一种新的寿命分布

高艳红,周秀轻

(南京师范大学数学科学学院,江苏 南京 210023)

[摘要] 本文介绍了一种带有两个参数的且危险率函数单调递增的分布,讨论了这种分布的各种性质,并给出了参数的极大似然估计. 利用 EM 算法得到了参数所满足的迭代方程组以及极大似然估计的渐近方差 - 协方差阵.

[关键词] EM 算法,Rayleigh 分布,危险率函数,几何分布,生存函数,极大似然估计

[中图分类号] O212.1 [文献标志码] A [文章编号] 1001-4616(2012)04-0018-07

A New Life Distribution

Gao Yanhong,Zhou Xiuqing

(School of Mathematical Sciences,Nanjing Normal University,Nanjing 210023,China)

Abstract: A two-parameter distribution with increasing failure rate is introduced. Various properties are discussed and the estimation of parameters is studied by the method of maximum likelihood. The estimates are attained by the EM algorithm and expressions for their asymptotic variances and covariances are obtained.

Key words: EM algorithm,Rayleigh distribution,failure rate,geometric distribution,lifetime distributions,maximum likelihood estimation

为了更好地拟合真实的寿命数据,很多人研究了带有两个参数的寿命分布. 寿命分布中一个非常重要的概念就是危险率(也称失效率). 根据危险率的不同得到了不同的密度函数,其中危险率单调递减和危险率为浴盆形状的两参数寿命分布问题研究的比较多.

当系统处于工作旺盛时期,人的器官免疫力较强的时候,危险率函数就会不断地降低. Adamidis K 和 Loukas A S[1]、Coskun Kus[2]、Rasoo Tahmasbia 和 Sadegh Rezaeib[3] 研究的是危险率递减的混合分布,分别为 EG 分布(指数 - 几何分布)、EP 分布(指数 - 泊松分布)、EL 分布(指数 - 对数几何分布). 他们研究了新的分布的各种性质如:生存函数、危险率函数、矩等等,利用 EM 算法得到了参数的极大似然估计以及渐近方差和协方差,并做了数值上的模拟.

通常一个系统的危险率随着时间的变化是连续的,开始的时候系统与子系统之间以及周围的环境之间的适应能力比较好,危险率函数单调递减. 随着使用时间的延长,系统的性能就会处于稳定阶段,这个时候危险率函数几乎不再发生变化. 但是,当系统处于老化阶段之后,系统的性能就会降低,危险率会逐渐地升高,这样的分布就是浴盆形状的寿命分布. Chen Zhenmin[4],Xie M、Tang Y 和 Goh T N[5]、Jong-Wuu wu 和 Chun-Hsien Wu 等[6] 主要研究的是累积分布函数为指数 - 威布尔分布的分布,给出了参数的似然估计、区间估计以及与其他分布之间的联系.

当系统处于老化阶段时,系统性能会逐渐降低;当人处于老年时期人的器官的免疫能力减弱,抵抗疾病的能力下降. 对于上面提到的情况以及其他的一些情况危险率函数都是单调递增的,而从事这方面的研究非常少,因此对于这种混合分布的研究还是比较有实际意义的. 本文主要研究的是危险率递增的 RG 分布(瑞丽 - 几何分布),例如处于老化阶段的电器的寿命分布以及系统中导体的寿命分布等等.

收稿日期:2011 - 12 - 33.
基金项目:国家自然科学基金(10771104).
通讯联系人:高艳红,硕士,助教,研究方向:生存分析. E-mail:goalgyh@163.com

1　分布

假设 $Y_1,Y_2,Y_3,\cdots,Y_z$ 为一列独立同分布的随机变量，其密度函数为

$$f(y,\beta)=2\beta y\mathrm{e}^{-\beta y^2}\quad \beta,y\in\mathbf{R}_+,$$

Z 为服从几何分布的随机变量，其概率分布函数为

$$P(z,p)=(1-p)p^{z-1},z\in\mathbf{N},p\in(0,1),$$

其中 Y_s 和 Z 是相互独立的. 记 $X=\min\{Y_1,Y_2,\cdots,Y_z\}$，这时可得

$$f(x\mid z,\beta)=2\beta zx\mathrm{e}^{-\beta zx^2}.$$

进而可得

$$F(x,\theta)=\sum_{k=1}^{+\infty}P(X\leqslant x\mid z=k,\theta)P(z=k\mid\theta)=\sum_{k=1}^{+\infty}\int_0^x 2\beta ks\mathrm{e}^{-\beta ks^2}\mathrm{d}s(1-p)p^{k-1}=\sum_{k=1}^{+\infty}(1-\mathrm{e}^{-\beta kx^2})(1-p)p^{k-1}=$$

$$(1-p)\left(\frac{1}{1-p}-\frac{\mathrm{e}^{-\beta x^2}}{1-p\mathrm{e}^{-\beta x^2}}\right)=1-\frac{(1-p)\mathrm{e}^{-\beta x^2}}{1-p\mathrm{e}^{-\beta x^2}},\tag{1}$$

其中 $\theta=(\beta,p)$，(1) 式中 $F(x,\theta)$ 对 x 求导数，则可以求得 X 的边缘密度函数为

$$f(x,\theta)=2\beta(1-p)x\mathrm{e}^{-\beta x^2}(1-p\mathrm{e}^{-\beta x^2})^{-2}.\tag{2}$$

称(2) 式所得到的分布为 Rayleigh-Geometric 分布(简称 RG 分布).

RG 是一个混合分布，是通过将 Rayleigh 分布与几何分布混合在一起得到的. Z 为系统中起初存在的同一类型缺陷的个数，Z 为未知的，缺陷的出现将导致系统的失效，Y_s 表示它们的寿命. 当 $0<p<\frac{\sqrt{\mathrm{e}}}{2}$ 时得到了危险率函数单调递增的 Rayleigh-Geometric 分布. 由于 $P(z\geqslant m)=\sum_{k=m}^{+\infty}p^{k-1}(1-p)=p^{m-1}$ 关于 p 是单调递增的，也就是说随着 p 的增大系统中存在多个缺陷的概率在增大，当概率接近于 1 的时候系统中存在的缺陷过多，那么就需要较多的维修费用，可能就没有维修的必要了，所以考虑 $0<p<\frac{\sqrt{\mathrm{e}}}{2}$ 的情形还是有意义的. 对于参数给定后，RG 分布的概率密度函数如图 1. 当 $p\to 0$ 时，RG 分布趋近于参数为 β 的 Rayleigh 分布.

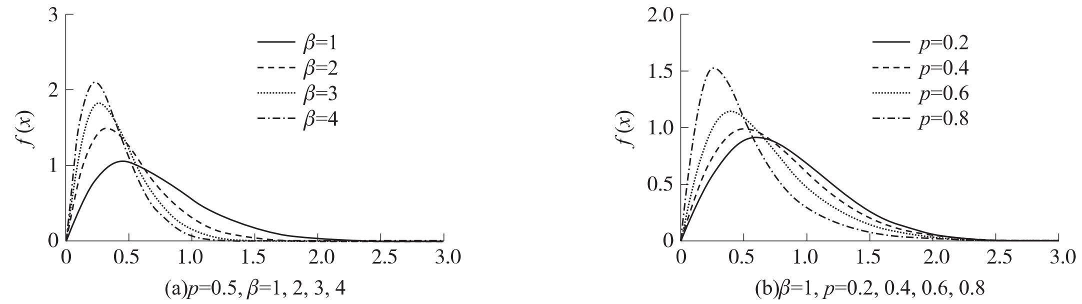

图 1　参数取不同值时，RG 分布的密度函数

Fig. 1　Probability density functions of the RG distribution

2　分布的性质

2.1　分布函数和矩

由(1) 式可知，RG 分布的分布函数为

$$F(x,\theta)=1-\frac{(1-p)\mathrm{e}^{-\beta x^2}}{1-p\mathrm{e}^{-\beta x^2}},$$

进而可以得到 X 的中位数为 $\sqrt{\frac{\ln(2-p)}{\beta}}$. 对于 $\forall r\in\mathbf{N}$，通过对(2) 式直接进行积分可以得到 X 的 r 阶矩为

$$E(X^r,\theta)=\int_0^{+\infty}x^r 2\beta(1-p)x\mathrm{e}^{-\beta x^2}(1-p\mathrm{e}^{-\beta x^2})^{-2}\mathrm{d}x=2\beta(1-p)\int_0^{+\infty}x^{r+1}\mathrm{e}^{-\beta x^2}\sum_{k=1}^{+\infty}kp^{k-1}\mathrm{e}^{-(k-1)\beta x^2}\mathrm{d}x=$$

$$2\beta(1-p)\sum_{k=1}^{+\infty}kp^{k-1}\int_0^{+\infty}x^{r+1}\mathrm{e}^{-k\beta x^2}\mathrm{d}x=\beta^{\frac{-r}{2}}(1-p)\Gamma\left(\frac{r}{2}+1\right)\sum_{k=1}^{+\infty}k^{-\frac{r}{2}}p^{k-1}=p^{-1}\beta^{-\frac{r}{2}}(1-p)\Gamma\left(\frac{r}{2}+1\right)L\left(p;\frac{r}{2}\right),\tag{3}$$

其中 $\Gamma(\cdot)$ 表示伽玛函数[7], $L\left(p;\frac{r}{2}\right)=\sum_{k=1}^{+\infty}k^{-\frac{r}{2}}p^k$. 于是可以得到 RG 分布的数学期望和方差分别为

$$E(X,\theta)=p^{-1}\beta^{-\frac{1}{2}}(1-p)\frac{\sqrt{\pi}}{2}L\left(p;\frac{1}{2}\right),$$

$$\mathrm{Var}(X,\theta)=p^{-2}\beta^{-1}(1-p)\left[p\ln(1-p)-\frac{\pi}{4}(1-p)\left(L\left(p,\frac{1}{2}\right)\right)^2\right],$$

进而可以求得变异系数为

$$C_v(X)=\frac{2}{\sqrt{\pi}}\left[p(p-1)^{-1}\ln(1-p)-\frac{\pi}{4}(p-1)L^2\left(p,\frac{1}{2}\right)\right]^{\frac{1}{2}}L^{-1}\left(p,\frac{1}{2}\right).$$

2.2 生存函数,危险率函数和平均剩余寿命

利用(1)式和(2)式,可得 RG 分布的生存函数(也称可靠性函数)和危险率函数(也称为效率函数)分别为

$$s(x,\theta)=1-F(x,\theta)=(1-p)\mathrm{e}^{-\beta x^2}(1-p\mathrm{e}^{-\beta x^2})^{-1},$$
$$h(x,\theta)=\frac{f(x,\theta)}{s(x,\theta)}=2\beta x(1-p\mathrm{e}^{-\beta x^2})^{-1},$$

则 $h(x,\theta)$ 对 x 求导数为

$$h'(x,\theta)=2\beta\left(\frac{x}{1-p\mathrm{e}^{-\beta x^2}}\right)'=2\beta\frac{\mathrm{e}^{\beta x^2}-p-2\beta p x^2}{(1-p\mathrm{e}^{-\beta x^2})^2\mathrm{e}^{\beta x^2}}.$$

当 $0<p<\frac{\sqrt{\mathrm{e}}}{2}$时,$h'(x,\theta)>0$,由此可以知道 $0<p<\frac{\sqrt{\mathrm{e}}}{2}$时 RG 分布的危险率是单调递增的,对于给定参数时,危险率函数 $h'(x,\theta)$ 的图像如图 2 所示.

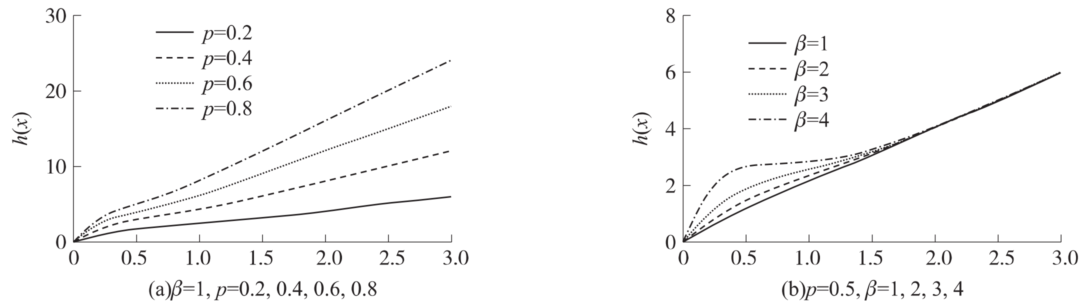

图 2　参数取不同值时,RG 分布的危险率函数

Fig. 2　Failure rate of the RG distribution

RG 分布的平均剩余寿命函数为

$$m(x_0;\theta)=E(X-x_0\mid X\geqslant x_0;\theta)=\frac{\sqrt{\pi}}{2}p^{-1}\beta^{-\frac{1}{2}}(\mathrm{e}^{\beta x_0^2}-p)\sum_{k=1}^{+\infty}k^{-\frac{1}{2}}p^k\mathrm{e}^{-k\beta x_0^2},$$

显然当 $x_0\to 0$ 时,$m(x_0;\theta)\to E(X;\theta)$;当 $x_0\to+\infty$ 时,$m(x_0;\theta)\to\frac{\sqrt{\pi}}{2}\beta^{-1}$.

3 参数估计

3.1 极大似然估计

由(3)式可以看到利用矩估计的方法很难得到 RG 分布的参数估计,下面将利用极大似然估计的方法得出 RG 分布的参数估计,记 $y_{\mathrm{obs}}=(x_i,i=1,2,\cdots,n)$ 为从 RG 分布中抽取的样本,由 RG 的分布密度函数

$$f(x,\theta)=2\beta(1-p)x\mathrm{e}^{-\beta x^2}(1-p\mathrm{e}^{-\beta x^2})^{-2},$$

可以得到其似然函数为

$$L(\theta,y_{\text{obs}})=n\ln[2\beta(1-p)]+\sum_{i=1}^{n}\ln x_i-\beta\sum_{i=1}^{n}x_i^2-2\sum_{i=1}^{n}\ln(1-p\mathrm{e}^{-\beta x_i^2}).$$

上式分别对β,p求偏导数可得

$$\frac{\partial L}{\partial\beta}=\frac{n}{\beta}+\sum_{i=1}^{n}x_i^2-2\sum_{i=1}^{n}\{x_i^2(1-p\mathrm{e}^{-\beta x_i^2})^{-1}\},\tag{4}$$

$$\frac{\partial L}{\partial p}=-\frac{n}{1-p}+2\sum_{i=1}^{n}\frac{\mathrm{e}^{-\beta x_i^2}}{1-p\mathrm{e}^{-\beta x_i^2}}.\tag{5}$$

令$\frac{\partial L}{\partial\beta}=0$,其解的情况由下面的定理给出.

定理 记(4)式右边为$g(\beta,p,y_{\text{obs}})$,其中$p$为参数真值,且$0<p<\frac{\sqrt{\mathrm{e}}}{2}$,则对于任给的$0<p<\frac{\sqrt{\mathrm{e}}}{2}$,由$g(\beta,p,y_{\text{obs}})=0$得到的$\beta$的估计值是惟一的,并且$\hat{\beta}\in\left((1-p)\left\{\frac{\sum_{i=1}^{n}x_i^2(1+p)}{n}\right\}^{-1},\left\{\frac{\sum_{i=1}^{n}x_i^2}{n}\right\}^{-1}\right)$.

证明 令

$$w(\beta,p,y_{\text{obs}})=2\sum_{i=1}^{n}\{x_i^2(1-p\mathrm{e}^{-\beta x_i^2})^{-1}\},$$

又$w(\beta,p,y_{\text{obs}})$对$\beta$的导数为

$$\frac{\partial w(\beta,p,y_{\text{obs}})}{\partial\beta}=2\sum_{i=1}^{n}\frac{-px_i^4\mathrm{e}^{-\beta x_i^2}}{(1-p\mathrm{e}^{-\beta x_i^2})^2}<0,$$

故$w(\beta,p,y_{\text{obs}})$关于$\beta$是单调递减的. 又

$$\lim_{\beta\to\infty}w(\beta,p,y_{\text{obs}})=2\sum_{i=1}^{n}x_i^2,$$

所以

$$g(\beta,p,y_{\text{obs}})<\frac{n}{\beta}+\sum_{i=1}^{n}x_i^2-\lim_{\beta\to\infty}w(\beta,p,y_{\text{obs}})=\frac{n}{\beta}-\sum_{i=1}^{n}x_i^2.$$

故当$\beta>\frac{n}{\sum_{i=1}^{n}x_i^2}$时,$g(\beta,p,y_{\text{obs}})<0$. 又因为

$$\lim_{\beta\to 0}w(\beta,p,y_{\text{obs}})=2(1-p)^{-1}\sum_{i=1}^{n}x_i^2,$$

于是有

$$g(\beta,p,y_{\text{obs}})>\frac{n}{\beta}+\sum_{i=1}^{n}x_i^2-\lim_{\beta\to 0}w(\beta,p,y_{\text{obs}})=\frac{n}{\beta}-(1+p)(1-p)^{-1}\sum_{i=1}^{n}x_i^2.$$

所以当$\beta<(1-p)\left\{\frac{\sum_{i=1}^{n}x_i^2(1+p)}{n}\right\}^{-1}$时,$g(\beta,p,y_{\text{obs}})>0$. 进而$g(\beta,p,y_{\text{obs}})=0$在区间$\left((1-p)\left\{\frac{\sum_{i=1}^{n}x_i^2(1+p)}{n}\right\}^{-1},\left\{\frac{\sum_{i=1}^{n}x_i^2}{n}\right\}^{-1}\right)$上至少有一个根.

下面证明$g(\beta,p,y_{\text{obs}})=0$的根是惟一的. 由已知

$$g(\beta,p,y_{\text{obs}})=\frac{n}{\beta}+\sum_{i=1}^{n}x_i^2-2\sum_{i=1}^{n}\{x_i^2(1-p\mathrm{e}^{-\beta x_i^2})^{-1}\},$$

上式对β求导得

$$g'(\beta,p,y_{\text{obs}})=\frac{\partial g(\beta,p,y_{\text{obs}})}{\partial\beta}=-\frac{n}{\beta^2}+2p\sum_{i=1}^{n}\{x_i^4\mathrm{e}^{-\beta x_i^2}(1-p\mathrm{e}^{-\beta x_i^2})^{-2}\}.$$

由此可知对于$\forall\beta>0$,$g'(\beta,p,y_{\text{obs}})$不总是单调的. 设$g'(\beta,p,y_{\text{obs}})=0$的一个根为$\beta^*$,则有

$$\frac{n}{\beta^*}=2\beta^* p\sum_{i=1}^{n}\{x_i^4 e^{-\beta^* x_i^2}(1-pe^{-\beta^* x_i^2})^{-2}\}.$$

于是

$$g(\beta^*,p,y_{\text{obs}})=2\beta^* p\sum_{i=1}^{n}\{x_i^4 e^{-\beta^* x_i^2}(1-pe^{-\beta^* x_i^2})^{-2}\}+\sum_{i=1}^{n}x_i^2-2\sum_{i=1}^{n}\{x_i^2(1-pe^{-\beta^* x_i^2})^{-1}\}=$$
$$\sum_{i=1}^{n}\frac{x_i^2(p^2e^{-2\beta^* x_i^2}+2\beta^* px_i^2e^{-\beta^* x_i^2}-1)}{(1-pe^{-\beta^* x_i^2})^2}.$$

记$f(p,\beta^*,x)=p^2e^{-2\beta^* x_i^2}+2\beta^* px_i^2e^{-\beta^* x_i^2}-1$. $f(p,\beta^*,x)$对p求导数得

$$f'(p,\beta^*,x)=2pe^{-2\beta^* x_i^2}+2\beta^* x_i^2e^{-\beta^* x_i^2}>0,$$

故$f(p,\beta^*,x)$关于p是单调递增的. 所以

$$p^2e^{-2\beta^* x_i^2}+2\beta^* px_i^2e^{-\beta^* x_i^2}-1<2e^{-\beta^* x_i^2}\{\beta^* x_i^2-sh(\beta^* x_i^2)\}<0,$$

其中$\text{sh}(\beta^* x_i^2)\frac{=e^{\beta^* x_i^2}-e^{-\beta^* x_i^2}}{2}$. 故对于所有的$\beta^*$, $x_i>0, i=1,2,\cdots,n, g(\beta^*,p,y_{\text{obs}})<0$,也就是说$g(\beta,p,y_{\text{obs}})$在其稳定点处的值是负的,又因为

$$\lim_{\beta\to\infty}g(\beta,p,y_{\text{obs}})=-\sum_{i=1}^{n}x_i^2,$$

所以$g(\beta,p,y_{\text{obs}})=0$的根最多只有一个. 而前面已经证明出$g(\beta,p,y_{\text{obs}})=0$在区间$\left((1-P)\left\{\frac{\sum_{i=1}^{n}x_i^2(1+p)}{n}\right\}^{-1},\left\{\frac{\sum_{i=1}^{n}x_i^2}{n}\right\}^{-1}\right)$上至少有一个根,故$g(\beta,p,y_{\text{obs}})=0$的根是惟一的,进而可得由$g(\beta,p,y_{\text{obs}})=0$得到的$\beta$的估计值是惟一的. 定理得证.

3.2 EM算法

EM算法是由Dempster, Laird, Rubin[8]于1977年提出的求参数极大似然估计的一种迭代方法,EM算法的主要特征是每个迭代都由两步组成:第一步是求期望;第二步是求极大值. 这种方法可以广泛地应用于缺损数据,截尾数据,成群数据以及带有讨厌参数的数据等所谓的不完全数据. 设$\mathcal{X}$和$\mathcal{Y}$为两个样本空间,$y(x)$为$\mathcal{X}$到$\mathcal{Y}$上的一个多对一的映射,即设$X\in\mathcal{X}$为完全数据,它是不能够观测到的,我们能观测到的仅仅是$\mathcal{Y}$内的Y,即所谓的不完全数据. 设参数$\theta\in\Theta$, X的密度函数为$f(x|\theta)$, $\mathcal{X}(y)\doteq\{x:y(x)=y\}$,则$y$的密度函数为

$$g(y|\theta)=\int_{\mathcal{X}}f(x|\theta)\mu(\mathrm{d}x).$$

EM算法[9]就是试图对$\theta\in\Theta$,极大化$\ln f(x|\theta)$在给定y和$\theta^{(p)}$($\theta^{(p)}$表示第p步θ的迭代值)处的条件期望. 记

$$Q(\theta|\theta')=E(\ln f(x,\theta)|y,\theta').$$

在给定参数θ的初值$\theta^{(0)}$后,对$k=0,1,2,\cdots$执行以下两步:

E步:计算$Q(\theta|\theta^{(k)})$;

M步:求$\theta^{k+1}\in\Theta$使$Q(\theta|\theta^{(k+1)})$极大化,直到满足一定的条件停止迭代.

下面将利用EM算法求参数β,p的估计.

记$(x_i,z_i;i=1,2,\cdots,n)$为完全数据集,它是不能被完全观测到的,只能观测到其中的x_i,记$y_{\text{obs}}=(x_i;i=1,2,\cdots,n)$. 完全数据的密度函数为

$$f(x,z;\theta)=2\beta zxe^{-\beta zx^2}p^{z-1}(1-p),$$

其中$x,\beta\in\mathbf{R}_+$, $z\in\mathbf{N}$, $p\in(0,1)$, $\theta=(\beta,p)$. 进一步可以得到

$$P(z|x;\theta)=\frac{f(x,z;\theta)}{f(x;\theta)}=zp^{z-1}e^{-\beta(z-1)x^2}(1-pe^{-\beta x^2})^2.$$

根据EM算法,设第n步的条件期望为$E(Z|X;\theta^{(n)})$,其中$\theta^{(n)}=(\beta^{(n)},p^{(n)})$. 则

$$E(Z|X;\theta^{(n)}) = \sum_{z=1}^{+\infty} zP(z|x;\theta^{(n)}) = 2(1-p^{(n)}\mathrm{e}^{-\beta^{(n)}x^2})^{-1}-1.$$

由于完全数据的密度函数为

$$f(x,z,\theta) = 2\beta zx\mathrm{e}^{-\beta zx^2(1-p)p^{z-1}},$$

故可以求得 $f(x,z,\theta)$ 的似然函数为

$$L(x,z|\theta) = n\ln 2\beta + \sum_{i=1}^{n}\ln z_i + 2\sum_{i=1}^{n}\ln x_i - \beta\sum_{i=1}^{n}x_i^2 z_i + n\ln(1-p) + \left(\sum_{i=1}^{n}z_i - n\right)\ln p.$$

上式分别对 β, p 求偏导数得

$$\begin{cases}\dfrac{\partial L}{\partial\beta} = \dfrac{n}{\beta} - \sum\limits_{i=1}^{n}x_i^2 z_i, \\ \dfrac{\partial L}{\partial p} = -\dfrac{n}{1-p} + \dfrac{\sum\limits_{i=1}^{n}z_i - n}{p}.\end{cases}$$

令 $\dfrac{\partial L}{\partial\beta}, \dfrac{\partial L}{\partial p}$ 分别为 0，则有

$$\begin{cases}\beta = n\left(\sum\limits_{i=1}^{n}x_i^2 z_i\right)^{-1}, \\ p = 1 - n\left(\sum\limits_{i=1}^{n}z_i\right)^{-1}.\end{cases}$$

根据 EM 算法的思想，用 $E(Z|X,\theta)$ 代替 z 得

$$\begin{cases}\beta^{(n+1)} = n\left\{\sum\limits_{i=1}^{n}x_i^2\left(\dfrac{1+p^{(n)}\mathrm{e}^{-\beta(n)x_i^2}}{1-p^{(n)}\mathrm{e}^{-\beta(n)x_i^2}}\right)\right\}^{-1}, \\ p^{(n+1)} = 1 - n\left\{\sum\limits_{i=1}^{n}\left(\dfrac{1+p^{(n)}\mathrm{e}^{-\beta(n)x_i^2}}{1-p^{(n)}\mathrm{e}^{-\beta(n)x_i^2}}\right)\right\}^{-1}.\end{cases}$$

这样就得到参数 β 和 p 的估计所满足的迭代方程组，解上面的非线性方程组即可得到估计值 $\hat{\beta}$ 和 $\hat{p}$.

3.3 极大似然估计的渐近方差和协方差

根据大样本的渐近正态性，参数 θ 的极大似然估计渐近于均值为 θ，方差 - 协方差阵为 $\boldsymbol{J}^{-1}(\theta)$ 的二元正态分布，其中 $\boldsymbol{J}(\theta) = \boldsymbol{E}(\boldsymbol{I};\theta)$，$\boldsymbol{I} = \boldsymbol{I}(\theta;y_{\mathrm{obs}})$ 为观测值的 Fisher 信息阵，$I_{ij} = -\dfrac{\partial^2 L}{\partial\theta_i\partial\theta_j}(i,j=1,2)$ 为 Fisher 信息阵中的元素，对(4)式和(5)式求导数，则可以得到

$$I_{11} = 2\sum_{i=1}^{n}\frac{px_i^4\mathrm{e}^{-\beta x_i^2}}{(1-p\mathrm{e}^{-\beta x_i^2})^2} - \frac{n}{\beta^2},$$

$$I_{12} = I_{21} = 2\sum_{i=1}^{n}\frac{x_i^2\mathrm{e}^{-\beta x_i^2}}{(1-p\mathrm{e}^{-\beta x_i^2})^2},$$

$$I_{22} = 2\sum_{i=1}^{n}\frac{\mathrm{e}^{-2\beta x_i^2}}{(1-p\mathrm{e}^{-\beta x_i^2})^2} - \frac{n}{\beta^2} - \frac{n}{(1-p)^2},$$

$$E\left\{\frac{X^r\mathrm{e}^{-s\beta X^2}}{(1-p\mathrm{e}^{\mathrm{e}^{-\beta X^2}})^2};\theta\right\} = (1-p)\beta^{-\frac{r}{2}}(s+1)^{-\frac{r}{2}-1}\Gamma\left(1+\frac{r}{2}\right).$$

为了求 $\boldsymbol{J}(\theta)$，先求出所用到的期望值如下

$$E\left\{\frac{X^4\mathrm{e}^{-\beta X^2}}{(1-p\mathrm{e}^{-\beta X^2})^2};\theta\right\} = \frac{1-p}{4\beta^2},$$

$$E\left\{\frac{X^2\mathrm{e}^{-\beta X^2}}{(1-p\mathrm{e}^{-\beta X^2})^2};\theta\right\} = \frac{1-p}{4\beta},$$

$$E\left\{\frac{\mathrm{e}^{-2\beta X^2}}{(1-p\mathrm{e}^{-\beta X^2})^2};\theta\right\} = \frac{1-p}{3},$$

于是可得

$$J_{11}=\frac{-n(1+p)}{2\beta^2},$$

$$J_{12}=\frac{n(1-p)}{2\beta},$$

$$J_{22}=\frac{2n(1-p)}{3}.$$

$J^{-1}(\hat{\theta})$ 便为参数 θ 的极大似然估计的渐近方差-协方差阵.

[参考文献]

[1] Adamidis K, Loukas A S. Lifetime distribution with decreasing failure rate [J]. Statistics Probability Letters, 1998, 39(1): 35-42.

[2] Coskun Kus. A new lifetime distribution [J]. Computational Statistics Data Analysis, 2007, 51(9): 4 497-4 509.

[3] Rasool Tahmasbia, Sadegh Rezaeib. A two-parameter lifetime distribution with decreasing failure rate [J]. Computational Statistics and Data Analysis, 2008, 52(8): 3 889-3 901.

[4] Chen Zhenmin. A new two-parameter lifetime distribution with bathtub shape or increasing failure rate function [J]. Statistics Probability Letters, 2000, 49(2): 155-161.

[5] Xie M, Tang Y, Goh T N. A modified Weibull extension with bathtub-shaped failure rate function [J]. Reliability Engineering and System Safety, 2002, 76(3): 279-285.

[6] Wu Zhongwu, Lu Hailin, Chen Chonghong, et al. Statistical inference about the shape parameter of the new two-parameter bathtub-shaped lifetime distribution [J]. Research, 2004, 20(6): 607-616.

[7] 刘式适, 刘式达. 特殊函数[M]. 北京: 气象出版社, 2002: 2-10.

[8] Dempster A P, Laird N M, Rubin D B. Maximum likelihood from incomplete data via the EM algorithm [J]. Journal of the Royal Statistical Society B, 1977, 39(1): 1-38.

[9] 王松桂. EM算法[J]. 应用数学与计算数学, 1983(6): 43-48.

[责任编辑: 丁　蓉]