第 34 卷第 3 期 南京师大学报(自然科学版) Vol. 34 No. 3
2011 年 9 月 JOURNAL OF NANJING NORMAL UNIVERSITY(Natural Science Edition) Sept, 2011

逐步Ⅰ型区间删失数据下的参数估计

任 瑞,周秀轻

(南京师范大学数学科学学院,江苏 南京 210046)

[摘要] 针对逐步Ⅰ型区间删失数据,用 EM 算法确定参数的极大似然估计,根据缺失信息原则计算出了 MLE 的渐近协方差,并在该数据下证明了 MLE 的相合性和渐近正态性.

[关键词] 逐步Ⅰ删失数据,EM 算法,缺失信息原则,相合性,渐近正态性

[中图分类号] O212.3 [文献标志码] A [文章编号] 1001-4616(2011)03-0007-06

Parameter Estimations Under Progressive Type-Ⅰ Interval Censoring

Ren Rui, Zhou Xiuqing

(School of Mathematical Sciences, Nanjing Normal University, Nanjing 210046, China)

Abstract: EM algorithm are used to obtain the maximum likelihood estimates of parameters when the data are progressively type-Ⅰ interval censoring. The consistency and asymptotic of the MLE are proved and asymptotic covariances are computed by means of the missing information principle.

Key words: progressively type-Ⅰ interval censoring, EM algorithm, missing information principle, consistency, asymptotic normality

在工业寿命检验和医疗生存分析中,通常有个体在失效或死亡之前丢失、撤离或者仅知道个体的寿命在某一区间内的情况,这样得到的样本为删失样本. 最常见的删失有Ⅰ型删失、Ⅱ型删失、逐步删失等. 有许多学者曾研究过删失样本的统计问题,Ng,Chan[1] 用 EM 算法估计逐步删失样本下的参数,并给出了 MLE 的渐近方差和协方差; Balakrishan[2] 讨论了在逐步Ⅱ型删失数据下最大似然估计的渐近性质,基于逐步Ⅱ型删失样本证明了 MLE 的相合性和渐近正态性.

逐步Ⅰ型区间删失是一种更复杂的删失方式,假定取 n 个产品进行寿命试验,它们的真实寿命为 $X_i(i=1,2,\cdots,n)$ 是独立同分布的,密度函数为 $f(x;\theta)$,分布函数为 $F(x;\theta)$,试验在 $t_0=0$ 时开始,在预先指定的 m 个时刻 $t_1<t_2<\cdots<t_m$ 时进行观察,在第 i 个观察时刻 t_i 时发现有 d_i 个个体在 $(t_{i-1},t_i]$ 内失效,同时有 $R_i(i=1,2,\cdots,m)$ 个个体随机地从试验中删失掉. 由于在 t_i 时仍存活的个体数 S_i 是随机变量,且应有 $R_i<S_i$,为此我们可以指定 $R_i=[p_is_i](i=1,2,\cdots,m)$,其中 $p_1,p_2,\cdots,p_{m-1},p_m$ 是预先给定的常数,满足

$$0<p_i<1, i=1,2,\cdots,m-1, p_m=1.$$

我们可以观察到逐步Ⅰ型区间删失数据 $\{(d_i,R_i,t_i),(i=1,2,\cdots,m)\}$,其中 d_i 和 R_i 满足 $a_j-\sum_{j=1}^{i-1}(d_i+R_i)=n$. 记 $\boldsymbol{d}=(d_1,d_2,\cdots,d_m)$,$\boldsymbol{R}=(R_1,R_2,\cdots,R_m)$,$\boldsymbol{T}=(t_1,t_2,\cdots,t_m)$,则我们有似然函数[3]

$$L(\theta;\boldsymbol{d},\boldsymbol{R},\boldsymbol{T})=c\prod_{i=1}^{m}[F(t_i)-F(t_{i-1})]^{d_i}[1-F(t_i)]^{R_i}. \tag{1}$$

Chen[4] 介绍了在逐步Ⅰ型区间删失下广义指数分布的参数的最大似然估计法、矩估计法和概率点法,且通过做模拟求均方误差和偏差比较了这些估计方法,但是没有给出这些估计的性质. 本文主要基于逐步Ⅰ型区间删失样本进行研究,在不要求总体的具体分布形式的条件下,首先用 EM 算法给出了未知参数的估计,并根据缺失信息原则计算出 MLE 的渐进协方差; 然后证明了在单参数情况下 MLE 的相合性和渐进正

收稿日期: 2010-10-11.

通讯联系人: 周秀轻,博士,副教授,研究方向: 生存分析. E-mail: zhouxiuqing@njnu.edu.cn

态性.

1　EM 算法及 MLE 的渐进协方差

逐步Ⅰ型区间删失模型的问题也是不完全数据问题的一种,适合用EM算法获得参数的极大似然估计. 我们把 $Y_{ij}(j=1,2\cdots,d_i)$ 记为区间$(t_{i-1},t_i]$ 内失效个体的寿命,$Z_{ik}(k=1,2\cdots,R_i)$ 记为在 t_i 时删失掉的个体的寿命,并记 $\boldsymbol{Y}_i=(Y_{i1},Y_{i2},\cdots,Y_{id_i})$,$\boldsymbol{Z}_i=(Z_{i1},Z_{i2},\cdots,Z_{iR_i})$,$\boldsymbol{Y}=\{\boldsymbol{Y}_1,\boldsymbol{Y}_2,\cdots,\boldsymbol{Y}_m\}$,$\boldsymbol{Z}=\{\boldsymbol{Z}_1,\boldsymbol{Z}_2,\cdots,\boldsymbol{Z}_m\}(i=1,2,\cdots,m)$. 则 $\boldsymbol{Y}$ 和 $\boldsymbol{Z}$ 的真实值不能被观察到,可认为是缺失数据. 完全数据 X 对应的似然函数为

$$f_X(x;\theta)=\prod_{i=1}^{n}f(x_i;\theta),\tag{2}$$

然而算法的 E 步需要计算 $E[f_X(x,\theta_{h+1}\mid\theta_h)]$,这就需要计算 $\boldsymbol{Y}$ 和 $\boldsymbol{Z}$ 在逐步Ⅰ型区间删失条件下的期望. 因此为了实施EM算法,需要确定在逐步Ⅰ型区间删失的情况下 $\boldsymbol{Y}$ 和 $\boldsymbol{Z}$ 的条件期望及参数的初始值.

定理1　在逐步Ⅰ型区间删失条件下,Y_{ij},Z_{ik} 的条件密度分别为:

$$f(y_{ij}\mid\theta,t_{i-1}<y_{ij}<t_i)=\frac{f(y_{ij};\theta)}{F(t_i)-F(t_{i-1})},\tag{3}$$

$$f(z_{ik}\mid\theta,t_i<z_{ik})=\frac{f(z_{ik};\theta)}{1-F(t_i)},\tag{4}$$

且$(\boldsymbol{Y},\boldsymbol{Z})$ 的各分量是相互独立的.

证明　完全样本被观察到的联合概率密度函数为: $\prod_{i=1}^{m}[\prod_{j=1}^{d_i}f(y_{ij};\theta)\prod_{k=1}^{R_i}f(z_{ik};\theta)]$.

逐步Ⅰ型区间删失数据被观察到的概率为: $\prod_{i=1}^{m}\{[F(t_i)-F(t_{i-1})]^{d_i}[1-F(t_i)]^{R_i}\}$.

则在逐步Ⅰ型区间删失条件下 $\boldsymbol{Y}$ 和 $\boldsymbol{Z}$ 的条件联合概率密度为:

$$\prod_{i=1}^{m}\left\{\prod_{j=1}^{d_i}\left[\frac{f(y_{ij};\theta)}{F(t_i)-F(t_{i-1})}\right]\prod_{k=1}^{R_i}\left[\frac{f(z_{ik};\theta)}{1-F(t_i)}\right]\right\}.$$

再利用分解定理,我们的定理得证.

由上述定理我们可以求得 $\boldsymbol{Y}$ 和 $\boldsymbol{Z}$ 的条件期望. 进行 M 步时,把在EM算法的第 $h+1$ 次迭代中使 $f_X(x,\theta\mid\theta_h)$ 达到最大的 θ 值记为 θ_{h+1}. θ 的MLE估计可通过重复计算 E 步和 M 步直到 θ_h 收敛得到. 初始值 $\theta^{(0)}$ 通过假设 $y_{ij}=t_{i-1}(j=1,2,\cdots,d_i)$,$z_{ik}=t_i(k=1,2,\cdots,R_i)$ 求得.

当样本容量 n 充分大时,记 $a_1=F(t_1)$,$b_1=(1-a_1)p_1$,$a_i=(1-\sum_{j=1}^{i-1}a_j-\sum_{j=1}^{i-1}b_j)\frac{F(t_i)-F(t_{i-1})}{1-F(t_{i-1})}$,$b_i=(1-\sum_{j=1}^{i}a_j-\sum_{j=1}^{i-1}b_j)p_i(i=2,3,\cdots,m)$. 则易证 $\sum_{i=1}^{m}(a_i+b_i)=1$,且 $\frac{d_i}{n}\xrightarrow{p}a_i$,$\frac{R_i}{n}\xrightarrow{p}b_i$,其中"$\xrightarrow{p}$"表示依概率收敛,$a_i$,$b_i$ 为确定的正常数.

用EM算法求解不完全数据问题中的极大似然估计时,Louis[5] 给出了一提取观察信息矩阵的程序. 即观察信息 = 完全信息 − 缺失信息,这一程序称为缺失信息准则. 我们使用这一准则计算在逐步Ⅰ型区间删失情形下MLE的渐进协方差.

首先完全信息为

$$I_X(\theta)=-E\left[\frac{\partial^2 f_X(x;\theta)}{\partial\theta^2}\right],\tag{5}$$

由定理1,类似于[1]中的解法我们容易求得缺失信息为

$$I_{Y,Z}(\theta)=-\sum_{i=1}^{m}\left\{na_iE\left[\frac{\partial^2 f_{y_i}(y_{i1};\theta)}{\partial\theta^2}\right]+nb_iE\left[\frac{\partial^2 f_{z_i}(z_{i1};\theta)}{\partial\theta^2}\right]\right\},\tag{6}$$

因此观察信息为

$$I(\theta)=I_X(\theta)-I_{Y,Z}(\theta),\tag{7}$$

θ 的MLE的渐进协方差为 $I^{-1}(\hat\theta)$.

2　极大似然估计的相合性与渐近正态性

极大似然估计的相合性与渐进正态性的证明与[2]有相同的思路,假设 θ 是单参数. 下面所涉及到的 I 都为一非退化区间,先给定一些条件:

A1: 对几乎所有 x, $\frac{\partial^i \ln f(x;\theta)}{\partial\theta^i}$ 存在且 $\left|\frac{\partial^i \ln f(x;\theta)}{\partial\theta^i}\right| \leqslant G_i^*(x)$,其中 $G_i^*(x)$ 满足对某常数 K 和 δ 及任一 $\theta \in I$ 有 $\int |G_i^*(x)|^{1+\delta} f(x;\theta)\,\mathrm{d}x \leqslant K, i = 1,2,3.$

A2: 对任一 $\theta \in I$,有 $\left|\frac{\partial^i f(x;\theta)}{\partial\theta^i}\right| \leqslant G_i(x)$,其中 $\int G_i(x)\,\mathrm{d}x < \infty, i = 1,2,3.$

A3: 对任一 $\theta \in I$,有 $\left|\frac{1}{1-F(x;\theta)}\right| \leqslant \eta(x)$, $\left|\frac{1}{F(t_i;\theta)-F(t_{i-1};\theta)}\right| \leqslant \eta_i(x)$, $i = 1,2,\cdots,m$. 其中 $\int\eta(x)f(x;\theta)\,\mathrm{d}x < M$, $\int\eta_i(x)f(x;\theta)\,\mathrm{d}x < M_i$, M 和 M_i 为正常数.

A4: 对任一 $\theta \in I$,有 $\gamma^2 = \int\left[\frac{\partial}{\partial\theta}\ln f(x;\theta)\right]^2 f(x;\theta)\,\mathrm{d}x, 0 < \gamma^2 < +\infty.$

注　方差已知,均值未知的正态分布满足以上条件.

为证明极大似然估计的相合性,还需以下引理.

引理1　假设条件 A2,A3 成立,则存在一可积函数 $T(x)$ $\left(\int T(x)\,\mathrm{d}x < \infty\right)$ 和一正常数 Q(与 θ 无关),有 $\frac{1}{n}\left|\frac{\partial^3}{\partial\theta^3}L(\theta;\boldsymbol{d},\boldsymbol{R},\boldsymbol{T})\right| \leqslant T(x)$, $E[T(x)] \leqslant Q$.

证明　类似于[2]中引理1的证明过程,由(1)式我们有

$$\frac{1}{n}\left|\frac{\partial^3}{\partial\theta^3}\ln L(\theta;\boldsymbol{d},\boldsymbol{R},\boldsymbol{T})\right| \leqslant \frac{1}{n}\sum_{i=1}^m\left\{d_i\left|\frac{\partial^3}{\partial\theta^3}\ln[F(t_i;\theta)-F(t_{i-1};\theta)]\right| + R_i\left|\frac{\partial^3}{\partial\theta^3}\ln[1-F(t_i;\theta)]\right|\right\}.$$

在条件 A3 下,

$$\begin{aligned}\left|\frac{\partial^3}{\partial\theta^3}\ln[F(t_i;\theta)-F(t_{i-1};\theta)]\right| = &\left|\frac{\frac{\partial^3}{\partial\theta^3}[F(t_i;\theta)-F(t_{i-1};\theta)]}{F(t_i;\theta)-F(t_{i-1};\theta)} - \right.\\ &\left. 3\frac{\frac{\partial}{\partial\theta}[F(t_i;\theta)-F(t_{i-1};\theta)]\frac{\partial^2}{\partial\theta^2}[F(t_i;\theta)-F(t_{i-1};\theta)]}{(F(t_i;\theta)-F(t_{i-1};\theta))^2} + 2\frac{\left(\frac{\partial}{\partial\theta}[F(t_i;\theta)-F(t_{i-1};\theta)]\right)^3}{(F(t_i;\theta)-F(t_{i-1};\theta))^3}\right| \leqslant\\ &\left[\left|\frac{\partial^3}{\partial\theta^3}F(t_i;\theta)\right| + \left|\frac{\partial^3}{\partial\theta^3}F(t_{i-1};\theta)\right|\right]\eta_i(x) + \\ &3\left[\left|\frac{\partial}{\partial\theta}F(t_i;\theta)\right| + \left|\frac{\partial}{\partial\theta}F(t_{i-1};\theta)\right|\right]\left[\left|\frac{\partial^2}{\partial\theta^2}F(t_i;\theta)\right| + \left|\frac{\partial^2}{\partial\theta^2}F(t_{i-1};\theta)\right|\right]\eta_i^2(x) + \\ &2\left[\left|\frac{\partial}{\partial\theta}F(t_i;\theta)\right| + \left|\frac{\partial}{\partial\theta}F(t_{i-1};\theta)\right|\right]^3\eta_i^3(x).\end{aligned} \tag{8}$$

根据条件 A2 和拉格朗日控制收敛定理,可知

$$\frac{\partial^i}{\partial\theta^i}F(t_i;\theta) = \frac{\partial^i}{\partial\theta^i}\int_{-\infty}^{x_i}f(x;\theta)\,\mathrm{d}x = \int_{-\infty}^{x_i}\frac{\partial^i}{\partial\theta^i}f(x;\theta)\,\mathrm{d}x, i = 1,2,3.$$

因此式(8)中不等式右边的项可以转化为

$$\begin{aligned}&\left[\left|\int_{-\infty}^{t_i}\frac{\partial^3}{\partial\theta^3}f(x)\,\mathrm{d}x\right| + \left|\int_{-\infty}^{t_{i-1}}\frac{\partial^3}{\partial\theta^3}f(x)\,\mathrm{d}x\right|\right]\eta_i(x) + \\ &3\left[\left|\int_{-\infty}^{t_i}\frac{\partial}{\partial\theta}f(x)\,\mathrm{d}x\right| + \left|\int_{-\infty}^{t_{i-1}}\frac{\partial}{\partial\theta}f(x)\,\mathrm{d}x\right|\right]\left[\left|\int_{-\infty}^{t_i}\frac{\partial^2}{\partial\theta^2}f(x)\,\mathrm{d}x\right| + \left|\int_{-\infty}^{t_{i-1}}\frac{\partial^2}{\partial\theta^2}f(x)\,\mathrm{d}x\right|\right]\eta_i^2(x) + \\ &2\left[\left|\int_{-\infty}^{t_i}\frac{\partial}{\partial\theta}f(x)\,\mathrm{d}x\right| + \left|\int_{-\infty}^{t_{i-1}}\frac{\partial}{\partial\theta}f(x)\,\mathrm{d}x\right|\right]^3\eta_i^3(x) \leqslant\end{aligned}$$

$$2\int G_3(x)\,\mathrm{d}x\eta_i(x) + 12\int G_1(x)\,\mathrm{d}x\int G_2(x)\,\mathrm{d}x\eta_i^2(x) + 2\int G_1(x)\,\mathrm{d}x^3\eta_i^3(x).$$

记

$$\varphi(x) = 2\int G_3(x)\,\mathrm{d}x\eta_i(x) + 12\int G_1(x)\,\mathrm{d}x\int G_2(x)\,\mathrm{d}x\eta_i^2(x) + 2\int G_1(x)\,\mathrm{d}x^3\eta_i^3(x),$$

则

$$\left|\frac{\partial^3}{\partial\theta^3}\ln[F(t_i;\theta) - F(t_{i-1};\theta)]\right| \leqslant \varphi(x).$$

再由条件A2,A3,可知 $E[\varphi(x)] \leqslant M^*$ ($M^* > 0$ 且与 θ 无关). 同理,我们可以证得存在 $\omega(x)$,满足

$$\left|\frac{\partial^3}{\partial\theta^3}\ln[1 - F(t_i;\theta)]\right| \leqslant \omega(x).$$

$E[\omega(x)] \leqslant N^*$ ($N^* > 0$ 且与 θ 无关). 令 $\boldsymbol{T}(x) = \frac{1}{n}\sum_{i=1}^{m}[d_i\varphi(x) + R_i\omega(x)]$,引理得证.

定理2　如果条件A1 ~ A4满足,那么若似然方程

$$\frac{\partial}{\partial\theta}\ln L(\theta;\boldsymbol{d},\boldsymbol{R},\boldsymbol{T}) = \sum_{i=1}^{m}\left\{d_i\frac{\partial}{\partial\theta}\ln[F(t_i) - F(t_{i-1})] + R_i\frac{\partial}{\partial\theta}\ln[1 - F(t_i)]\right\} = 0 \tag{9}$$

有解 θ_n,则 $\theta_n \to \theta_0$,其中 θ_0 为真实参数.

证明　由泰勒展式和引理1,可知

$$\frac{1}{n}\frac{\partial}{\partial\theta}\ln L(\theta;\boldsymbol{d},\boldsymbol{R},\boldsymbol{T}) = \frac{1}{n}\frac{\partial}{\partial\theta}\ln L(\theta;\boldsymbol{d},\boldsymbol{R},\boldsymbol{T})\Big|_{\theta=\theta_0} + \frac{(\theta-\theta_0)}{n}\frac{\partial^2}{\partial\theta^2}\ln L(\theta;\boldsymbol{d},\boldsymbol{R},\boldsymbol{T})\Big|_{\theta=\theta_0} +$$
$$\frac{1}{2}\Delta(\theta-\theta_0)^2\boldsymbol{T}(x) = B_0 + B_1(\theta-\theta_0) + \frac{1}{2}\Delta(\theta-\theta_0)^2B_2, \tag{10}$$

其中 $|\Delta| < 1$.

下面我们只需证明 $B_0 \to 0$, $B_1 \to -\zeta_1^2$, $B_2 \to \zeta_2$ 即可,其中 ζ_1^2, ζ_2 为常量.

首先考虑 B_0. 在逐步I型区间删失样本下,因为 $\frac{d_i}{n}\xrightarrow{p}a_i$, $\frac{R_i}{n}\xrightarrow{p}b_i$,所以

$$B_0 = \frac{1}{n}\sum_{i=1}^{m}\left\{d_i\frac{\partial}{\partial\theta}\ln[F(t_i) - F(t_{i-1})] + R_i\frac{\partial}{\partial\theta}\ln[1 - F(t_i)]\right\}\Big|_{\theta=\theta_0} \to$$
$$\sum_{i=1}^{m}\left\{a_i\frac{\partial}{\partial\theta}\ln[F(t_i) - F(t_{i-1})] + b_i\frac{\partial}{\partial\theta}\ln[1 - F(t_i)]\right\}\Big|_{\theta=\theta_0}.$$

又由前一节所提到的缺失信息原则及[2]中定理1,可知

$$B_0 = \left\{\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\ln f(x_i;\theta) - \frac{1}{n}\sum_{i=1}^{m}\left[\sum_{j=1}^{d_i}\frac{\partial}{\partial\theta}\ln f(y_{ij};\theta) + \sum_{k=1}^{R_i}\frac{\partial}{\partial\theta}\ln f(z_{ik};\theta)\right]\right\}\Big|_{\theta=\theta_0} \to$$
$$E\left[\frac{\partial}{\partial\theta}f(x_i;\theta)\right] - \sum_{i=1}^{m}\left\{a_iE\left[\frac{\partial}{\partial\theta}\ln f(Y_{i1};\theta)\right] + b_iE\left[\frac{\partial}{\partial\theta}\ln f(Z_{i1};\theta)\right]\right\}\Big|_{\theta=\theta_0},$$

即

$$B_0 \to -\sum_{i=1}^{m}\left\{a_i\frac{\partial}{\partial\theta}\ln[F(t_i) - F(t_{i-1})] + b_i\frac{\partial}{\partial\theta}\ln[1 - F(t_i)]\right\}\Big|_{\theta=\theta_0}.$$

所以 $B_0 \to 0$.

下面证明 $B_1 \to -\zeta_1^2$. 一方面

$$B_1 = \frac{1}{n}\sum_{i=1}^{m}\left\{d_i\frac{\partial^2}{\partial\theta^2}\ln[F(t_i) - F(t_{i-1})] + R_i\frac{\partial^2}{\partial\theta^2}\ln[1 - F(t_i)]\right\}\Big|_{\theta=\theta_0} \to$$
$$\sum_{i=1}^{m}\left\{a_i\frac{\partial^2}{\partial\theta^2}\ln[F(t_i) - F(t_{i-1})] + b_i\frac{\partial^2}{\partial\theta^2}\ln[1 - F(t_i)]\right\}\Big|_{\theta=\theta_0}.$$

记 $\alpha = \sum_{i=1}^{m}\left\{a_i\frac{\partial^2}{\partial\theta^2}\ln[F(t_i) - F(t_{i-1})] + b_i\frac{\partial^2}{\partial\theta^2}\ln[1 - F(t_i)]\right\}\Big|_{\theta=\theta_0}$.

另一方面根据缺失信息原则

$$B_1=\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta^2}\ln f(x_i;\theta)-\frac{1}{n}\sum_{i=1}^{m}\left[\sum_{j=1}^{d_i}\frac{\partial^2}{\partial\theta^2}\ln f(y_{ij};\theta)+\sum_{k=1}^{R_i}\frac{\partial^2}{\partial\theta^2}\ln f(z_{ik};\theta)\right]\right\}\Bigg|_{\theta=\theta_0}\to$$

$$E\left[\frac{\partial^2}{\partial\theta^2}\ln f(\boldsymbol{X};\theta)\right]-\sum_{i=1}^{m}\left\{a_iE\left[\frac{\partial^2}{\partial\theta^2}\ln f(Y_{i1};\theta)\right]+b_iE\left[\frac{\partial^2}{\partial\theta^2}\ln f(Z_{i1};\theta)\right]\right\}\Bigg|_{\theta=\theta_0}.$$

记 $\beta=E\left[\frac{\partial^2}{\partial\theta^2}\ln f(\boldsymbol{X};\theta)\right]-\sum_{i=1}^{m}\left\{a_iE\left[\frac{\partial^2}{\partial\theta^2}\ln f(Y_{i1};\theta)\right]+b_iE\left[\frac{\partial^2}{\partial\theta^2}\ln f(Z_{i1};\theta)\right]\right\}\Bigg|_{\theta=\theta_0}$. 所以应有 $\alpha=\beta$. 由条件 A2 ~ A4 可得

$$\frac{\partial^2}{\partial\theta^2}\ln[F(t_i)-F(t_{i-1})]=\frac{\frac{\partial^2}{\partial\theta^2}[F(t_i)-F(t_{i-1})]}{F(t_i)-F(t_{i-1})}-\left\{E\left[\frac{\partial}{\partial\theta}\ln f(Y_{i1};\theta)\right]\right\}^2,\tag{11}$$

$$E\left\{\frac{\partial^2}{\partial\theta^2}\ln[f(Y_{i1};\theta)]\right\}=\int_{t_{i-1}}^{t_i}\frac{\frac{\partial^2}{\partial\theta^2}f(y;\theta)}{f(y;\theta)}\frac{f(y;\theta)}{F(t_i)-F(t_{i-1})}\mathrm{d}y-\int_{t_{i-1}}^{t_i}\left[\frac{\partial}{\partial\theta}\ln f(y;\theta)\right]^2\frac{f(y;\theta)}{F(t_i)-F(t_{i-1})}\mathrm{d}y=$$

$$\frac{\frac{\partial^2}{\partial\theta^2}[F(t_i)-F(t_{i-1})]}{F(t_i)-F(t_{i-1})}-\int_{t_{i-1}}^{t_i}\left[\frac{\partial}{\partial\theta}\ln f(y;\theta)\right]^2\frac{f(y;\theta)}{F(t_i)-F(t_{i-1})}\mathrm{d}y.\tag{12}$$

从式(11),(12)知

$$\frac{\partial^2}{\partial\theta^2}\ln[F(t_i)-F(t_{i-1})]=E\left\{\frac{\partial^2}{\partial\theta^2}\ln[f(Y_{i1};\theta)]\right\}+\mathrm{var}\left[\frac{\partial}{\partial\theta}\ln f(Y_{i1;\theta})\right].$$

同理有

$$\frac{\partial^2}{\partial\theta^2}\ln[1-F(t_i)]=E\left\{\frac{\partial^2}{\partial\theta^2}\ln[f(Z_{i1};\theta)]\right\}+\mathrm{var}\left[\frac{\partial}{\partial\theta}\ln f(Z_{i1;\theta})\right].$$

那么我们容易得出

$$\alpha=\beta=\frac{1}{2}\left\{E\left[\frac{\partial^2}{\partial\theta^2}\ln f(\boldsymbol{X};\theta)\right]+\sum_{i=1}^{m}\left\{a_i\mathrm{var}\left[\frac{\partial}{\partial\theta}\ln f(Y_{i1;\theta})\right]+b_i\mathrm{var}\left[\frac{\partial}{\partial\theta}\ln f(Z_{i1;\theta})\right]\right\}\right\},$$

所以

$$B_1=\frac{1}{2}\left\{E\left[\frac{\partial^2}{\partial\theta^2}\ln f(\boldsymbol{X};\theta)\right]+\sum_{i=1}^{m}\left\{a_i\mathrm{var}\left[\frac{\partial}{\partial\theta}\ln f(Y_{i1;\theta})\right]+b_i\mathrm{var}\left[\frac{\partial}{\partial\theta}\ln f(Z_{i1;\theta})\right]\right\}\right\}=$$

$$-\frac{1}{2}\left\{\gamma^2-\sum_{i=1}^{m}\left\{a_i\mathrm{var}\left[\frac{\partial}{\partial\theta}\ln f(Y_{i1;\theta})\right]+b_i\mathrm{var}\left[\frac{\partial}{\partial\theta}\ln f(Z_{i1;\theta})\right]\right\}\right\}.$$

从条件 A_3, A_4 可知 $\mathrm{var}\left[\frac{\partial}{\partial\theta}\ln f(y_{i1};\theta)\right]<\gamma^2$, $\mathrm{var}\left[\frac{\partial}{\partial\theta}\ln f(z_{i1};\theta)\right]<\gamma^2$ 是有界的且与 θ 无关，所以

$$\sum_{i=1}^{m}\left\{a_i\mathrm{var}\left[\frac{\partial}{\partial\theta}\ln f(Y_{i1;\theta})\right]+b_i\mathrm{var}\left[\frac{\partial}{\partial\theta}\ln f(Z_{i1;\theta})\right]\right\}\to\Lambda,\tag{13}$$

Λ 为一有界常数. 令 $\zeta_1^2=\frac{1}{2}(\gamma^2-\Lambda)$，则 $B_1\to-\zeta_1^2$.

应用引理 1 可类似证明 $B_2\xrightarrow{p}\zeta_2$，$\zeta_2$ 为一常数. 定理得证.

下面证明 MLE 的渐近正态性.

定理 3 如果条件 A1 ~ A4 满足，且似然方程(9)有解 $\hat{\theta}_n$，则 $\sqrt{n}(\hat{\theta}_n-\theta)\xrightarrow{d}N(0,\gamma^2+\Lambda)$.

证明 由(10)式知

$$\zeta_1\sqrt{n}(\hat{\theta}-\theta_0)=\frac{\frac{\sqrt{n}}{\zeta_1}B_0}{-\frac{B_1}{\zeta_1^2}-\frac{\Delta}{2\zeta_1^2}B_2(\hat{\theta}-\theta_0)}=$$

$$\frac{1}{\zeta_1\sqrt{n}}\sum_{i=1}^{m}\left\{d_i\frac{\partial}{\partial\theta}\ln\left[F(t_i)-F(t_{i-1})\right]+R_i\frac{\partial}{\partial\theta}\ln\left[1-F(t_i)\right]\right\}\Big/-\frac{B_1}{\zeta_1^2}-\frac{\Delta}{2\zeta_1^2}B_2(\hat{\theta}-\theta_0), \tag{14}$$

容易看出式(14)的分母收敛到1,所以我们只需证明分子是渐近服从均值为0的正态分布.

$$\sqrt{n}B_0=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\ln f(x_i;\theta)-\frac{1}{\sqrt{n}}\sum_{i=1}^{m}\left[\sum_{j=1}^{d_i}\frac{\partial}{\partial\theta}\ln f(y_{ij};\theta)+\sum_{k=1}^{R_i}\frac{\partial}{\partial\theta}\ln f(z_{ik};\theta)\right]\Bigg|_{\theta=\theta_0}.$$

首先$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\ln f(x_i;\theta)\xrightarrow{d}N(0,\gamma^2)$[6],又由条件A1,A3知$E\left|\frac{\partial}{\partial\theta}\ln f(Y_{i1};\theta)-E\left[\frac{\partial}{\partial\theta}\ln f(Y_{i1};\theta)\right]\right|^3$,$E\left|\frac{\partial}{\partial\theta}\ln f(Z_{i1};\theta)-E\left[\frac{\partial}{\partial\theta}\ln f(Z_{i1};\theta)\right]\right|^3$是有界的且与$\theta$无关.所以由(13)式、[2]中定理3和定理5可知

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{m}\left[\sum_{j=1}^{d_i}\frac{\partial}{\partial\theta}\ln f(y_{ij};\theta)+\sum_{k=1}^{R_i}\frac{\partial}{\partial\theta}\ln f(z_{ik};\theta)\right]\Bigg|_{\theta=\theta_0}\xrightarrow{d}N(0,\Lambda),$$

再利用Slutsky′s定理,定理得证.

[参考文献]

[1] Ng H K T, Chan P S, Balakrishnan N. Estimation of parameters from progressively censored data using EM algorithm[J]. Computational Statistics and Data Analysis, 2002, 39: 371-386.

[2] Lin Chientai, Balakrishnan N. Asymptotic properties of maximum likelihood estimators based on progressive Type-Ⅱ censoring[J]. Metrika, 2010-03-25. Online First.

[3] Aggarwala R. Progressively interval censoring: Some mathematical results with application to inference[J]. Communications in Statistics-Theory and Methods, 2001, 30: 1 921-1 935.

[4] Chen D G, Lio Y L. Parameter estimations for generalized exponential distribution under progressive type-Ⅰ interval censoring[J]. Computational Statistics and Data Analysis, 2010, 54: 1 581-1 591.

[5] Louis T A. Finding the observed information matrix when using the EM algorithm[J]. R Stat Soc Series B, 1982, 44: 226-233.

[6] Cramér H. Mathematical Methods of Statistics[M]. Princeton: Princeton University Press, 1946.

[责任编辑:丁　蓉]