

文章编号: 1000 5013(2010)05-0597-04

指数分布恒加试验定时截尾试验 数据缺失时的 Bayes 分析

田 霆

(华侨大学 数学科学学院, 福建 泉州 362021)

摘要: 运用 Gibbs 抽样迭代方法, 解决 Bayes 分析中的后验边际分布的计算问题, 得到满足顺序约束的参数的 Bayes 估计. 通过 Monte-Carlo 模拟表明, 在各场合存在先验信息的情况下, Bayes 估计的相对偏差和相对均方误差都小于极大似然估计; 而对于没有先验信息的情况, Bayes 估计跟极大似然估计的效果差不多.

关键词: 指数分布; 恒定应力; 加速寿命; 定时截尾; 数据缺失; Bayes 分析

中图分类号: O 213.2

文献标识码: A

在短时间内对产品的可靠性指标进行评定, 恒定应力加速寿命(恒加试验)是加速寿命试验中最常用的一种, 故对其研究也较多^[1-2]. 当时, 以往的研究都是在没有缺失数据的情况下进行的. 在试验中, 由于各种原因常会遇到试验数据缺失的现象, 缺失数据一般较难处理. 对一般寿命试验数据缺失时的统计分析, 已有相关的结果^[3-4]. 在加速寿命试验中, 产品承受的应力高于正常应力水平, 失效加快, 数据缺失现象更容易发生; 然而, 对于加速寿命试验中的试验数据缺失问题, 却很少有人进行研究. 文[5]讨论了定数截尾恒加试验中数据缺失时统计分析方法, 并对各种估计的优良性进行模拟比较. 茆诗松等^[6]讨论了当寿命分布是指数分布时, 定数截尾场合下恒加应力加速寿命试验中常见的几类数据类型(完全样本、分组样本、删失样本)的 Bayes 统计分析方法. 同时, 他们运用 Gibbs 抽样迭代算法, 解决了 Bayes 分析中极为复杂的后验边际分布的计算问题, 得到满足顺序约束的参数的 Bayes 估计. 本文仅讨论定时截尾恒加试验中, 试验数据缺失时统计分析方法.

1 基本假定

选择 l 个应力水平 $S_1 < S_2 < \dots < S_l$, l 个应力水平均高于正常应力水平 S_0 ; 然后, 从一产品批中随机抽取 n_i 个产品放在应力 S_i 下进行定时截尾寿命试验. 当试验时间进行到 τ_i (τ_i 是预先给定的正数)时刻停止. 设产品的次序失效数据为 $0 = t_{i,0} < t_{i,1} < \dots < t_{i,r_i} \leq \tau_i, i = 1, 2, \dots, l$.

考虑试验数据缺失的情况, 假设第 i 个应力水平下的试验数据只剩下 k_i 个, 其取值依次为

$$0 = t_{i,r_{i,0}} < t_{i,r_{i,1}} < \dots < t_{i,r_{i,k_i}} \leq \tau_i, \quad i = 1, 2, \dots, l. \quad (1)$$

下面给出恒加试验的两个基本假定:

(I) 在正常应力水平 S_0 和加速应力水平 S_i 下, 产品的寿命都服从指数分布, 不同的仅在参数 θ 上. 即在应力水平 S_i 下, 产品寿命 T 的分布函数为 $F_i(t) = 1 - \exp(-\lambda t)$. 其中: $t > 0; i = 0, 1, \dots, l; \lambda$ 为失效率.

(II) 产品的平均寿命 θ 与所加应力 S_i 之间满足 $\ln \theta = a + b\phi(S_i)$. 其中: a, b 为未知的待估参数; $\phi(S_i)$ 为应力水平的已知单调函数. 常用的 Arrhenius 模型和逆幂律模型均可写成上述形式. 为后面行文方便, 记 $\phi = \phi(S_i)$.

收稿日期: 2008-11-19

通信作者: 田霆(1972), 男, 讲师, 主要从事产品可靠性的研究. E-mail: tianting1972928@sohu.com.

基金项目: 福建省自然科学基金资助项目(Z0511027)

在上述两个基本假定下, 讨论数据(1)的统计方法.

2 参数的 Bayes 估计

在上述两个基本假定下, 应力 S_i 下, 用概率元方法^[6] 求出试验数据的似然函数为

$$L_i(\lambda \mid t_{i,r_{i,1}}, \dots, t_{i,r_{i,k_i}}) = C \prod_{j=1}^{k_i} [f(t_{i,r_{i,j}})] \times \\ \prod_{j=0}^{k_i-1} [F(t_{i,r_{i,j+1}}) - F(t_{i,r_{i,j}})]^{r_{i,j+1}-r_{i,j}-1} \cdot [F(\tau_i) - F(t_{i,r_{i,k_i}})]^{r_{i,k_i}-r_{i,k_i}-1} \cdot [1 - F(\tau_i)]^{n_i-r_{i,k_i}}.$$

上式中, C 是与 λ 无关的常数. 即有

$$L_i(\lambda \mid t_{i,r_{i,1}}, \dots, t_{i,r_{i,k_i}}) = C \prod_{j=0}^{k_i-1} [\exp(-\lambda t_{i,r_{i,j}}) - \exp(-\lambda t_{i,r_{i,j+1}})]^{r_{i,j+1}-r_{i,j}-1} \times \\ \lambda^{k_i} \exp\{-\lambda [\sum_{j=1}^{k_i} t_{i,r_{i,j}} + (n_i - r_i) \tau_i]\} \cdot [\exp(-\lambda t_{i,r_{i,k_i}}) - \exp(-\lambda \tau_i)]^{r_{i,k_i}-r_{i,k_i}-1},$$

而 $[\exp(-\lambda t_{i,r_{i,k_i}}) - \exp(-\lambda \tau_i)]^{r_{i,k_i}-r_{i,k_i}-1}$ 可化^[7] 为

$$[\exp(-\lambda t_{i,r_{i,k_i}}) - \exp(-\lambda t_{i,r_{i,k_i+1}})]^{r_{i,k_i+1}-r_{i,k_i}-1}.$$

令 $r_i = r_{i,k_i+1} - 1$, 则可归并到 $\prod_{j=0}^{k_i-1} [\exp(-\lambda t_{i,r_{i,j}}) - \exp(-\lambda t_{i,r_{i,j+1}})]^{r_{i,j+1}-r_{i,j}-1}$ 项中. 因此, 上式可化为

$$L_i(\lambda \mid t_{i,r_{i,1}}, \dots, t_{i,r_{i,k_i}}) = C \prod_{j=0}^{k_i} [\exp(-\lambda t_{i,r_{i,j}}) - \exp(-\lambda t_{i,r_{i,j+1}})]^{r_{i,j+1}-r_{i,j}-1} \times \\ \lambda^{k_i} \exp\left\{-\lambda \left[\sum_{j=1}^{k_i} t_{i,r_{i,j}} + (n_i - r_i) \tau_i\right]\right\}.$$

设 $m_{i,j} = r_{i,j+1} - r_{i,j} - 1$, $t^{(i)} = \sum_{j=1}^{k_i} t_{i,r_{i,j}} + (n_i - r_i) \tau_i$. 由于在各应力水平下的试验是相互独立的, 故此试验数据的似然函数是

$$L = \prod_{i=1}^l L_i = \prod_{i=1}^l \prod_{j=0}^{k_i} [\exp(-\lambda t_{i,r_{i,j}}) - \exp(-\lambda t_{i,r_{i,j+1}})]^{m_{i,j}} \prod_{i=1}^l \lambda^{k_i} \exp\left\{-\sum_{i=1}^l \lambda t^{(i)}\right\}. \quad (2)$$

Bayes 分析中至关重要的一步是先验分布的选取. 其通常取决于分布(似然函数)的具体形式及样本(历史数据)关于参数的信息. 参见文[8]提出的准则, 取参数 λ , $i = 1, \dots, l$. 即有

$$\pi(\lambda, \dots, \lambda) \propto \prod_{i=1}^l \frac{1}{\lambda}.$$

为使问题一般化, 可采用理论和工程常用的伽马分布 $\Gamma(\alpha, \beta)$, 即

$$g(\lambda; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} \exp(-\beta \lambda), \quad i = 1, 2, \dots, l.$$

作为应力 s_i 下参数 λ 的先验分布, 并假定分布间相互独立, 从而参数的联合先验分布为

$$g(\lambda, \alpha, \beta) = \prod_{i=1}^l \frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)} \lambda_i^{\alpha_i-1} \exp(-\beta_i \lambda_i), \quad \lambda \in H_l.$$

式中: $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_l)$, $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_l)$, $\beta = (\beta_1, \beta_2, \dots, \beta_l)$. 由此可得到参数的联合后验分布为

$$g(\lambda \mid \alpha, \beta, t_{i,r_{i,1}}, \dots, t_{i,r_{i,k_i}}) \propto \prod_{i=1}^l \prod_{j=0}^{k_i} [\exp(-\lambda t_{i,r_{i,j}}) - \exp(-\lambda t_{i,r_{i,j+1}})]^{m_{i,j}} \times \\ \prod_{i=1}^l \lambda_i^{\alpha_i+k_i-1} \exp\left\{-\sum_{i=1}^l \lambda_i [t^{(i)} + \beta_i]\right\}$$

其中: $\lambda \in H_l$, $\alpha_i > 0$, $\beta_i > 0$.

先求 λ , $i = 1, 2, \dots, l$ 的边际后验分布为

$$f(\lambda \mid \alpha, \beta, t_{i,r_{i,1}}, \dots, t_{i,r_{i,k_i}}) \propto \prod_{j=0}^{k_i} [\exp(-\lambda t_{i,r_{i,j}}) - \exp(-\lambda t_{i,r_{i,j+1}})]^{m_{i,j}} \times$$

$$\lambda^{k+\alpha-1} \exp\{-\lambda[t^{(\beta)} + \beta]\}.$$

此时分布的形式复杂, 且是一个非“标准”分布, 直接抽样有难度. 因此, 可以将所有的观察样本 y 看作是一个未知“参数”, 即 $y_{i,r_{ij}} = (y_{i,r_{ij},1}, \dots, y_{i,r_{ij},m_{ij}})$, $i = 1, \dots, l, j = 0, 1, \dots, k_i; y = (y^{1,r_{1,0}}, y^{1,r_{1,1}}, y^{1,r_{1,2}}, \dots, y^{1,r_{1,k_1}}, \dots, y^{l,r_{l,0}}, y^{l,r_{l,1}}, \dots, y^{l,r_{l,k_l}})$. 根据 Gibbs 抽样方法, 求 y 的全条件分布, 则有

$$f(y|\lambda, \alpha, \beta, \tau) \propto \prod_{i=1}^l \left\{ \prod_{j=0}^{k_i} [f(y_{i,r_{ij}}) \prod_{s=1}^{m_{ij}} f(y_{i,r_{ij},s}) / [F(t_{i,r_{ij+1}}) - F(t_{i,r_{ij}})]^{m_{ij}}] \times I(t_{i,r_{ij}} \leq y_{i,r_{ij},1} \leq \dots \leq y_{i,r_{ij}+m_{ij}} \leq t_{i,r_{ij+1}}) \right\}. \tag{3}$$

为获得 y 的样本, 依次从失效率 λ 定义在区间 $[t_{i,r_{ij}}, t_{i,r_{ij+1}}]$ 上的截断型指数分布 $\exp(-\lambda \cdot)$, 抽取 m_{ij} 个有序样本 $y_{i,r_{ij}} = (y_{i,r_{ij},1}, \dots, y_{i,r_{ij},m_{ij}})$. 对于变量 $y_{i,r_{ij}}$ 的分布可看成是 $t_{i,r_{ij}}$ 的退化分布, 取值为 $t_{i,r_{ij}}$, 这样就完成了 y 的抽样. 其中, 对于定义在区间 $[t_{i,r_{ij}}, t_{i,r_{ij+1}}]$ 的指数分布 $\exp(-\lambda \cdot)$ 的一个样本可按下式获得. 即

$$f(y|\lambda, \alpha, \beta, \tau) \propto \prod_{i=1}^l \left\{ \prod_{j=0}^{k_i} [f(y_{i,r_{ij}}) \prod_{s=1}^{m_{ij}} f(y_{i,r_{ij},s}) / [F(t_{i,r_{ij+1}}) - F(t_{i,r_{ij}})]^{m_{ij}}] \times I(t_{i,r_{ij}} \leq y_{i,r_{ij},1} \leq \dots \leq y_{i,r_{ij}+m_{ij}} \leq t_{i,r_{ij+1}}) \right\} = \prod_{i=1}^l \left\{ \lambda^{\alpha_i} \prod_{j=0}^{k_i} \frac{\exp(-\lambda \sum_{s=1}^{r_{ij+1}} y_{i,j,s}) I(y_{i,r_{ij}} \in [t_{i,r_{ij}}, t_{i,r_{ij+1}}])}{[\exp(-\lambda t_{i,r_{ij}}) - \exp(-\lambda t_{i,r_{ij+1}})]^{r_{ij+1}}} \right\}. \tag{4}$$

因此, 所获得 $\lambda, i = 1, 2, \dots, l$ 的完全条件分布为

$$f(\lambda|\alpha, \beta, y) \propto \prod_{j=0}^{k_i} [\exp(-\lambda t_{i,r_{ij}}) - \exp(-\lambda t_{i,r_{ij+1}})]^{m_{ij}} \lambda^{k+\alpha-1} \exp\{-\lambda[t^{(\beta)} + \beta]\}. \tag{5}$$

式(5)中: $\lambda \in S_i, S_i$ 为 H_1 沿 λ 方向的横截面, 具有区间形式 $\lambda^L < \lambda < \lambda^U, S_1 = \{0 < \lambda < \lambda_1\}, S_i = \{\lambda_{i-1} < \lambda < \lambda_{i+1}\}, S_l = \{\lambda_{l-1} < \lambda < +\infty\}$. 因此 $f(\lambda|\alpha, \beta)$ 为区间 (λ^L, λ^U) 上的截断伽玛分布, 其样本可由

$$\lambda = G_i^{-1} [G_i(\lambda^L) + u(G_i(\lambda^U) - G_i(\lambda^L))]$$

获得, 其中 u 为 $(0, 1)$ 上均匀分布随机数, $G_i(\cdot)$ 为伽玛分布 $T(\lambda|n_i + \alpha_i, \sum_{j=1}^{n_i} t_{i,j} + \beta_i)$. 类似地不难获得 $(\lambda_1, \lambda_2, \dots, \lambda_l)$ 的 q 个独立同分布的 Gibbs 样本 $(\lambda_p^{(1)}, \dots, \lambda_p^{(q)}), p = 1, \dots, q$, 其中 t 表示达到收敛后 Gibbs 样本的迭代次数. 利用上述 Gibbs 样本, 求得 $\lambda(i = 1, \dots, l)$ 的边际密度估计(Rao-Blackwell)为

$$f(\lambda) = \frac{1}{q} \sum_{p=1}^q f(\lambda_p|\alpha, \beta).$$

由于有

$$E(\lambda, |\alpha, \beta) = \int_{\lambda^L}^{\lambda^U} \lambda f(\lambda_p|\alpha, \beta) d\lambda = \frac{G_i(\lambda^U|\alpha'_i + 1, \beta_i) - G_i(\lambda^L|\alpha'_i + 1, \beta_i)}{G_i(\lambda^U|\alpha_i, \beta_i) - G_i(\lambda^L|\alpha_i, \beta_i)},$$

其中: $\alpha'_i = n_i + \alpha, \beta_i = \sum_{j=1}^{n_i} t_{i,j} + \beta_i$. 因此, λ 的 Bayes 估计为

$$\lambda = \frac{1}{q} \sum_{p=1}^q E(\lambda_p|\alpha, \beta). \tag{6}$$

参数 $\lambda_1, \dots, \lambda_l$ 的另一种形式的 Bayes 估计为 Gibbs 样本均值, 即有

$$\bar{\lambda} = \frac{1}{q} \sum_{p=1}^q \lambda_p^{(t)}, \quad i = 1, \dots, l.$$

如果应力 S_i 与 λ 满足线性关系, 即 $\ln(1/\lambda) = a + b\phi(s_i)$, 则不难获得参数 a, b 的最小二乘估计 \hat{a}, \hat{b} . 因此, 可以获得正常应力下失效率 λ 的 Bayes 估计为 $\lambda_0 = \exp(\hat{a} + \hat{b}\phi(s_0))$.

3 Monte-Carlo 模拟比较

为了考察不同类型的微型电机(寿命参数 λ 不同)在各应力水平下的可靠性特征, 参数值分别取为

$\lambda_1=5.2\times 10^{-5}$, $\lambda_2=4.4\times 10^{-4}$, $\lambda_3=2.5\times 10^{-4}$, $\lambda_4=1.0\times 10^{-4}$, 各应力水平下分别投入 50 个产品参加试验, 直到全部失效为止. 在数据缺失场合下, 假定应力 s_i 只记录到 k_i 个失效数据, $0=t_{i,0}<t_{i,1}<\dots<t_{i,k_i}\leqslant \tau$, $i=1,2,3,4$. 为考察对 Bayes 估计及近似 Bayes 的效果, 把它们与极大似然估计作 Monte-Carlo 模拟比较.

参数估计值的相对偏差 $\frac{\text{Bias}(\hat{\lambda}_i)}{\lambda_i}$ 及其相对均方误差 $(\frac{\text{MSE}(\hat{\lambda}_i)}{\lambda_i^2})$, 结果如表 1 所示. 另外为了方便比较, 参数 $\lambda_1, \dots, \lambda_4$ 的先验分布取为无信息分布和有信息先验分布 $\Gamma(5, 2.5\times 10^5)$. 不同场合下的试验结果表明, 各场合存在先验信息的情况下, Bayes 估计的相对偏差和相对均方误差都小于极大似然估计; 而对于没有先验信息的情况, Bayes 估计跟极大似然估计的效果差不多.

表 1 参数估计值及其相对均方误差表

Tab. 1 Relativities of the bias and MSE for the estimation

参数	$\frac{\text{Bias}(\hat{\lambda}_1)}{\lambda_1}$	$\frac{\text{MSE}(\hat{\lambda}_1^2)}{\lambda_1^2}$	$\frac{\text{Bias}(\hat{\lambda}_2)}{\lambda_2}$	$\frac{\text{MSE}(\hat{\lambda}_2^2)}{\lambda_2^2}$	$\frac{\text{Bias}(\hat{\lambda}_3)}{\lambda_3}$	$\frac{\text{MSE}(\hat{\lambda}_3^2)}{\lambda_3^2}$	$\frac{\text{Bias}(\hat{\lambda}_4)}{\lambda_4}$	$\frac{\text{MSE}(\hat{\lambda}_4^2)}{\lambda_4^2}$
Bayes 估计(有先验)	0.021 0	0.040 4	0.035 1	0.047 2	0.026 8	0.054 1	0.045 2	0.056 7
Bayes 估计(无先验)	0.026 1	0.057 8	0.065 7	0.076 8	0.039 7	0.056 4	0.067 5	0.078 3
极大似然估计	0.024 5	0.041 7	0.036 7	0.048 6	0.027 1	0.058 3	0.046 4	0.058 9

参考文献:

[1] 茆诗松, 王玲玲. 加速寿命试验[M]. 北京: 科学出版社, 1997.

[2] NELSON W B. Accelerated testing: Statistical models, test plans and data analysis[M]. New York: John Wiley & Sons, 1990.

[3] 王乃生, 王玲玲. 定数截尾数据缺失场合下指数分布参数的 Bayes 估计[J]. 应用概率统计, 2001, 8(3): 229-235.

[4] BALAKRISHNAN N. On the maximum likelihood estimation of the location and scale parameters of exponential distribution based on multiply type-II censored samples[J]. J Appl Statist, 1990, 17(1): 55-61.

[5] 王乃生, 王玲玲. 恒定应力加速寿命试验数据缺失时的统计分析[J]. 华东师范大学学报: 自然科学版, 2002, 3(1): 35-44.

[6] 茆诗松, 王静龙, 濮晓龙, 等. 高等数理统计[M]. 北京: 高等教育出版社, 2002: 35-38

[7] 田霆, 刘次华. 定时截尾缺失数据下指数分布的统计推断[J]. 华侨大学学报: 自然科学版, 2006, 27(1): 20-23.

[8] 顾龙全, 周晓东, 汤银才, 等. 指数分布场合恒加试验缺失数据的 Bayes 统计分析[J]. 高校应用数学学报: A 辑, 2006, 21(2): 183-190.

Bayes Analysis of Parameter of Exponential Distribution
under Constant Stress Accelerated Life Testing
and Multiply Type-I Censoring

TIANG Ting

(School of Mathematical Sciences, Huaqiao University, Quanzhou 362021, China)

Abstract: This paper discusses the Bayesian statistical analysis of the parameter of exponential distribution under constant stress accelerated life testing and multiply type-I censoring. The computation of the complicated post-marginal distributions involved in the Bayesian using Gibbs sampling iteration algorithm is solved. By the Monte-Carlo simulation, this method is feasible.

Keywords: exponential distribution; constant stress; accelerated life; type-I censoring; data missing; Bayesian analysis

(责任编辑: 陈志贤 英文审校: 张金顺, 黄心中)