

$$\phi_1(u,v)=\rho(u,v)-\int_0^\infty\int_0^v \rho(u,v)g(u,v)\mathrm{d}u\mathrm{d}v$$

满足 (2.2) 的第一个式子．

说明 2 选取连续函数 $\rho(\cdot,\cdot)$ 满足 $\mathsf{E}\rho(U,V)<\infty$, 下面有几个满足 (2.2) 的比较一般的例子：

$$(1)\quad \begin{cases}\phi_1(u,v)=\rho(u,v)-\displaystyle\int_0^\infty\int_0^v \rho(u,v)g(u,v)\mathrm{d}u\mathrm{d}v,\\ \phi_2(u,v)=\phi_1(u,v),\\ \phi_3(u,v)-\phi_2(u,v)=(rv^{r-1})\Big/\Big[\displaystyle\int_0^v g(u,v)\mathrm{d}u\Big].\end{cases}$$

特别地，可取 $\phi_1(u,v)=0$, $\phi_2(u,v)=0$, $\phi_3(u,v)=(rv^{r-1})\Big/\Big[\displaystyle\int_0^v g(u,v)\mathrm{d}u\Big]$.

$$(2)\quad \begin{cases}\phi_1(u,v)=\rho(u,v)-\displaystyle\int_0^\infty\int_0^v \rho(u,v)g(u,v)\mathrm{d}u\mathrm{d}v,\\ \phi_2(u,v)-\phi_1(u,v)=(ru^{r-1})\Big/\Big[\displaystyle\int_u^\infty g(u,v)\mathrm{d}v\Big],\\ \phi_3(u,v)=\phi_2(u,v).\end{cases}$$

特别地，可取 $\phi_1(u,v)=0$, $\phi_2(u,v)=\phi_3(u,v)=(ru^{r-1})\Big/\Big[\displaystyle\int_u^\infty g(u,v)\mathrm{d}v\Big]$.

$$(3)\quad \begin{cases}\phi_1(u,v)=\rho(u,v)-\displaystyle\int_0^\infty\int_0^v \rho(u,v)g(u,v)\mathrm{d}u\mathrm{d}v,\\ \phi_2(u,v)-\phi_1(u,v)=(ru^{r-1})\Big/\Big[2\displaystyle\int_u^\infty g(u,v)\mathrm{d}v\Big],\\ \phi_3(u,v)-\phi_2(u,v)=(rv^{r-1})\Big/\Big[2\displaystyle\int_0^v g(u,v)\mathrm{d}u\Big].\end{cases}$$

特别地, 可取 $\phi_1(u,v)=0$, $\phi_2(u,v)=(ru^{r-1})\Big/\Big[2\displaystyle\int_u^\infty g(u,v)\mathrm{d}v\Big]$, $\phi_3(u,v)=(ru^{r-1})\Big/\Big[2\displaystyle\int_u^\infty g(u,v)\mathrm{d}v\Big]$ $+(rv^{r-1})\Big/\Big[2\displaystyle\int_0^v g(u,v)\mathrm{d}u\Big]$.

当 $g(u,v)$ 的具体表达式已知时，很容易就可以得到 $\phi_1(u,v)$, $\phi_2(u,v)$, $\phi_3(u,v)$ 的一些显式表示．

假定

$$g(u,v)=\begin{cases}\dfrac{1}{v}\lambda e^{-\lambda v} & 0<u\le v;\\ 0 & \text{其它},\end{cases}$$

如果选择 $\phi_1(u,v)=0$, (2.2) 的第一个方程成立，第二个方程变成

$$\int_y^{\infty}\phi_2(y,v)\frac{\lambda e^{-\lambda v}}{v}\,\mathrm{d}v+\int_0^y[\phi_3(u,y)-\phi_2(u,y)]\frac{\lambda e^{-\lambda y}}{y}\,\mathrm{d}u=ry^{r-1},$$

而且令 $\phi_2(u,v)=v$, 那么

$$\int_0^y\phi_3(u,y)\mathrm{d}u=\frac{ry^r}{\lambda}e^{\lambda y}-\frac{y}{\lambda}+y^2.$$

这个积分方程有很多解，容易验证

$$\phi_3(u,v)=rv^re^{-\lambda(u-v)}+\frac{rv^{r-1}}{\lambda}+v-\frac{1}{\lambda}$$

是其中一个解．

说明 3 下面的例子给出了一组形式特别简单的解．令 $h(t)$ 满足

$$\int_0^{\infty}h(t)\mathrm{d}t=1,$$

因为 $g(u,v)>0$ 对任意的 $0<u\le v<\infty$ 成立，那么

$$\begin{cases}\phi_2(y,v)-\phi_1(y,v)=\dfrac{ry^{r-1}h(v)}{g(y,v)}\\ \phi_3(u,y)-\phi_2(u,y)=\dfrac{ry^{r-1}h(u)}{g(u,y)}\end{cases}$$

满足 (2.2) 的第二个方程．注意到，在 (2.2) 的第一个方程中，当 $u>v$ 时 $g(u,v)=0$, 若 $T(u,v)$ 为 u,v 的任意一个连续函数，并且满足

$$\int_{v=0}^{\infty}\int_{u=0}^{v}T(u,v)g(u,v)\mathrm{d}u\mathrm{d}v=C<\infty,$$

那么

$$\begin{cases}\phi_1(u,v)=T(u,v)-C\\ \phi_2(u,v)=T(u,v)-C+\dfrac{ru^{r-1}h(v)}{g(u,v)}\\ \phi_3(u,v)=T(u,v)-C+\dfrac{ru^{r-1}h(v)}{g(u,v)}+\dfrac{rv^{r-1}h(u)}{g(u,v)}\end{cases}\tag{2.3}$$

是 (2.2) 的一个解．

将所有满足 (2.2) 的 (ϕ_1,ϕ_2,ϕ_3) 组成的类，用 $H_r^{(\mathrm{II})}$ 表示．为简单起见，也用 $H_r^{(\mathrm{II})}$ 来表示所有 $(\phi_1,\phi_2,\phi_3)\in H_r^{(\mathrm{II})}$ 的形如 (2.1) 的 $Y_i^{(r)}$ 所组成的“估计”类．在这个估计类 $H_r^{(\mathrm{II})}$ 中所有的元素都是 $\mathsf{E}Y^r$ 的无偏估计．

构造出了一类均值为 $\mathsf{E}Y^r$ 的随机变量后，自然要对它们的方差进行研究了.

$$\begin{aligned}
& \mathsf{E}[\phi_1(U,V)I_{Y\in(0,U)}+\phi_2(U,V)I_{Y\in[U,V]}+\phi_3(U,V)I_{Y\in(V,\infty)}]^2 \\
= & \int_0^\infty \Big[\int_{v=y}^\infty\int_{u=y}^v \phi_1^2(u,v)g(u,v)\mathrm{d}u\mathrm{d}v+\int_{v=y}^\infty\int_{u=0}^y \phi_2^2(u,v)g(u,v)\mathrm{d}u\mathrm{d}v \\
& +\int_{v=0}^y\int_{u=0}^v \phi_3^2(u,v)g(u,v)\mathrm{d}u\mathrm{d}v\Big]\mathrm{d}F(y).
\end{aligned}$$

因为

$$\begin{aligned}
& \int_{v=y}^\infty\int_{u=y}^v g(u,v)\mathrm{d}u\mathrm{d}v+\int_{v=y}^\infty\int_{u=0}^y g(u,v)\mathrm{d}u\mathrm{d}v+\int_{v=0}^y\int_{u=0}^v g(u,v)\mathrm{d}u\mathrm{d}v \\
= & \int_{v=y}^\infty\int_{u=0}^v g(u,v)\mathrm{d}u\mathrm{d}v+\int_{v=0}^y\int_{u=0}^v g(u,v)\mathrm{d}u\mathrm{d}v \\
= & \int_{v=0}^\infty\int_{u=0}^v g(u,v)\mathrm{d}u\mathrm{d}v=1,
\end{aligned}$$

并且由 Hölder 积分不等式,

$$\begin{aligned}
\Big|\int_y^\infty\int_y^v \phi_1(u,v)g(u,v)\mathrm{d}u\mathrm{d}v\Big| &\le \Big[\int_y^\infty\int_y^v \phi_1^2(u,v)g(u,v)\mathrm{d}u\mathrm{d}v\cdot\int_y^\infty\int_y^v g(u,v)\mathrm{d}u\mathrm{d}v\Big]^{1/2}, \\
\Big|\int_y^\infty\int_0^y \phi_2(u,v)g(u,v)\mathrm{d}u\mathrm{d}v\Big| &\le \Big[\int_y^\infty\int_0^y \phi_2^2(u,v)g(u,v)\mathrm{d}u\mathrm{d}v\cdot\int_y^\infty\int_0^y g(u,v)\mathrm{d}u\mathrm{d}v\Big]^{1/2}, \\
\Big|\int_0^y\int_0^v \phi_3(u,v)g(u,v)\mathrm{d}u\mathrm{d}v\Big| &\le \Big[\int_0^y\int_0^v \phi_3^2(u,v)g(u,v)\mathrm{d}u\mathrm{d}v\cdot\int_0^y\int_0^v g(u,v)\mathrm{d}u\mathrm{d}v\Big]^{1/2}.
\end{aligned}$$

再利用 Hölder 和不等式，容易得到

$$\begin{aligned}
y^{2r} = & \Big[\int_{v=y}^\infty\int_{u=y}^v \phi_1(u,v)g(u,v)\mathrm{d}u\mathrm{d}v+\int_{v=y}^\infty\int_{u=0}^y \phi_2(u,v)g(u,v)\mathrm{d}u\mathrm{d}v \\
& +\int_{v=0}^y\int_{u=0}^v \phi_3(u,v)g(u,v)\mathrm{d}u\mathrm{d}v\Big]^2 \\
\le & \Big\{\Big[\int_{v=y}^\infty\int_{u=y}^v \phi_1^2(u,v)g(u,v)\mathrm{d}u\mathrm{d}v\cdot\int_{v=y}^\infty\int_{u=y}^v g(u,v)\mathrm{d}u\mathrm{d}v\Big]^{1/2} \\
& +\Big[\int_{v=y}^\infty\int_{u=0}^y \phi_2^2(u,v)g(u,v)\mathrm{d}u\mathrm{d}v\cdot\int_{v=y}^\infty\int_{u=0}^y g(u,v)\mathrm{d}u\mathrm{d}v\Big]^{1/2} \\
& +\Big[\int_{v=0}^y\int_{u=0}^v \phi_3^2(u,v)g(u,v)\mathrm{d}u\mathrm{d}v\cdot\int_{v=0}^y\int_{u=0}^v g(u,v)\mathrm{d}u\mathrm{d}v\Big]^{1/2}\Big\}^2 \\
\le & \int_{v=y}^\infty\int_{u=y}^v \phi_1^2(u,v)g(u,v)\mathrm{d}u\mathrm{d}v+\int_{v=y}^\infty\int_{u=0}^y \phi_2^2(u,v)g(u,v)\mathrm{d}u\mathrm{d}v \\
& +\int_{v=0}^y\int_{u=0}^v \phi_3^2(u,v)g(u,v)\mathrm{d}u\mathrm{d}v,
\end{aligned}$$

也就是说 $\mathsf{Var}\,Y_i^{(r)}\ge\mathsf{Var}\,Y_i^r$. 新构造出来的变量 $Y_i^{(r)}$ 的方差比被截断变量 Y_i^r 的方差大，可以理解为 Y_i^r 被截断所带来的损失.

既然 $H_r^{(\mathrm{II})}$ 中所有的元素都是 $\mathsf{E}Y_i^r$ 的无偏估计，自然会想到在 $H_r^{(\mathrm{II})}$ 中寻找一个对任意的

未知分布 $F(\cdot)$ 而言一致最小方差的 "估计". 记

$$\begin{aligned} S_1 \;=\; & \int_{v=y}^{\infty}\int_{u=y}^{v}\phi_1^2(u,v)g(u,v)\mathrm{d}u\mathrm{d}v+\int_{v=y}^{\infty}\int_{u=0}^{y}\phi_2^2(u,v)g(u,v)\mathrm{d}u\mathrm{d}v \\ & +\int_{v=0}^{y}\int_{u=0}^{v}\phi_3^2(u,v)g(u,v)\mathrm{d}u\mathrm{d}v+\lambda\Big[\int_{v=y}^{\infty}\int_{u=y}^{v}\phi_1(u,v)g(u,v)\mathrm{d}u\mathrm{d}v \\ & +\int_{v=y}^{\infty}\int_{u=0}^{y}\phi_2(u,v)g(u,v)\mathrm{d}u\mathrm{d}v+\int_{v=0}^{y}\int_{u=0}^{v}\phi_3(u,v)g(u,v)\mathrm{d}u\mathrm{d}v-y^r\Big], \end{aligned}$$

其中 λ 可能依赖 y. 假如有 $(\phi_1^0,\phi_2^0,\phi_3^0)\in H_r^{(\mathrm{II})}$, 它使得 $Y_i^{(r)}$ 对所有的 F 而言有最小的二阶矩, 令 $\phi_1=\phi_1^0+\rho\Delta_1$, $\phi_2=\phi_2^0+\delta\Delta_2$, $\phi_3=\phi_3^0+\eta\Delta_3$, 其中 $\Delta_1,\Delta_2,\Delta_3$ 是 u,v 的任意函数, ρ,δ,η 是参数. 那么

$$\begin{aligned} S_1 \;=\; & \int_{v=y}^{\infty}\int_{u=y}^{v}[\phi_1^0(u,v)+\rho\Delta_1]^2g(u,v)\mathrm{d}u\mathrm{d}v+\int_{v=y}^{\infty}\int_{u=0}^{y}[\phi_2^0(u,v)+\delta\Delta_2]^2g(u,v)\mathrm{d}u\mathrm{d}v \\ & +\int_{v=0}^{y}\int_{u=0}^{v}[\phi_3^0(u,v)+\eta\Delta_3]^2g(u,v)\mathrm{d}u\mathrm{d}v+\lambda\Big\{\int_{v=y}^{\infty}\int_{u=y}^{v}[\phi_1^0(u,v)+\rho\Delta_1]g(u,v)\mathrm{d}u\mathrm{d}v \\ & +\int_{v=y}^{\infty}\int_{u=0}^{y}[\phi_2^0(u,v)+\delta\Delta_2]g(u,v)\mathrm{d}u\mathrm{d}v+\int_{v=0}^{y}\int_{u=0}^{v}[\phi_3^0(u,v)+\eta\Delta_3]g(u,v)\mathrm{d}u\mathrm{d}v-y^r\Big\}, \end{aligned}$$

并且

$$\begin{aligned} \frac{\partial S_1}{\partial\rho}\Big|_{\rho=0} &= \int_{v=y}^{\infty}\int_{u=y}^{v}2\phi_1^0(u,v)\Delta_1g(u,v)\mathrm{d}u\mathrm{d}v+\lambda\int_{v=y}^{\infty}\int_{u=y}^{v}\Delta_1g(u,v)\mathrm{d}u\mathrm{d}v, \\ \frac{\partial S_1}{\partial\delta}\Big|_{\delta=0} &= \int_{v=y}^{\infty}\int_{u=0}^{y}2\phi_2^0(u,v)\Delta_2g(u,v)\mathrm{d}u\mathrm{d}v+\lambda\int_{v=y}^{\infty}\int_{u=0}^{y}\Delta_2g(u,v)\mathrm{d}u\mathrm{d}v, \\ \frac{\partial S_1}{\partial\eta}\Big|_{\eta=0} &= \int_{v=0}^{y}\int_{u=0}^{v}2\phi_3^0(u,v)\Delta_3g(u,v)\mathrm{d}u\mathrm{d}v+\lambda\int_{v=0}^{y}\int_{u=0}^{v}\Delta_3g(u,v)\mathrm{d}u\mathrm{d}v. \end{aligned}$$

$(\phi_1^0,\phi_2^0,\phi_3^0)$ 在 $H_r^{(\mathrm{II})}$ 中对任意 F 有最小二阶矩的充要条件是上述三个方程均等于零, 也就是说,

$$\begin{cases} \displaystyle\int_{v=y}^{\infty}\int_{u=y}^{v}[2\phi_1^0(u,v)+\lambda]\Delta_1g(u,v)\mathrm{d}u\mathrm{d}v=0 \\ \displaystyle\int_{v=y}^{\infty}\int_{u=0}^{y}[2\phi_2^0(u,v)+\lambda]\Delta_2g(u,v)\mathrm{d}u\mathrm{d}v=0 \\ \displaystyle\int_{v=0}^{y}\int_{u=0}^{v}[2\phi_3^0(u,v)+\lambda]\Delta_3g(u,v)\mathrm{d}u\mathrm{d}v=0 \end{cases}$$

对任意的 $\Delta_1,\Delta_2,\Delta_3$ 成立. 因此

$$2\phi_1^0(u,v)=2\phi_2^0(u,v)=2\phi_3^0(u,v)=-\lambda.$$

考虑到约束条件

$$\int_{v=y}^{\infty}\int_{u=y}^{v}\phi_1(u,v)g(u,v)\mathrm{d}u\mathrm{d}v+\int_{v=y}^{\infty}\int_{u=0}^{y}\phi_2(u,v)g(u,v)\mathrm{d}u\mathrm{d}v+\int_{v=0}^{y}\int_{u=0}^{v}\phi_3(u,v)g(u,v)\mathrm{d}u\mathrm{d}v=y^r,$$

得到

$$\Big(\int_{v=y}^{\infty}\int_{u=y}^{v}+\int_{v=y}^{\infty}\int_{u=0}^{y}+\int_{v=0}^{y}\int_{u=0}^{v}\Big)\lambda g(u,v)\mathrm{d}u\mathrm{d}v=-2y^r,$$

$$\int_{v=0}^{\infty}\int_{u=0}^{v}\lambda g(u,v)\mathrm{d}u\mathrm{d}v=-2y^r,$$

即

$$\lambda=-2y^r,$$

因此

$$\phi_1^0(u,v)=\phi_2^0(u,v)=\phi_3^0(u,v)=y^r.$$

这和原假定 "ϕ_1,ϕ_2,ϕ_3 与 y 无关" 产生了矛盾． #

定理 2.2 在前文的假定条件下，并不存在 $(\phi_1^0,\phi_2^0,\phi_3^0)\in H_r^{(\mathrm{II})}$ 使得 $Y_i^{(r)}$ 的方差在类 $H_r^{(\mathrm{II})}$ 中对所有的 F 一致的最小．

说明 证明过程也意味着在最小方差意义下 (ϕ_1,ϕ_2,ϕ_3) 的最佳选择就是 y^r, 这也用另一种方法证明了 $\mathsf{Var}\,Y_i^r\le\mathsf{Var}\,Y_i^{(r)}$.

在上述讨论的基础上，很容易就可得出 $\mathsf{E}Y^r$ 的一个优良估计．选取 $Y_i^{(r)}\in H_r^{(\mathrm{II})}$, $i=1,2,\cdots,n$, 用

$$\overline{Y^{(r)}}=\frac{1}{n}\sum_{i=1}^{n}Y_i^{(r)}$$

来估计 $\mathsf{E}Y^r$. 由于它是独立同分布随机变量和的形式，所以下面的定理显然成立．

定理 2.3

(i) $\mathsf{E}\overline{Y^{(r)}}=\mathsf{E}Y^r$;

(ii) $\overline{Y^{(r)}}$ 是强相合的；

(iii) $\overline{Y^{(r)}}$ 的强相合收敛速度为 $n^{-1/2}(\log\log n)^{1/2}$;

(iv) 当 $n\to\infty$ 时， $\sqrt{n}(\overline{Y^{(r)}}-\mathsf{E}Y^r)\to N(0,\sigma^2)$, 其中 $\sigma^2=\mathsf{Var}\,(Y_1^{(r)})$.

§3. 模拟计算

假定 Y 服从均匀分布 $U(0,2)$, 即 $\mathsf{E}Y^2=4/3$, U,V 的联合分布密度函数为

$$g(u,v)=\frac{1}{v}\lambda e^{-\lambda v},\qquad 0<u\le v<\infty, \tag{3.1}$$

产生 (U,V,Y) 的一组随机数后，可以生成模拟的观测值 (U,V,δ_1,δ_2), 下面将根据文中所提供的方法来对 $\mathsf{E}Y^2$ 进行估计．

令

$$Y^*=\phi_1(U,V)\delta_1+\phi_2(U,V)\delta_2+\phi_3(U,V)(1-\delta_1-\delta_2), \tag{3.2}$$

其中 $(\phi_1,\phi_2,\phi_3)\in H_2^{(\mathrm{II})}$. Y^* 的均值即为所求的估计值．

选取不同的 (ϕ_1,ϕ_2,ϕ_3) 会得到不同的估计值，表 3.1 给出了下面四组 (ϕ_1,ϕ_2,ϕ_3) 所对应的模拟结果．

$$\text{(i)}\quad\begin{cases}\phi_1(u,v)=u-\dfrac{1}{2\lambda},\\ \phi_2(u,v)=u-\dfrac{1}{2\lambda},\\ \phi_3(u,v)=u-\dfrac{1}{2\lambda}+\dfrac{2v}{\lambda}e^{\lambda v};\end{cases}\qquad\qquad \text{(ii)}\quad\begin{cases}\phi_1(u,v)=u-\dfrac{1}{2\lambda},\\ \phi_2(u,v)=u-\dfrac{1}{2\lambda}+v,\\ \phi_3(u,v)=u-\dfrac{1}{2\lambda}+v+\dfrac{1}{\lambda}(2ve^{\lambda v}-1);\end{cases}$$

(iii) $\begin{cases}\phi_1(u,v)=0,\\ \phi_2(u,v)=0,\\ \phi_3(u,v)=\dfrac{2v}{\lambda}e^{\lambda v};\end{cases}$ (iv) $\begin{cases}\phi_1(u,v)=0,\\ \phi_2(u,v)=v,\\ \phi_3(u,v)=v+\dfrac{1}{\lambda}(2ve^{\lambda v}-1);\end{cases}$

令 $\lambda=1$, 下面是 100 次随机模拟所得估计的平均值和标准差.

表 3.1 不同 (ϕ_1,ϕ_2,ϕ_3) 下 $\mathsf{E}Y^2$ 的估计 (真实值 $1.333\dot{3}$)

		完全观测	(i)	(ii)	(iii)	(iv)
$n=100$	均值	1.3418	1.2508	1.2559	1.2666	1.2717
	标准差	0.1106	0.2516	0.2577	0.2482	0.2511
$n=200$	均值	1.3379	1.3364	1.3493	1.3489	1.3618
	标准差	0.0890	0.1899	0.1877	0.1866	0.1818
$n=400$	均值	1.337	1.3605	1.3632	1.3680	1.3707
	标准差	0.0538	0.1491	0.1549	0.1461	0.1500
$n=800$	均值	1.3418	1.3351	1.3393	1.3400	1.3442
	标准差	0.0405	0.1025	0.1148	0.0977	0.1085

"完全观测" 对应的是随机变量 Y 没有被截断的情形.

固定 $\phi_1(u,v)=0$, $\phi_2(u,v)=0$, $\phi_3(u,v)=(2v)\Big/\Big[\int_0^v g(u,v)\mathrm{d}u\Big]$, 表 3.2 给出了 (U,V) 服从不同分布时模拟计算的结果.

表 3.2 不同 $g(u,v)$ 时 $\mathsf{E}Y^2$ 的估计 (真实值 $1.333\dot{3}$)

		$g(u,v)=\dfrac{1}{2v}$, $0<u<v<2$	$g(u,v)=\dfrac{1}{4v}$, $0<u<v<4$	$g(u,v)=\dfrac{1}{v}e^{-v}$, $0<u<v<\infty$	$g(u,v)=\dfrac{1}{2v}e^{-1/(2v)}$, $0<u<v<\infty$
$n=100$	均值	1.3111	1.3259	1.3281	1.3446
	标准差	0.1563	0.2673	0.2670	0.2372
$n=200$	均值	1.1956	1.2671	1.3493	1.3383
	标准差	0.0969	0.1842	0.2015	0.1692
$n=400$	均值	1.2163	1.3210	1.3516	1.3293
	标准差	0.0695	0.1413	0.1501	0.1399
$n=800$	均值	1.3132	1.3299	1.3574	1.3258
	标准差	0.0519	0.0913	0.1045	0.1008
$n=2000$	均值	1.2873	1.3007	1.3334	1.3220
	标准差	0.0326	0.0558	0.0707	0.0684

参 考 文 献

[1] Buckley, J. and James, I.R., Linear regression with censored data, *Biometrika*, **66**(1979), 429–436.

[2] Huang, J. and Wellner, J.A., Interval censored survival data: a review of recent progress, *Proceeding of the First Seattle Symposium in Biostatistics: Survival Analysis* (D. Lin and T. Fleming, eds), 123–170, Springer, New York, 1997.

[3] James J.R. and Smith, P.J., Consistency results for linear regression with censored data, *the Annals of Statistics*, **12**(1984), 590–600.

[4] Koul, H., Susarla, V. and Van Ryzin, J., Regression analysis with randomly right-censored data, *Ann. Statist.*, **9**(1981), 1276–1288.

[5] Zheng, Z.K., Regression analysis with censored data, Ph.D. Dissertation, Columbia University, U.S.A., 1984

[6] Zheng, Z.K., A class of estimators of the mean survival time from interval censored data with application to linear regression, manuscript, 2003.

The Estimation of The rth Original Mement of Interval Censored Data

DENG WENLI

(*School of Mathematics and Information Science, Jiangxi Normal University, Nanchang, 330027*)

ZHENG ZUKANG

(*Department of Statistics, School of Management, Fudan University, Shanghai, 200433*)

In reliability and survival analysis, the data obtained in modelling failure times are often interval censored. When interval censoring arises, almost all traditional statistical methods cannot work any more.

In this paper unbiased transformation method will be used to estimate the rth original mement of interval censored data. When the distribution of the censoring variables is specified, a class of estimators having strong consistency (with the rate of $O(n^{-1/2}(\log\log n)^{1/2})$) and asymptotic normality have been got.