正态分布场合下只有一个失效数据的统计分析

张 志 华
(海军工程学院, 武汉, 430033)

摘 要

本文利用母体分布函数的凹凸性给出了各检测时刻失效概率的Bayes估计, 该估计计算简便可行. 最后将此方法应用于实际例子, 所得结果较为稳健.

关键词: 正态分布, Bayes 估计, 只有一个失效数据
学科分类号: 213.2

§1. 引 言

在评定产品的可靠性时, 由于受实际试验条件的限制, 人们常常采用各种定时截尾试验方法对产品进行可靠性试验. 在产品质量不太高的情况下, 所获得的定时截尾数据的样品失效个数一般有$r \geq 2$, 此时处理定时截尾数据已有很多成熟的统计方法[2]. 但是对于高可靠、长寿命产品, 在规定的试验时间内所有参试样本均没有失效(即无失效数据)或只有个别产品失效(即只有一个失效数据Data only one failure, 简记DOOF数据)二种情况常常发生, 因此研究无失效数据及DOOF数据的统计方法是十分重要的. 对于无失效数据, 在综合产品的验前信息基础上, 文献[1, 3]给出了几种处理方法. 对于DOOF数据, 在产品寿命分布已知情况下, 可以利用极大似然估计方法进行处理, 但是文献[4]指出: 所获得的参数MLE并不稳健. 为此文献[4]采用多层Bayes方法对DOOF数据进行处理, 得到的参数估计具有较好的稳健性.

本文利用母体分布函数的凹凸性给出了DOOF数据一种新的 Bayes处理方法. 在§2中描述了DOOF数据的统计模型. 在§3中, 利用正态分布函数的凹凸性确定了各检测时刻的失效概率$p_i\,(i=1,2,\cdots,k)$的可能变化范围, 由此较方便的获得$p_i\,(i=1,2,\cdots,k)$的先验分布及其Bayes估计, 且此估计计算简便可行. 最后利用本文方法对某型号轴承寿命试验中出现的DOOF数据进行处理, 结果表明本文给出的方法较为稳健.

§2. DOOF数据的模型

设产品寿命T服从正态分布$N(\mu,\sigma^2)$, 其中μ,σ^2为未知参数. 为评定产品的可靠性, 随机的选取n个样品进行寿命试验, 现将这些产品分成k组, 各组样品数分别为$n_1,\cdots,n_k$, $(\sum\limits_{i=1}^{k} n_i = n)$. 各组试验的起始时刻相同(记为时刻0), 而结束时刻不相同. 事先给定k个常数$t_1,t_2,\cdots,t_k$, 其中$t_1<t_2<\cdots<t_k$, 对于第i组产品. 试验进行到预先给定时刻t_i时停止这n_i个样品试验, 而在整个试验中全部n个参试样品只有一个失效, 设该产品的失效发生在(t_{m-1},t_m)之间, 称这类数据

本文1996年9月20日收到, 1997年4月29日收到修改稿.

为DOOF数据. 记$s_l = \sum_{i=l}^{k} n_i$, 则DOOF数据可以表示为:

$$(t_l, s_l, r_l), \qquad l = 1, 2, \cdots, k, \tag{1}$$

其中s_l, r_l分别表示在时刻t_l时的样品参试个数及失效个数, 此时$s_1 \geq s_2 \geq \cdots \geq s_k$, 当$l \leq m-1$时, $r_l = 0$, 当$l \geq m$时, $r_l = 1$.

综合上述试验, 数据(1)提供的信息有:

1^o: $t = 0$时产品失效概率$p_0 = 0$(或接近为0);

2^o: 由于$0 < t_1 < t_2 < \cdots < t_k$, 记$p_i = \mathrm{P}(T < t_i)$, 则有

$$p_1 < p_2 < \cdots < p_k, \tag{2}$$

在s_k较大时, $p_i\,(i = 1, 2, \cdots, k)$都较小.

我们的问题: 在假设产品寿命T服从正态分别的情况下, 如何利用上述信息估计未知参数μ, σ^2及其它可靠性指标.

§3. DOOF数据的统计分析

§3.1 $p_i\,(i = 1, 2, \cdots, k)$的先验分布

为了确定$p_i\,(i = 1, 2, \cdots k)$的先验分布, 我们首先确定产品在检验时刻$t_i$时失效概率$p_i$的可能取值范围, 由于多个样品试验到时刻$t_k$时只有一个样品失效, 则可以判定样品在$(0, t_k)$内的可靠度很高, 即失效概率$p_k$不会很大, 特别当$s_k$较大时, p_k更接近于零. 因此产品在时刻t_k之前的失效概率满足$p_k \leq 0.5$. 在一般情况下, 通过收集有关专家的设计使用经验信息可确定出p_k的更加精确上界λ_k, 此时p_k的取值范围为$(0, \lambda_k)$, 并且$\lambda_k \leq 0.5$.

由于母体寿命服从正态分布$N(\mu, \sigma^2)$, 并且

$$\frac{\partial f(t, \mu, \sigma^2)}{\partial t} = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(t-\mu)^2}{2\sigma^2}\right\} \frac{(\mu - t)}{\sigma^2}, \tag{3}$$

当$t < \mu$时, $\dfrac{\partial f(t, \mu, \sigma^2)}{\partial t} > 0$, 即当$t < \mu$时, 母体分布函数$F(t)$是$t$的凹函数. 由于$p_k < \lambda_k \leq 0.5$, 则$t_1 < t_2 < \cdots < t_k < \mu$, 由凹函数的性质可知:

$$0 < \frac{p_1 - p_0}{t_1 - t_0} < \frac{p_2 - p_0}{t_2 - t_0} < \cdots < \frac{p_k - p_0}{t_k - t_0}, \tag{4}$$

其中t_0是给定的一点, 且$t_0 < t_1, p_0 = \mathrm{P}(T < t_0)$. 当$t_0 = 0$时, p_0很小, 可近似取为0. 这样上式可写成:

$$0 < \frac{p_1}{t_1} < \frac{p_2}{t_2} < \cdots < \frac{p_k}{t_k}. \tag{5}$$

因$p_k < \lambda_k \leq 0.5$, 则

$$0 < \frac{p_1}{t_1} < \frac{p_2}{t_2} < \cdots < \frac{p_k}{t_k} < \frac{\lambda_k}{t_k}. \tag{6}$$

由此可见: 当母体服从正态分布时, $p_1, \cdots, p_k$应满足(6), 由[6]易知p_i的可能取值范围是$[0, \lambda_i]$(其中$\lambda_i = \dfrac{\lambda_k t_i}{t_k}$), 这是在估计失效概率$p_i$时可利用的先验信息. 若$p_i$无其它先验信息可用, 可认为$p_i$在$[0, \lambda_i]$上取值是等可能的. 则$p_i$的先验分布为:

$$\pi(p_i) = \begin{cases} \dfrac{1}{\lambda_i}, & 0 < p_i < \lambda_i, \\ 0, & \text{其它}. \end{cases} \tag{7}$$

在实际中,通过专家经验给出失效概率p_k的上界是十分方便的,因此由(6)确定p_i的先验分布是可行的,从而避免了在处理实际问题时取p_i的先验分布的技巧性.

§3.2 p_i的Bayes估计

由于在时刻t_i内有s_i个样品参加试验,有r_i个产品失效,则其似然函数为:

$$L(r_i,p_i)=\binom{s_i}{r_i}p_i^{r_i}(1-p_i)^{s_i-r_i}. \tag{8}$$

把上述先验分布(7)与似然函数结合,使用Bayes公式,立即可得到p_i的后验分布:

$$\pi(p_i|s_i,r_i)=\frac{L(r_i,p_i)}{\int_0^{\lambda_i}L(r_i,p_i)dp_i},\qquad i=1,2,\cdots,k. \tag{9}$$

在平方损失函数下,p_i的Bayes估计为: $$\widehat{p}_i=\frac{\int_0^{\lambda_i}p_iL(r_i,p_i)dp_i}{\int_0^{\lambda_i}L(r_i,p_i)dp_i}.$$

当$i<m$时,$r_i=0$,则p_i的后验分布及Bayes估计为:

$$\pi(p_i|s_i,r_i)=\frac{(s_i+1)(1-p_i)^{s_i}}{1-(1-\lambda_i)^{s_i+1}},\qquad 0<p_i<\lambda_i,$$

$$\widehat{p}_i=1-\frac{s_i+1}{s_i+2}\cdot\frac{1-(1-\lambda_i)^{s_i+2}}{1-(1-\lambda_i)^{s_i+1}},\qquad i=1,2,\cdots,m-1. \tag{10}$$

当$i\geq m$时,$r_i=1$,则p_i的后验分布及Bayes估计为:

$$\pi(p_i|s_i,r_i)=\frac{s_i(s_i+1)p_i(1-p_i)^{s_i-1}}{1-(1+s_i\lambda_i)(1-\lambda_i)^{s_i}},\qquad 0<p_i<\lambda_i,$$

$$\widehat{p}_i=\frac{2-[2+2s_i\lambda_i+s_i(s_i+1)\lambda_i^2](1-\lambda_i)^{s_i}}{(s_i+2)[1-(s_i+1)(1-\lambda_i)^{s_i}+s_i(1-\lambda_i)^{s_i+1}]},\qquad i=m,m+1,\cdots,k. \tag{11}$$

由于个检验时刻的失效概率$p_1,\cdots,p_k$满足(2),由此从(10)及(11)得到$p_1,\cdots,p_k$的估计$\widehat{p}_1,\cdots,\widehat{p}_k$也应保证不出现"倒挂"现象,即满足关系

$$\widehat{p}_1\leq\widehat{p}_2\leq\cdots\leq\widehat{p}_k, \tag{12}$$

(12)式的证明见附录A.

§3.3 可靠度估计

当产品寿命T服从正态分布$N(\mu,\sigma^2)$时,产品在时刻t_i的失效概率p_i可表示为:

$$p_i=\mathrm{P}(T\leq t_i)=\Phi(\frac{t_i-\mu}{\sigma}),\qquad i=1,2,\cdots,k,$$

其中$\Phi(\cdot)$为标准正态分布函数. 记$\Phi^{-1}(\cdot)$为其反函数,则上述可表示为:

$$t_i=\mu+\sigma\Phi^{-1}(p_i),\qquad i=1,2,\cdots,k.$$

若用$\widehat{p}_i\ (i=1,2,\cdots,k)$代替上式中的$p_i$,则上式可改写为:

$$t_i=\mu+\sigma\Phi^{-1}(\widehat{p}_i)+\varepsilon_i,\qquad i=1,2,\cdots,k,$$

其中ε_i是用$\widehat{p}_i$代替p_i所引起的误差. 利用最小二乘法,可得到参数μ,σ的估计

$$\widehat{\mu}=\frac{BC-AD}{kB-A^2},\qquad \widehat{\sigma}=\frac{kG-AC}{kB-A^2},$$

其中

$$A=\sum_{i=1}^{k}\Phi^{-1}(\widehat{p}_i),\qquad B=\sum_{i=1}^{k}[\Phi^{-1}(\widehat{p}_i)]^2,\qquad C=\sum_{i=1}^{k}t_i,\qquad D=\sum_{i=1}^{k}t_i\Phi^{-1}(\widehat{p}_i).$$

从而可得到任意时刻t的可靠度估计为: $\widehat{R}(t)=1-\Phi(\frac{t-\widehat{\mu}}{\widehat{\sigma}}).$

§4. 例

在某产品的可靠性试验中, 随机选取20套进行定时截尾试验, 可在规定实际522小时内仅有一个样品失效, 失效发生在(432, 462)小时内. 为获得更多失效数据, 希望试验能继续进行, 但由于经济原因, 只得对部分产品分批停止试验, 其余一直试验到1336小时全部停止, 但在522小时后的部分样品的延长试验中没有产品失效, 即在整个试验中, 只有一个样品失效, 全部试验数据列于表1:

i	t_i	s_i	r_i
1	432	20	0
2	462	20	1
3	522	20	1
4	742	16	1
5	962	12	1
6	1184	10	1
7	1336	5	1

表1: 试验数据(单位: 小时)

为了对试验数据进行处理, 必须首先确定失效概率p_7的上界λ_7. 为了研究λ_7对产品的可靠度的影响, 在本例中分别选取$\lambda_7=0.3, 0.4, 0.445, 0.5$等四种情况, 在$\lambda_7$选定的情况下, 分别用§3的结果计算得到诸$p_i$的Bayes估计, 由此可得到参数$\mu,\sigma$的估计和可靠度的估计, 全部估计结果列于表2.

λ_7	0.5	0.445	0.4	0.3
μ	2049.53	2032.16	2176.24	2390.44
σ	1015	966.26	1067.37	1150.35
$R(100)$	0.9726	0.9772	0.9738	0.9767
$R(200)$	0.9656	0.9706	0.9678	0.9713
$R(400)$	0.9479	0.9545	0.9515	0.9582
$R(800)$	0.8907	0.8980	0.9015	0.9162
$R(1200)$	0.7983	0.8051	0.8186	0.8485
$R(1500)$	0.7054	0.7088	0.7357	0.7794

表2: 参数估计及可靠度估计

从表2结果看出: 尽管选取的λ_7的变化范围较大, 但是所得参数μ,σ的估计波动并不很大, 特别在$T\le 1200$小时时, $\widehat{R(t)}$的波动更小, 说明此估计方法是稳健的. 而在实际问题中, 通过专家经验所确定的λ_7的变化范围一般较小, 即使在没有专家经验等先验信息的情况下所确定的λ_7的

变化范围一般不会有如此大, 因此在λ_7的可能变化范围内所得到的估计结果变化幅度更小, 从而说明本文方法是实用的.

附录A

证明 $\widehat{p}_1 \le \widehat{p}_2 \le \cdots \le \widehat{p}_k.$ (12)

引理1 $s \ge 1$为固定常数, 当$\mu \in (0.1)$时有:

$$\frac{s}{s+2}\frac{1-(s+2)\mu^{s+1}+(s+1)\mu^{s+2}}{1-(s+1)\mu^{s}+s\mu^{s+1}} \le \frac{s+1}{s+3}\frac{1-(s+3)\mu^{s+2}+(s+2)\mu^{s+3}}{1-(s+2)\mu^{s+1}+(s+1)\mu^{s+2}}, \tag{13}$$

$$\frac{s+1}{s+2}\frac{1-\mu^{s+2}}{1-\mu^{s+1}} \le \frac{s+2}{s+3}\frac{1-\mu^{s+3}}{1-\mu^{s+2}}, \tag{14}$$

$$\frac{(s+1)(1-\mu^{s+2})}{1-\mu^{s+1}} \ge \frac{s[1-(s+2)\mu^{s+1}+(s+1)\mu^{s+2}]}{1-(s+1)\mu^{s}+s\mu^{s+1}}. \tag{15}$$

证明 为证明不等式(13), 首先研究函数

$$\begin{aligned} g(\mu) =&2s\mu^{s+4}-2(2s+3)\mu^{s+3}+2(s+3)\mu^{s+2}+(s+2)(s+3)\mu^{3} \\ &-(s+3)(3s+2)\mu^{2}+s(3s+7)\mu-(s+1)s, \end{aligned}$$

在$\mu \in (0,1)$上的单调性, 对函数$g(\mu)$求三阶导数可得:

$$\frac{d^3g(\mu)}{d\mu^3} = 2(s+3)(s+2)[s(s+4)\mu^{s+1}-(2s+3)(s+1)\mu^{s}+s(s+1)\mu^{s-1}+3].$$

可以证明: 当$\mu \in (0,1)$时, $\frac{d^3g(\mu)}{d\mu^3} > 0$, 同时可以验证: $\frac{d^2g(\mu)}{d\mu^2}\Big|_{\mu=1} = \frac{dg(\mu)}{d\mu}\Big|_{\mu=1} = 0$, 因此$g(\mu)$在$\mu \in (0,1)$上是增函数. 又由于$g(1)=0$, 故当$\mu \in (0,1)$时, $g(\mu) < 0$. 令:

$$\begin{aligned} G(\mu) =&(s+1)(s+2)[1-(s+1)\mu^{s}+s\mu^{s+1}][1-(s+3)\mu^{s+2}+(s+2)\mu^{s+3}] \\ &-s(s+3)[1-(s+2)\mu^{s+1}+(s+1)\mu^{s+2}]^2, \end{aligned}$$

则

$$\frac{dG(\mu)}{d\mu} = (s+1)(s+2)\mu^{s-1}g(\mu),$$

由此可知: $G(\mu)$是减函数. 又因$g(1)=0$, 则当$\mu \in (0,1)$时, $G(\mu) > 0$, 从而说明不等式(13)成立.

不等式(14), (15)同理可证明.

引理2 (1) 函数$G_1(\mu,s) = \frac{1-(s+2)\mu^{s+1}+(s+1)\mu^{s+2}}{1-(s+1)\mu^{s}+s\mu^{s+1}}$关于$\mu \in (0,1)$是增函数;

(2) 函数$G_2(\mu,s) = \frac{1-\mu^{s+2}}{1-\mu^{s+1}}$关于$\mu \in (0,1)$是增函数.

证明 (1) 令$\psi(\mu) = s\mu^{s+2}-(s+2)\mu^{s+1}+(s+2)\mu-s$, 则

$$\frac{d^2\psi(\mu)}{d\mu^2} = s(s+1)(s+2)(\mu-1)\mu^{s-1},$$

当$\mu\in(0,1)$时$\frac{d^2\psi(\mu)}{d\mu^2}<0$, 又由于$\frac{d\psi(\mu)}{d\mu}\Big|_{\mu=1}=0, \psi(1)=0$, 则当$\mu\in(0,1)$时, $\psi(\mu)<0$. 又因为

$$\frac{dG_1(\mu)}{d\mu}=\frac{(s+1)(s+2)\mu^s(\mu-1)\psi(\mu)}{[1-(s+1)\mu^s+s\mu^{s+1}]^2},$$

当$\mu\in(0,1)$时$\frac{dG_1(\mu)}{d\mu}\geq 0$, 即$G_1(\mu)$关于$\mu\in(0,1)$是增函数.

同理可证明(2).

不等式(12)的证明:

由引理1及引理2可知, 函数

$$\phi_1(s,\mu)=1-\frac{s+1}{s+2}\frac{1-\mu^{s+2}}{1-\mu^{s+1}},$$

$$\phi_2(s,\mu)=1-\frac{s}{s+2}\frac{1-(s+2)\mu^{s+1}+(s+1)\mu^{s+2}}{1-(s+1)\mu^s+s\mu^{s+1}},$$

在$\mu\in(0,1)$固定下关于自然数s是减函数, 在d固定下关于$\mu\in(0,1)$是减函数.

令$\mu_l=1-\lambda_l, l=1,2,\cdots,k$, 则当$l\leq m-1$时, p_l的Bayes估计可表示为

$$\widehat{p}_l=\phi_1(s_l,\mu_l),\qquad l=1,2,\cdots,m-1.$$

令$\widehat{p}'_m=\phi_1(s_m,\mu_m)$, 由于$s_1>s_2>\cdots>s_{m-1}>s_m, \mu_1>\mu_2>\cdots>\mu_{m-1}>\mu_m$, 则可知

$$\widehat{p}_1\leq\widehat{p}_2\leq\widehat{p}_{m-1}\leq\widehat{p}'_m. \tag{16}$$

当$l\geq m$时, p_l的Bayes估计可表示为: $\widehat{p}_l=\phi_2(s_l,\mu_l),\qquad l=m,m+1,\cdots,k.$

由于$s_m>s_{m+1}>\cdots>s_k, \mu_m>\mu_{m+1}>\cdots>\mu_k$, 则有

$$\widehat{p}_m\leq\widehat{p}_{m+1}\leq\cdots\leq\widehat{p}_k. \tag{17}$$

由不等式(15)立得$\widehat{p}'_m\leq\widehat{p}_m$, 从而有 $\widehat{p}_1\leq\widehat{p}_2\leq\cdots\leq\widehat{p}_{m-1}\leq\widehat{p}_m\leq\cdots\leq\widehat{p}_k.$ 证毕.

参 考 文 献

[1] 茆诗松, 罗朝斌, 无失效数据的可靠性分析, 数理统计与应用概率, 4(4)(1989).
[2] 戴树森等, 可靠性实用及其统计分析, 国防工业出版社, 1985.
[3] 张志华, 无失效数据的统计分析, 数理统计与应用概率, 10(1995).
[4] 茆诗松, 陈军, 只有一个失效数据的Bayes分析, 投《应用概率统计》.

Statistical Analysis of Data Only One Failure Under Normal Case

ZHANG ZHIHUA

(*Naval Academy of Engineering, Wuhan*)

In this paper, we study a new Bayesian method of data only one failure from normal distribution, making use of concave-convex character of distribution function. And the estimation of parameters has been obtained simply and conveniently. Finally an example is given.