文章编号:1673-1549(2017)04-0081-06　　DOI:10.11863/j. suse.2017.04.15

EM 算法在基于区间数据的加速寿命试验统计分析中的应用

邱雷鞶

(闽江师范高等专科学校计算机系, 福州 350002)

摘　要:在步进加速寿命试验中,当寿命服从广义指数分布且获得的数据是区间数据时,给出试验安排,并通过"各应力下产品失效机理保持不变"等几个基本模型假定,得出各应力水平下形状参数相等的结论以及不同应力水平下的试验时间 $t_1 - t_0, t_2 - t_1, \cdots, t_{i-1} - t_{i-2}$ 折算到某一应力水平下的时间 τ_{i-1} 的折算公式,进而得出求相关参数的极大似然估计的隐性表达式。进一步讨论利用 EM 算法求解,先由 E 步求出期望,再由 M 步求出使得期望极大化的点,给出具体的迭代过程。最后采用 Monte Carlo 数据模拟方法分别在大样本和小样本场合下给出了参数真值的估计值,结果表明,该方法在大样本场合下更具有效性。

关键词:区间数据;广义指数分布;EM 算法

中图分类号:0212.7　　**文献标志码**:A

引　言

加速寿命试验,是指将样品置于超过正常应力水平下,通过观察样品的失效时间(即寿命),从而利用统计方法推断在正常应力条件下产品的各项可靠性指标的一种寿命试验。目前常见的类型有恒定应力加速寿命试验,步进应力加速寿命试验,序进应力加速寿命试验。文献[1]中讨论了加速寿命试验的类型和理论基础。本文讨论步进应力加速寿命试验,简称步加试验,是先选定一组加速应力水平 $S_1 < \cdots < S_m$,它们都高于正常应力水平,试验开始时把一定数量的样品都置于应力水平 S_1 下进行寿命试验,经过一段时间,把应力提高到 S_2,将未失效的样品在 S_2 下继续进行寿命试验,如此继续下去,直到有一定数量的样品发生失效为止。文献[2-3]对步进加速寿命试验做出了详细的阐述。

在步进加速寿命试验的实际操作过程中通常不能获得完全数据,只能了解寿命数据 X 是否位于某一个时间区间内,此时得到的数据是区间数据。在数据处理中,按时间顺序 $t_0 < t_1 < \cdots < t_n < \infty$ 将时间分成若干段 $[t_0, t_1), [t_1, t_2), \cdots, [t_n, \infty)$,考察产品、系统或组件的寿命长度出现在哪个时段中,或者考虑每个时间段中产品、系统或组件的个数,然后作出统计推断。因此,近年来出现了关于寿命数据为区间数据的加速寿命试验的研究,如 Dempster A P[4] 等人给出了不完全数据中相关参数的极大似然估计,其中,首次提出了 EM 迭代算法,即每一次迭代由两步组成: E 步(求期望)和 M 步(极大化)。记 θ 为未知参数,一般的,以 $P(\theta | Z_1)$ 表示 θ 的基于观测数据 Z_1 的后验分布密度函数,称为观测后

收稿日期:2017-06-06
基金项目:福建省教育厅中青年课题(JAT160827)
作者简介:邱雷鞶(1980-),女,福建石狮人,讲师,硕士,主要从事概率与数理统计方面的研究,(E-mail)qiuleipin@163.com

验分布, $P(\theta|Z_1,Z_2)$ 表示添加数据 Z_2 后得到的关于 θ 的后验分布密度函数, 称为添加后验分布, $P(Z_2|\theta,Z_1)$ 表示在给定 θ 和观测数据 Z_1 下潜在数据 Z_2 的条件分布密度函数。统计分析的目的是计算观测后验分布 $P(\theta|Z_1)$ 的众数。文献[5]详细介绍了该算法并讨论了收敛性。文献[6]对该算法做了推广。文献[7]系统地介绍了 EM 算法。EM 算法被广泛运用到计算极大似然估计,特别是当观测值是不完全数据的时候。文献[8]对 EM 算法做了改进,提出了多重集 EM 算法,可以更好地体现 EM 算法的收敛性。近两年,当产品寿命服从其它简单分布时,区间数据的加速寿命试验的统计分析问题也都有了相应的研究。文献[9]讨论了利用 EM 算法解决威布尔分布下缺失数据似然估计的统计推断,文献[10]列出了 EM 算法的最大似然法被广泛应用于诸如高斯混合模型等隐藏结构模型中参数的估计。

然而,针对 Gupta 和 Kundu 提出的广义指数分布[11],利用 EM 算法对区间数据的统计分析的研究并不多见。文献[12]指出,在分析很多寿命数据时,广义指数分布往往比其它分布更能有效利用。近年来,双参数广义指数分布已经很广泛的运用到分析寿命数据。文献[13]研究了广义指数分布的尺度参数的极大似然估计基于随机截尾模型。文献[14]介绍了形状参数的增加,使得这一分布族产生了更多的灵活性,使它可以被用来分析删失数据,如区间数据。

本文将主要讨论当获得的寿命数据为区间数据,而寿命分布服从广义指数分布时相关参数的统计分析。

1 试验安排及基本模型假定

1.1 试验安排

假设 S_0 为正常条件下的应力水平, $S_1,S_2,\cdots,S_m$ 为加速应力水平,且有 $0<S_0<S_1<\cdots<S_m$, 试验开始时,安排 n 个产品在应力 S_1 下做寿命试验直至时刻 t_1 时为止,经检测知有 n_0 个产品失效,此时将应力水平由 S_1 上升到 S_2, 余下的未失效产品 $n-n_0$ 在应力水平 S_2 下继续做试验,到时刻 t_2 时停止试验,经检测,在 S_2 下有 n_1 个产品失效。如此进行下去,直到应力上升到 S_m, 到时刻 t_m 为止有 n_{m-1} 个产品在 S_m 下失效,记 $N=\sum_{i=0}^{m-1}n_i$, 而 n_m 为尚未失效的产品个数,则 $n=\sum_{i=0}^{m}n_i$, 试验结束。

1.2 基本假定

假定 1 在应力水平 $S_i(i=1,2,\cdots,m)$ 下,产品的寿命分布为广义指数分布,分布函数为 $F_i(t)=(1-e^{-\lambda_i t})^{\alpha_i}$, 密度函数为 $f_i(t)=\alpha_i\lambda_i(1-e^{-\lambda_i t})^{\alpha_i-1}\cdot e^{-\lambda_i t}$, 其中 λ_i 和 α_i 分别是尺度参数的倒数和形状参数,和应力水平 S_i 有关。

假定 2 在应力水平 $S_i(i=1,2,\cdots,m)$ 下,产品的失效机理保持不变[1]。

假定 3 产品在应力水平 S 下,满足线性加速模型 $\ln\lambda_i=a+b\varphi(S_i)$, 其中 $\varphi(S_i)$ 是与应力水平 S_i 有关的已知函数,下记 φ_i, $i=1,2,\cdots,m$。

假定 4 产品的剩余寿命仅依赖于当时已累积失效的部分和当时的应力水平,而与累积的方式无关[15]。

2 参数估计

引理 1 在应力水平 $S_i(i=1,2,\cdots,m)$ 下,产品的寿命分布为 $F_i(t)=(1-e^{-\lambda_i t})^{\alpha_i}$, 则 $\alpha_i\equiv\alpha(i=1,2,\cdots,m)$。

证明 设 $R(t)$ 为可靠度函数,即寿命超过时间 t 的产品在总体中所占的概率,则它的可靠寿命 t_R 满足方程 $1-F(t_R)=R$, 由基本假定,产品寿命服从广义指数分布,因此 $F(t_R)=(1-e^{-\lambda t_R})^{\alpha}$, 解得 $t_R=-\frac{1}{\lambda}\ln[1-(1-R)^{\frac{1}{\alpha}}]$。从而, S_i 对 S_j 的 R 可靠寿命的加速系数为

$$K_{ij}(R)=t_{R,j}/t_{R,i}=\frac{\lambda_i}{\lambda_j}\cdot\frac{\ln[1-(1-R)^{\frac{1}{\alpha_j}}]}{\ln[1-(1-R)^{\frac{1}{\alpha_i}}]}$$

由基本假定 2,在应力水平 $S_i(i=1,2,\cdots,m)$ 下,产品的失效机理保持不变,即加速系数是与 R 无关的常数,因此要求 $\alpha_i\equiv\alpha\quad(i=1,2,\cdots,m)$。

引理 2 若记 τ_{i-1} 为产品经历了水平 $S_1,S_2,\cdots,S_{i-1}$ 的步加试验后,各段试验延续时间 $t_1-t_0,t_2-t_1,\cdots,t_{i-1}-t_{i-2}$ 折算到 S_i 应力水平下总的折算时间,则

$$\begin{cases}\tau_0=0\\ \tau_{i-1}=\sum_{j=1}^{i-1}e^{b(\varphi_j-\varphi_i)}(t_j-t_{j-1}),i=2,3,\cdots,m\end{cases}$$

证明 由基本假定 4 知 τ_{i-1} 满足

$$\begin{cases}\tau_0=0\\ F_i(\tau_{i-1})=F_{i-1}(t_{i-1}-t_{i-2}+\tau_{i-2}),i=2,3,\cdots,m\end{cases}$$

又由假定 1 和假定 2 得:

$$(1-e^{-\lambda_i\tau_{i-1}})^{\alpha}=(1-e^{-\lambda_{i-1}(t_{i-1}-t_{i-2}+\tau_{i-2})})^{\alpha}$$

得到,

$$e^{-\lambda_i\tau_{i-1}}=e^{-\lambda_{i-1}(t_{i-1}-t_{i-2}+\tau_{i-2})}$$

由此得到递推公式

$$\tau_{i-1}=\frac{\lambda_{i-1}}{\lambda_i}(t_{i-1}-t_{i-2})+\frac{\lambda_{i-1}}{\lambda_i}\tau_{i-2},i=2,3,\cdots m$$

又由 $\tau_0=0$ 得

$$\tau_{i-1}=\frac{\lambda_{i-1}}{\lambda_i}(t_{i-1}-t_{i-2})+\frac{\lambda_{i-1}}{\lambda_i}\cdot\frac{\lambda_{i-2}}{\lambda_{i-1}}(t_{i-2}-t_{i-3})+\cdots+\frac{\lambda_2}{\lambda_i}\cdot\frac{\lambda_1}{\lambda_2}(t_1-t_0)=\frac{1}{\lambda_i}\sum_{j=1}^{i-1}\lambda_j(t_j-t_{j-1})$$

其中 ,$i=2,3,\cdots,m$。由基本假定知,

$$\begin{cases}\tau_0=0\\ \tau_{i-1}=\sum_{j=1}^{i-1}e^{b(\varphi_j-\varphi_i)}(t_j-t_{j-1}),i=2,3,\cdots,m\end{cases}$$

所以,τ_{i-1} 为 b 的函数,下记 $\tau_{i-1}(b)$,$i=1,2,\cdots,m$。

估计 a,b,α 三个参数。记 n 个样品的失效时间为 $X_k,k=1,2,\cdots,n$,它们是独立同分布的随机变量,密度函数为 $f(x)=\alpha\lambda(1-e^{-\lambda x})^{\alpha-1}\cdot e^{-\lambda x}$,$x>0,\lambda>0,\alpha>0$。只能观测到落在区间 $[t_i,t_{i+1})$ 中的样本数 n_i,其中 $i=0,1,2,\cdots,m$,$0=t_0<t_1<t_2<\cdots<t_m<t_{m+1}=\infty$。$x$ 落在区间 $[t_i,t_{i+1})$ 中的概率为

$$P(x\in[t_i,t_{i+1}))=F_{i+1}(t_{i+1})-F_i(t_i)=[1-e^{-\lambda_{i+1}(t_{i+1}-t_i+\tau_i(b))}]^{\alpha}-[1-e^{-\lambda_i(t_i-t_{i-1}+\tau_{i-1}(b))}]^{\alpha}$$

其中记 $t_{-1}=\tau_{-1}(b)=0,\lambda_0=\lambda_{m+1}=0$。

定理 1　参数 a,b,α 的极大似然估计可以由隐性表达式求解。

$$\sum_{i=0}^{m}\frac{n_iH_{i+1}^{\alpha}\cdot\ln H_{i+1}}{H_{i+1}^{\alpha}-H_i^{\alpha}}-\sum_{i=0}^{m}\frac{n_iH_i^{\alpha}\cdot\ln H_i}{H_{i+1}^{\alpha}-H_i^{\alpha}}=0\tag{1}$$

$$\sum_{i=0}^{m}\frac{n_iH_{i+1}^{\alpha-1}K_{i+1}\lambda_{i+1}(t_{i+1}-t_i+\tau_i(b))}{H_{i+1}^{\alpha}-H_i^{\alpha}}-\sum_{i=0}^{m}\frac{n_iH_i^{\alpha-1}K_i\lambda_i(t_i-t_{i-1}+\tau_{i-1}(b))}{H_{i+1}^{\alpha}-H_i^{\alpha}}=0\tag{2}$$

$$\sum_{i=0}^{m}\frac{n_iH_{i+1}^{\alpha-1}K_{i+1}\lambda_{i+1}\varphi_{i+1}M_{i+1}}{H_{i+1}^{\alpha}-H_i^{\alpha}}-\sum_{i=0}^{m}\frac{n_iH_i^{\alpha-1}K_i\lambda_i\varphi_iM_i}{H_{i+1}^{\alpha}-H_i^{\alpha}}=0\tag{3}$$

其中,

$$H_i\triangleq H_i(a,b,\alpha)=1-e^{-\lambda_i(t_i-t_{i-1}+\tau_{i-1}(b))}$$

$$K_i\triangleq K_i(a,b)=e^{-\lambda_i(t_i-t_{i-1}+\tau_{i-1}(b))}$$

$$M_i\triangleq M_i(b)=t_i-t_{i-1}+\tau_{i-1}(b)+\sum_{j=1}^{i-1}e^{b(\varphi_j-\varphi_i)}(t_j-t_{j-1})$$

证明　容易得到似然函数为

$$L=\prod_{i=0}^{m}[[1-e^{-\lambda_{i+1}(t_{i+1}-t_i+\tau_i(b))}]^{\alpha}-[1-e^{-\lambda_i(t_i-t_{i-1}+\tau_{i-1}(b))}]^{\alpha}]^{n_i}$$

对数似然函数为

$$\ln L=\sum_{i=0}^{m}n_i\ln[[1-e^{-\lambda_{i+1}(t_{i+1}-t_i+\tau_i(b))}]^{\alpha}-[1-e^{-\lambda_i(t_i-t_{i-1}+\tau_{i-1}(b))}]^{\alpha}]$$

利用时间折算公式,对数似然函数分别对 a,b,α 求偏导,并令 $\frac{\partial\ln L}{\partial a}=0,\frac{\partial\ln L}{\partial b}=0,\frac{\partial\ln L}{\partial\alpha}=0$,得到式(1)、式(2)和式(3)。

3　*EM* 算法求极大似然估计

式(1)、式(2)和式(3)无法得出 a,b,α 的显性表达式,于是考虑用 *EM* 算法实现。

为了便于表述,n 个样品的失效时间 $X_k,k=1,2,\cdots,n$ 的全体记为 X,观测结果即落在区间 $[t_i,t_{i+1})$ 中的样本数目 n_i,$i=0,1,2,\cdots,m$,$0=t_0<t_1<t_2<\cdots<t_m<t_{m+1}=\infty$ 记为 Y。记 X_{ih} 为落入区间 $[t_{i-1},t_i)$ 的随机变量。

E 步:$P(a,b,\alpha|Y)$,$P(a,b,\alpha|Y,X)$ 分别表示关于参数 a,b,α 的观测后验分布与添加后验分布。事实上,X 实际上已经包含了 Y 所有的信息,所以有 $P(a,b,\alpha|Y,X)=P(a,b,\alpha|X)$,由广义指数分布的密度函数可以得到

$$\ln P(a,b,\sigma|X)=n\ln\alpha+\sum_{i=1}^{m}n_{i-1}(a+b\varphi_i)-\sum_{i=1}^{m}n_{i-1}e^{a+b\varphi_i}(x_{ih}-t_{i-1}+\tau_{i-1})+(\alpha-1)\sum_{i=1}^{m}n_{i-1}\ln\{1-\exp[-e^{a+b\varphi_i}(x_{ih}-t_{i-1}+\tau_{i-1})]\}$$

于是

$$Q(a,b,\alpha|a^{(k)},b^{(k)},\alpha^{(k)},Y)=E[\ln P(a,b,\alpha|X)|a^{(k)},b^{(k)},\alpha^{(k)},Y]=n(\ln\alpha+a)+b\sum_{i=1}^{m}n_{i-1}\varphi_i-\sum_{i=1}^{m}n_{i-1}E[e^{a+b\varphi_i}\cdot$$

$$(x_{ih}-t_{i-1}+\tau_{i-1}(b)))\,|\,a^{(k)},b^{(k)},\alpha^{(k)},Y]+$$

$$(\alpha-1)\sum_{i=1}^{m}n_{i-1}E[\ln(1-\exp(-e^{a+b\varphi_i}\cdot$$

$$(x_{ih}-t_{i-1}+\tau_{i-1}(b))))\,|\,a^{(k)},b^{(k)},\alpha^{(k)},Y]$$

记 X_{ih} 的条件密度函数为 $P_i^{(k)}(t)$, $t\in[t_{i-1},t_i)$, $i=1,2,\cdots,m$, 则有,

$$P_i^{(k)}(t)\triangleq f(t\mid a^{(k)},b^{(k)},\alpha^{(k)},Y)=$$

$$\frac{1}{C_i^{(k)}}\cdot\alpha^{(k)}\cdot e^{a^{(k)}+b^{(k)}\varphi_i}\cdot$$

$$[1-\exp[-e^{a^{(k)}+b^{(k)}\varphi_i}(t-t_{i-1}+\tau_{i-1}(b^{(k)}))]]^{\alpha^{(k)}-1}\cdot$$

$$\exp[-e^{a^{(k)}+b^{(k)}\varphi_i}(t-t_{i-1}+\tau_{i-1}(b^{(k)}))]$$

其中,

$$C_i^{(k)}\triangleq C_i(a^{(k)},b^{(k)},\alpha^{(k)})=$$

$$[1-\exp[-e^{a^{(k)}+b^{(k)}\varphi_i}(t_i-t_{i-1}+\tau_{i-1}(b^{(k)}))]]^{\alpha^{(k)}}-$$

$$[1-\exp[-e^{a^{(k)}+b^{(k)}\varphi_i}\tau_{i-1}(b^{(k)})]]^{\alpha^{(k)}}$$

所以

$$Q(a,b,\alpha\,|\,a^{(k)},b^{(k)},\alpha^{(k)},Y)=n(\ln\alpha+a)+$$

$$\sum_{i=1}^{m}n_{i-1}\Big[b\varphi_i+\int_{t_{i-1}}^{t}[\ln(1-A_i)+(\alpha-1)\ln A_i]P_i^{(k)}(t)dt\Big]$$

其中,

$$A_i\triangleq A_i(a,b,t)\triangleq 1-\exp[-e^{a+b\varphi_i}(t-t_{i-1}+$$

$$\tau_{i-1}(b))]$$

M 步:将期望 $Q(a,b,\alpha\,|\,a^{(k)},b^{(k)},\alpha^{(k)},Y)$ 分别对 a,b,α 求偏导,求出使得 $Q(a,b,\alpha|a^{(k)},b^{(k)},\alpha^{(k)},Y)$ 极大化的点 $(a^{(k+1)},b^{(k+1)},\alpha^{(k+1)})$。

$$\frac{\partial Q}{\partial\alpha}=\frac{n}{\alpha}+\sum_{i=1}^{m}n_{i-1}\int_{t_{i-1}}^{t}(\ln A_i)P_i^{(k)}(t)\,dt$$

令 $\frac{\partial Q}{\partial\alpha}=0$, 得到

$$\alpha=-\frac{n}{\sum_{i=1}^{m}n_{i-1}\int_{t_{i-1}}^{t}(\ln A_i)P_i^{(k)}(t)\,dt}\tag{4}$$

$$\frac{\partial Q}{\partial a}=n+\sum_{i=1}^{m}n_{i-1}\int_{t_{i-1}}^{t}\frac{1+\alpha A_i-\alpha}{A_i}$$

$$(\ln(1-A_i))P_i^{(k)}(t)\,dt$$

令 $\frac{\partial Q}{\partial a}=0$, 可得,

$$a=$$

$$\ln\Big[n+(\alpha-1)\sum_{i=1}^{m}n_{i-1}\int_{t_{i-1}}^{t}\frac{1-A_i}{A_i}\ln(1-A_i)P_i^{(k)}(t)dt\Big]-$$

$$\ln\sum_{i=1}^{m}n_{i-1}\int_{t_{i-1}}^{t}e^{b\varphi_i}(t-t_{i-1}+\tau_{i-1}(b))P_i^{(k)}(t)\,dt\tag{5}$$

$$\frac{\partial Q}{\partial b}=$$

$$\sum_{i=1}^{m}n_{i-1}\varphi_i\Big[1+\int_{t_{i-1}}^{t}\frac{1+\alpha A_i-\alpha}{A_i}(\ln(1-A_i))P_i^{(k)}(t)dt\Big]\tag{6}$$

迭代过程:

(1)当 $k=0$ 时,给定参数初始值 $a^{(0)}$, $b^{(0)}$, $\alpha^{(0)}$, 及任意小的正数 ε。

(2)当 $k\geqslant 1$ 时,现有的参数估计为 $a^{(k)}$, $b^{(k)}$, $\alpha^{(k)}$。

①取式(4)右边的, $b=b^{(k)}$, 即可找到 $\alpha^{(k+1)}$,

$$\alpha^{(k+1)}=-\frac{n}{\sum_{i=1}^{m}n_{i-1}\int_{t_{i-1}}^{t}\ln A_i^{(k)}P_i^{(k)}(t)\,dt}$$

②取式(5)右边的 $b=b^{(k)}$, $a=a^{(k)}$, $\alpha=\alpha^{(k)}$, 即可找到 $a^{(k+1)}$,

$$a^{(k+1)}=\ln[n+(\alpha^{(k+1)}-1)$$

$$\sum_{i=1}^{m}n_{i-1}\int_{t_{i-1}}^{t}\frac{1-A_i^{(k)}}{A_i^{(k)}}\ln(1-A_i^{(k)})P_i^{(k)}(t)\,dt\Big]-$$

$$\ln\sum_{i=1}^{m}n_{i-1}\int_{t_{i-1}}^{t}e^{b^{(k)}\varphi_i}(t-t_{i-1}+\tau_{i-1}(b^{(k)}))P_i^{(k)}(t)\,dt$$

③令式(6)为零,即可解出 b 的显示表达式,令表达式中 $a=a^{(k)}$, $b=b^{(k)}$, $\alpha=\alpha^{(k)}$ 找到 $b^{(k+1)}$,

$$b^{(k+1)}=\frac{\sum_{i=1}^{m}n_{i-1}\varphi_i\Big[\int_{t_{i-1}}^{t}\frac{1}{A_i^{(k)}}\ln(1-A_i^{(k)})P_i^{(k)}(t)\,dt+1\Big]}{\sum_{i=1}^{m}n_{i-1}\varphi_i\int_{t_{i-1}}^{t}\frac{\alpha^{(k)}\exp[-e^{a^{(k)}}(t-t_{i-1}+\tau_{i-1}(b^{(k)}))]}{A_i^{(k)}}\ln(1-A_i^{(k)})P_i^{(k)}(t)\,dt}$$

其中,

$$A_i^{(k)}\triangleq A_i^{(k)}(a^{(k)},b^{(k)},t)\triangleq 1-\exp[-e^{a^{(k)}+b^{(k)}\varphi_i}$$

$$(t-t_{i-1}+\tau_{i-1}(b^{(k)}))]$$

(3)重复过程(2),直到 $\|(a^{(k+1)},b^{(k+1)},\alpha^{(k+1)})-(a^{(k)},b^{(k)},\alpha^{(k)})\|<\varepsilon$ 时为止。

满足迭代终止的条件后继续进行 20 次迭代,取所

得到的 20 个参数迭代值的平均值作为对真实参数的估计 $\hat{a}, \hat{b}, \hat{\alpha}$。从而可得产品在正常使用条件下的寿命分布及特征寿命的估计,还可以得到其它的信息,如平均寿命 $ET = e^{\hat{a}+\hat{b}\varphi(S_0)} \cdot [\psi(\hat{\alpha}+1) - \psi(1)]$,其中,$\psi(\alpha) = \frac{d[\log\Gamma(\alpha)]}{d\alpha}$,$\Gamma(\alpha)$ 是 Gamma 函数,进而可以得到可靠度、失效率,以及分位寿命等估计。

4　数据模拟

考虑步进加速寿命试验过程中的应力为温度,其中 $T_1 = 240$, $T_2 = 350$, $T_3 = 470$, $\varphi_i = \varphi(T_i) = 1/T_i$, $t_0 = 0$, $t_1 = 1.20 \times 10^4$, $t_2 = 1.55 \times 10^4$, $t_3 = 1.66 \times 10^4$, 参数真值为 $a = -20$, $b = 10^4$, $\alpha = 2$, 取迭代初始值 $a^{(0)} = -18$, $b^{(0)} = 12000$, $\alpha^{(0)} = 2.5$。

在小样本场合下,取 $n_0 = n_1 = n_2 = 5$, $\varepsilon = 10^{-3}$, 在大样本的场合下,取 $n_0 = n_1 = n_2 = 100$, $\varepsilon = 10^{-3}$ 运用以上迭代过程得到参数真值的估计,见表 1。

表 1　模拟结果

样本量	真值	估计值	相对误差
$n = 15$	$\alpha = 2$	$\hat{\alpha} = 2.1095$	0.1095
	$a = -20$	$\hat{a} = -19.8093$	0.1807
	$b = 10^4$	$\hat{b} = 1.1201 \times 10^4$	0.1201×10^4
$n = 300$	$\alpha = 2$	$\hat{\alpha} = 1.9983$	0.0017
	$a = -20$	$\hat{a} = -20.0961$	0.0961
	$b = 10^4$	$\hat{b} = 1.1035 \times 10^4$	0.1035×10^4

由数据模拟结果表 1 可知,在大样本场合下,估计值与参数真值的相对误差有所减小。因此,在产品寿命服从广义指数分布的情况下,基于获得的数据为区间数据的步进加速寿命试验,在大样本场合下利用 *EM* 算法给出参数的最大似然估计显得更为有效。

参 考 文 献:

[1] YANG Y H,ZHOU Y Q.Theoretical foundation of accelerated life testing[J].Journal of Propulsion and Technology,2001,22(5):353-356.

[2] ZHAO W,ELSAYED E A.A general accelerated life model for step-stress testing[J].IIE Trans,2005,37:1059-1069.

[3] MOHAMED T.Step-Stress Accelerated Life Tests[C]// Miodrag Lovric.International Encyclopedia of Statistical Science,Springer,America:2011:1547-1517.

[4] DEMPSTER A P,LAIRD N M,RUBIN D B.Maximum likelihood from incomplete data via the EM algorithm [J].Math SciNet Google Scholar.JR Stat Soc Ser B, 1977,39:1-22.

[5] WU C F.On the convergence properties of the EM algorithm[J].The Annals of Statistics,1983,11(1):95-103.

[6] AMBROISE C.The EM Algorithm and Extensions[J]. Journal of classification,1988,15(1):154-156.

[7] CHARLES B,PAUL P B.Handbook of Mathematical Methods in Imaging[M].New York: Springer-Verlag, 2015.

[8] HUANG W H,CHEN Y G.The multiset EM algorithm[J]. Statistics & Probability Letters,2017,126:41-48.

[9] SUVRA P,BALAKRISHNAN N.Destructive negative binomial cure rate model and EM-based likelihood inference under Weibull lifetime[J].Statistics & Probability Letters,2016,116:9-20.

[10] JEAN-PATRICK B,GILLES C.EM for mixtures[J]. Statistics and Computing,2015,25(4):713-726.

[11] GUPTA R D,KUNDU D.Generalized exponential distributions[J].Australian & New Zealand Journal of Statistics,1999,41:173-188.

[12] CHIANG-SHENG L,HSINE-JEN T.A note on the generalized linear exponential distribution[J].Statistics & Probability Letters,2017,124(5):49-54.

[13] NAMHYUN K.Approximate MLE for the scale parameter of the generalized exponential distribution under random censoring[J].Journal of the Korean Statistical Society,2014,43(1):119-131.

[14] DEBASIS K,RAMESHWAR D G.An extension of the generalized exponential distribution[J].Statistical

Methodology,2011,8(6):485-496.

[15] NELSON W B. Accelerated life testing-step-stress models and data analysis[J].IEEE Transaction on Reliability,1980,29(2):103-108.

Statistical Analysis Based on EM Algorithm for Accelerated Life Testing Under Interval Censored Samples

QIU Leipin

(Department of Computer Science, Minjiang Teachers College, Fuzhou 350002, China)

Abstract: An accelerated life test carried out by step stress is considered when the the life time follows a generalized exponential distribution and the failure data obtained is interval censored. The test procedure is presented and the exchange formula of time is given by some assumptions. With the help of the expectation-maximization (EM) algorithm which is widely used when the observations can be viewed as incomplete data, the maximum likelihood estimates are computed. Moreover, an example by Mote Carlo data simulation is given to illustrate the procedure and show that this method is available, especially in large sample case.

Key words: interval censored; generalized exponential distribution; EM algorithm