应用概率统计　第 38 卷
第 5 期　2022 年 10 月

Chinese Journal of Applied Probability and Statistics
Oct., 2022, Vol. 38, No. 5, pp. 633-646
doi: 10.3969/j.issn.1001-4268.2022.05.001

双定时混合截尾下两参数 Pareto 分布的统计推断 *

龙　兵*
(荆楚理工学院数理学院, 荆门, 448000)

张忠占
(北京工业大学理学部, 北京, 100124)

摘　要: 基于本文中提出的新截尾试验方案 — 双定时混合截尾, 得到了两参数 Pareto 分布参数的极大似然估计, 根据 Fisher 信息量推导出参数 θ 的渐近置信区间. 当 α 已知时, 取 Gamma 先验分布的情况下, 求出了不同损失函数下参数 θ 的 Bayes 和 E-Bayes 估计以及可靠度函数的 Bayes 估计. 当 α, θ 都未知时, 取联合无信息先验分布, 在平方误差损失函数下计算出 α, θ 的 Bayes 估计. 利用蒙特卡洛方法模拟出双定时混合截尾样本, 得到了未知参数及可靠度函数的估计, 并计算出相对误差, 随着样本量的增加相对误差和置信区间的长度逐渐减小. 最后对一个数值例子进行了统计分析.

关键词: 两参数 Pareto 分布; 双定时混合截尾; 先验分布; Bayes 估计; E-Bayes 估计

中图分类号: O213.2

英文引用格式: LONG B, ZHANG Z Z. Statistical inference of two-parameter Pareto distribution under double type-I hybrid censoring scheme [J]. Chinese J Appl Probab Statist, 2022, 38(5): 633–646. (in Chinese)

§1. 引　　言

在生存分析和可靠性工程等领域中, 截尾寿命试验是一类应用广泛的试验方法. 常见的截尾试验有定时截尾、定数截尾、逐步增加的定数截尾及逐步增加首失效截尾等, 基于这几类试验数据的统计理论比较成熟研究成果较多, 国内外很多学者基于这类数据对一些寿命分布总体进行了统计推断, 如文献 [1–8]. 随着基础理论的发展和实际应用中的需要, 近年来出现了一种混合截尾试验方案, 即定时截尾试验和定数截尾试验的混合, 选择试验终止的时间为 $T_1=\min(X_{r:n},T)$, 其中 $X_{r:n}$ 为第 r 个元件的失效时刻, T 为指定的截尾时刻, 这样获得的数据称为混合 I 型截尾数据. 在这个方案中可以保证试验时间不超过 T, 但是可能得到较少的失效数据, 从而影响统计推断的准确性. 如果选择试验的终止时间为 $T_2=\max(X_{r:n},T)$, 这样获得的数据称为混合 II 型截尾数据. 此方案能保证得到至少 r 个失效数据, 但是不能控制试验的终止时刻, 尤其对一些高可靠性的产品. 在此基础上 Kundu 和 Joarder [9] 与 Childs 等 [10] 分别提出了逐步 I 型混合截尾试验和逐步 II 型混合截尾试验. 相关的研究成果如文献 [9–14]. 在定时截尾试验中, 如果在预先确定的试验终止时

*国家自然科学基金项目 (批准号: 11771032) 和荆楚理工学院科研团队项目 (批准号: TD202006) 资助.

*通讯作者, E-mail: qh-longbing@163.com.

本文 2020 年 6 月 30 日收到, 2021 年 4 月 13 日收到修改稿.

刻失效样品数太少会影响到统计推断的精度. 当我们遇到新研发的产品, 由于对该产品的性能没有先验信息可以利用, 可能会设计一个较短的截尾时间, 从而导致得到很少的失效数据. 为了改善估计的精度, 在试验时间和经费允许的条件下可以考虑适当延长试验时间, 重新确定试验的终止时刻, 以利于得到较多的失效数据从而了解新产品的准确信息. 因此在本文中首次提出了一种新的寿命试验方案, 即两个定时截尾试验的混合, 在这里被称为双定时混合截尾试验, 具体的方案将在第 2 节中介绍.

Pareto 分布最初是用来分析社会经济和一些自然现象, 目前该分布模型被广泛应用于生存分析、可靠性理论等多方面. 该分布的分布函数、概率密度函数和可靠度函数分别为

$$F(x;\alpha,\theta)=1-\alpha^{\theta}x^{-\theta},\qquad x\geqslant\alpha,\tag{1}$$

$$f(x;\alpha,\theta)=\theta\alpha^{\theta}x^{-(\theta+1)},\qquad x\geqslant\alpha,\tag{2}$$

$$R(x;\alpha,\theta)=\alpha^{\theta}x^{-\theta},\qquad x\geqslant\alpha,\tag{3}$$

式中 $\alpha>0$, $\theta>0$, 分别被称为尺度参数和形状参数.

本文假设元件的寿命服从两参数 Pareto 分布 (1), 在双定时混合截尾试验数据下对未知参数及可靠度函数进行统计分析, 用随机模拟的方法验证了所用统计方法的合理性, 最后对一个真实的数据集进行了分析.

§2. 模型描述

在可靠性试验中双定时混合截尾是可能在 2 个时刻进行截尾的一种试验方案, 其模型描述如下:

假设把 n 个独立同分布的元件投入试验, 正整数 $m<n$ 且 m 事先确定, 截尾时刻为 t_1,t_2, 满足 $0<t_1<t_2$. 设在时刻 t_1 之前的失效元件数为 m_1 个, 如果 $m_1\geqslant m$, 则在时刻 t_1 停止试验, 没有失效的 $n-m_1$ 个元件退出试验, 其中 $0<X_{1:n}\leqslant X_{2:n}\leqslant\cdots\leqslant X_{m_1:n}\leqslant t_1$ 为次序失效时刻; 如果 $m_1<m$, 则在时刻 t_2 停止试验, 此时的失效元件数为 m_2, 没有失效的 $n-m_2$ 个元件退出试验, 其中 $0<X_{1:n}\leqslant X_{2:n}\leqslant\cdots\leqslant X_{m_2:n}\leqslant t_2$ 为次序失效时刻. 显然这里 $m_1\leqslant m_2$ 且都为随机变量. 不失一般性, 假设 $X_{1:n}<t_1$.

在本文中把上述两种情形分别表示为 Case I 和 Case II, 并且称其为双定时混合截尾试验方案, 从而得到如下的观测数据:

$$\text{Case I: }(X_{1:n},X_{2:n},\cdots,X_{m_1:n}),\qquad \text{若 } m_1\geqslant m;$$

$$\text{Case II: }(X_{1:n},X_{2:n},\cdots,X_{m_2:n}),\qquad \text{若 } m_1<m.$$

从上面的描述可以看到, 当试验进行到预先确定的终止时刻 t_1 时, 如果失效数没有达到预期, 那么将根据实际情况对试验进行调整, 重新确定试验的终止时刻为 t_2. 与混合 I 型

截尾及混合 II 型截尾相比较, 双定时混合截尾试验方案可以在控制试验时间的同时兼顾统计推断的精度.

§3. 极大似然估计及渐近置信区间

首先讨论在双定时混合截尾试验下, Pareto 分布中未知参数 α 和 θ 的极大似然估计. 设

$$k=\begin{cases} m_1, & \text{Case I};\\ m_2, & \text{Case II}, \end{cases}\qquad t=\begin{cases} t_1, & \text{Case I};\\ t_2, & \text{Case II}. \end{cases}$$

另外设双定时截尾试验所得观测数据为 $x^*=(x_{1:n},x_{2:n},\cdots,x_{k:n})$.

基于上述观测数据, 省略与未知参数无关的常量, 则在 Case I 和 Case II 两种情形下的似然函数可以表示为

$$\begin{aligned} L(\alpha,\theta)&=\prod_{i=1}^{k} f(x_{i:n})[1-F(t)]^{n-k}I(x_{1:n}\geqslant\alpha)\\ &=\theta^k\alpha^{n\theta}\prod_{i=1}^{k}x_{i:n}^{-(\theta+1)}t^{-(n-k)\theta}I(x_{1:n}\geqslant\alpha), \end{aligned}\tag{4}$$

其中 $I(\cdot)$ 为示性函数. 由于 $L(\alpha,\theta)$ 是关于 α 的单调递增函数, 要使似然函数 $L(\alpha,\theta)$ 的值最大, 需取 $\alpha=x_{1:n}$, 因此得到参数 α 的极大似然估计为

$$\widehat{\alpha}=x_{1:n}.$$

当 $\alpha=x_{1:n}$ 时, 则对数似然函数为

$$\ln L(\theta)=k\ln\theta+n\theta\ln x_{1:n}-(\theta+1)\sum_{i=1}^{k}\ln x_{i:n}-(n-k)\theta\ln t.$$

求对数似然函数关于 θ 的导数得

$$\frac{\mathrm{d}\ln L(\theta)}{\mathrm{d}\theta}=\frac{k}{\theta}+n\ln x_{1:n}-\sum_{i=1}^{k}\ln x_{i:n}-(n-k)\ln t.$$

因此参数 θ 和 α 的极大似然估计分别为

$$\widehat{\theta}=\frac{k}{\sum\limits_{i=1}^{k}\ln x_{i:n}+(n-k)\ln t-n\ln x_{1:n}},\qquad \widehat{\alpha}=x_{1:n}.$$

根据极大似然估计的不变性, 可靠度函数 $R(x)$ 的极大似然估计为

$$\widehat{R}(x)=\widehat{\alpha}^{\widehat{\theta}}x^{-\widehat{\theta}}.$$

Fisher 信息反映总体模型本身所提供的信息量, 常被用来求参数的置信区间.

当 α 已知时, 根据似然函数 (4) 式可以得到

$$\frac{\mathrm{d}\ln L(\theta)}{\mathrm{d}\theta}=\frac{k}{\theta}+n\ln\alpha-\sum_{i=1}^{k}\ln x_{i:n}-(n-k)\ln t,$$

则参数 θ 的极大似然估计为

$$\widehat{\theta}_{\mathrm{M}}=\frac{k}{\sum\limits_{i=1}^{k}\ln x_{i:n}+(n-k)\ln t-n\ln\alpha},$$

进而

$$\frac{\mathrm{d}^2\ln L(\theta)}{\mathrm{d}\theta^2}=-\frac{k}{\theta^2},$$

因此 Fisher 信息量为

$$I(\theta)=-\mathsf{E}\Big[\frac{\mathrm{d}^2\ln L(\theta)}{\mathrm{d}\theta^2}\Big]=\frac{\mathsf{E}(k)}{\theta^2}.$$

由于 $k\sim\mathrm{B}(n,p)$, 则 $\mathsf{E}(k)=np$, 其中 $p=1-\alpha^{\theta}t^{-\theta}$, 于是

$$I(\theta)=\frac{n(1-\alpha^{\theta}t^{-\theta})}{\theta^2}.$$

由于在大样本下参数 θ 的极大似然估计具有渐近正态性, 于是有

$$\widehat{\theta}_{\mathrm{M}}\sim\mathrm{N}(\theta,\,I^{-1}(\widehat{\theta}_{\mathrm{M}})).$$

因此, 当 $0<\gamma<1$ 时, 参数 θ 的置信水平为 $100(1-\gamma)\%$ 的渐近置信区间为 $(\theta_{\mathrm{L}},\,\theta_{\mathrm{U}})$, 其中 $\theta_{\mathrm{L}}=\widehat{\theta}_{\mathrm{M}}-Z_{\gamma/2}I^{-1/2}(\widehat{\theta}_{\mathrm{M}})$, $\theta_{\mathrm{U}}=\widehat{\theta}_{\mathrm{M}}+Z_{\gamma/2}I^{-1/2}(\widehat{\theta}_{\mathrm{M}})$, 这里 $Z_{\gamma/2}$ 表示标准正态分布的上 $\gamma/2$ 分位数.

§4. Bayes 估计

如果有先验信息可以利用, 用 Bayes 方法可以提高估计的精度. 当 α 已知、θ 未知时, 本文在 Gamma 先验分布下给出 θ 及可靠度函数的 Bayes 估计; 当参数 α 和 θ 均未知时, 在联合无信息先验分布下给出了参数 α 和 θ 的 Bayes 估计.

1) Gamma 先验分布下的 Bayes 及 E-Bayes 估计

在这里我们取 θ 的先验分布为 Gamma 分布, 其概率密度函数为

$$\pi(\theta)=\frac{b^a}{\Gamma(a)}\theta^{a-1}\mathrm{e}^{-b\theta},\qquad\theta>0,\tag{5}$$

这里超参数 $a>0$, $b>0$, $\Gamma(\cdot)$ 表示 Gamma 函数.

当 α 已知时, 由 (4)–(5) 式, 根据 Bayes 公式可得到 θ 的后验密度函数为

$$\pi(\theta \,|\, x^*)=\frac{(A+b)^{k+a}}{\Gamma(k+a)}\theta^{k+a-1}\mathrm{e}^{-\theta(A+b)}, \qquad \theta>0, \tag{6}$$

其中 $A=\sum\limits_{i=1}^{k}\ln x_{i:n}+(n-k)\ln t-n\ln\alpha$.

引理 1　设 $y=(y_1,y_2,\cdots,y_n)$ 是来自某总体的样本观测值, θ 为其未知参数, 则有

(i) 在平方损失函数 $L_1(\theta,\delta)=(\theta-\delta)^2$ 下, 未知参数 θ 的 Bayes 估计为

$$\widehat{\theta}_{\mathrm{B1}}=\mathsf{E}(\theta \,|\, y);$$

(ii) 在熵损失函数 $L_2(\theta,\delta)\propto\delta/\theta-\ln(\delta/\theta)-1$ 下, 未知参数 θ 的 Bayes 估计为

$$\widehat{\theta}_{\mathrm{B2}}=[\mathsf{E}(\theta^{-1} \,|\, y)]^{-1};$$

(iii) 在加权平方损失函数 $L_3(\theta,\delta)=\theta^{-2}(\theta-\delta)^2$ 下, 未知参数 θ 的 Bayes 估计为

$$\widehat{\theta}_{\mathrm{B3}}=\frac{\mathsf{E}(\theta^{-1} \,|\, y)}{\mathsf{E}(\theta^{-2} \,|\, y)}.$$

这里 δ 是未知参数 θ 的一个估计.

利用 Bayes 方法求参数的点估计, 通常需要根据先验信息去估计先验分布中的超参数. 为了得到较稳健的 Bayes 估计, 韩明[15] 在 2005 年提出了 E-Bayes 的概念, 即利用超参数的概率分布, 对所得到的 Bayes 估计再求数学期望, 因此也称为期望 Bayes 估计. 具体定义如下:

定义 2　对 $(a,b)\in D$, 若参数 θ 的 Bayes 估计 $\widehat{\theta}_{\mathrm{B}}(a,b)$ 是关于 a,b 的连续函数, 则称

$$\widehat{\theta}_{\mathrm{EB}}=\iint_D\widehat{\theta}_{\mathrm{B}}(a,b)\pi(a,b)\mathrm{d}a\mathrm{d}b$$

是未知参数 θ 的 E-Bayes 估计 (expected Bayesian estimation), 其中 $\pi(a,b)$ 是超参数 a,b 在集合 D 上的联合密度函数.

定理 3　设 $x^*=(x_{1:n},x_{2:n},\cdots,x_{k:n})$ 为来自两参数 Pareto 分布 (1) 的双定时混合截尾数据, 当 α 已知时, 若 θ 的后验分布由 (6) 式给出, 超参数 a, b 的先验密度函数为

$$\pi(a,b)=\frac{1}{c_1c_2}, \qquad 0<a<c_1,\ 0<b<c_2, \tag{7}$$

则有下列结论:

(i) 在平方损失函数下, θ 的 Bayes 估计为 $\widehat{\theta}_{\mathrm{B1}}(a,b)=(k+a)/(A+b)$, θ 的 E-Bayes 估计为

$$\widehat{\theta}_{\mathrm{EB1}}=\frac{1}{c_2}\Big(k+\frac{c_1}{2}\Big)\ln\Big(1+\frac{c_2}{A}\Big);$$

(ii) 在熵损失函数下, θ 的 Bayes 估计为 $\widehat{\theta}_{\mathrm{B2}}(a,b)=(k+a-1)/(A+b)$, θ 的 E-Bayes 估计为

$$\widehat{\theta}_{\mathrm{EB2}}=\frac{1}{c_2}\Big(k-1+\frac{c_1}{2}\Big)\ln\Big(1+\frac{c_2}{A}\Big);$$

(iii) 在加权平方损失函数下, θ 的 Bayes 估计为 $\widehat{\theta}_{\mathrm{B3}}(a,b)=(k+a-2)/(A+b)$, θ 的 E-Bayes 估计为

$$\widehat{\theta}_{\mathrm{EB3}}=\frac{1}{c_2}\Big(k-2+\frac{c_1}{2}\Big)\ln\Big(1+\frac{c_2}{A}\Big).$$

其中 $A=\sum\limits_{i=1}^{k}\ln x_{i:n}+(n-k)\ln t-n\ln\alpha$.

证明: (i) 在平方损失函数下, θ 的 Bayes 估计为其后验分布的均值, 则

$$\begin{aligned}\widehat{\theta}_{\mathrm{B1}}(a,b)&=\int_0^{+\infty}\theta\pi(\theta\,|\,x^*)\mathrm{d}\theta=\frac{(A+b)^{k+a}}{\Gamma(k+a)}\int_0^{+\infty}\theta^{k+a}\mathrm{e}^{-\theta(A+b)}\mathrm{d}\theta\\&=\frac{k+a}{A+b},\end{aligned}$$

因此 θ 的 E-Bayes 估计为

$$\begin{aligned}\widehat{\theta}_{\mathrm{EB1}}&=\int_0^{c_1}\int_0^{c_2}\widehat{\theta}_{\mathrm{B1}}(a,b)\pi(a,b)\mathrm{d}b\mathrm{d}a=\frac{1}{c_1c_2}\int_0^{c_1}\int_0^{c_2}\frac{k+a}{A+b}\mathrm{d}b\mathrm{d}a\\&=\frac{1}{c_2}\Big(k+\frac{c_1}{2}\Big)\ln\Big(1+\frac{c_2}{A}\Big).\end{aligned}$$

(ii)

$$\mathsf{E}(\theta^{-1}\,|\,x^*)=\frac{(A+b)^{k+a}}{\Gamma(k+a)}\int_0^{+\infty}\theta^{k+a-2}\mathrm{e}^{-\theta(A+b)}\mathrm{d}\theta=\frac{A+b}{k+a-1},$$

因此在熵损失函数下, θ 的 Bayes 估计为

$$\widehat{\theta}_{\mathrm{B2}}(a,b)=[\mathsf{E}(\theta^{-1}\,|\,x^*)]^{-1}=\frac{k+a-1}{A+b},$$

θ 的 E-Bayes 估计为

$$\widehat{\theta}_{\mathrm{EB2}}=\frac{1}{c_1c_2}\int_0^{c_1}\int_0^{c_2}\frac{k+a-1}{A+b}\mathrm{d}b\mathrm{d}a=\frac{1}{c_2}\Big(k-1+\frac{c_1}{2}\Big)\ln\Big(1+\frac{c_2}{A}\Big).$$

(iii)

$$\mathsf{E}(\theta^{-2}\,|\,x^*)=\frac{(A+b)^{k+a}}{\Gamma(k+a)}\int_0^{+\infty}\theta^{k+a-3}\mathrm{e}^{-\theta(A+b)}\mathrm{d}\theta=\frac{(A+b)^2}{(k+a-1)(k+a-2)},$$

因此在加权平方损失函数下, θ 的 Bayes 估计为

$$\widehat{\theta}_{\mathrm{B3}}(a,b)=\frac{\mathsf{E}(\theta^{-1}\,|\,x^*)}{\mathsf{E}(\theta^{-2}\,|\,x^*)}=\frac{k+a-2}{A+b},$$

θ 的 E-Bayes 估计为

$$\widehat{\theta}_{\mathrm{EB3}}=\frac{1}{c_1c_2}\int_0^{c_1}\int_0^{c_2}\frac{k+a-2}{A+b}\mathrm{d}b\mathrm{d}a=\frac{1}{c_2}\Big(k-2+\frac{c_1}{2}\Big)\ln\Big(1+\frac{c_2}{A}\Big).$$

定理得证. □

定理 4 设 $x^*=(x_{1:n},x_{2:n},\cdots,x_{k:n})$ 为来自两参数 Pareto 分布 (1) 的双定时混合截尾数据, 当 α 已知时, 若 θ 的后验分布由 (6) 式给出, 则有下列结论:

(i) 在平方损失函数下, 可靠度函数 $R(x)$ 的 Bayes 估计为

$$\widehat{R}_{\mathrm{B1}}(x)=\frac{(A+b)^{k+a}}{(A+b+\ln x-\ln\alpha)^{k+a}}.$$

(ii) 在熵损失函数下, 可靠度函数 $R(x)$ 的 Bayes 估计为

$$\widehat{R}_{\mathrm{B2}}(x)=\frac{(A+b+\ln\alpha-\ln x)^{k+a}}{(A+b)^{k+a}}.$$

(iii) 在加权平方损失函数下, 可靠度函数 $R(x)$ 的 Bayes 估计为

$$\widehat{R}_{\mathrm{B3}}(x)=\frac{(A+b+2\ln\alpha-2\ln x)^{k+a}}{(A+b+\ln\alpha-\ln x)^{k+a}}.$$

其中 $A=\sum\limits_{i=1}^{k}\ln x_{i:n}+(n-k)\ln t-n\ln\alpha$.

证明: (i) 在平方损失下, 可靠度函数 $R(x)$ 的 Bayes 估计为

$$\widehat{R}_{\mathrm{B1}}(x)=\int_0^{+\infty}R(x)\pi(\theta\,|\,x^*)\mathrm{d}\theta=\frac{(A+b)^{k+a}}{(A+b+\ln x-\ln\alpha)^{k+a}}.$$

(ii)

$$\begin{aligned}\mathsf{E}[R^{-1}(x)\,|\,x^*]&=\frac{(A+b)^{k+a}}{\Gamma(k+a)}\int_0^{+\infty}\theta^{k+a-1}\mathrm{e}^{-\theta(A+b+\ln\alpha-\ln x)}\mathrm{d}\theta\\&=\frac{(A+b)^{k+a}}{(A+b+\ln\alpha-\ln x)^{k+a}},\end{aligned}$$

因此在熵损失函数下, 可靠度函数 $R(x)$ 的 Bayes 估计为

$$\widehat{R}_{\mathrm{B2}}(x)=\{\mathsf{E}[R^{-1}(x)\,|\,x^*]\}^{-1}=\frac{(A+b+\ln\alpha-\ln x)^{k+a}}{(A+b)^{k+a}}.$$

(iii)

$$\begin{aligned}\mathsf{E}[R^{-2}(x)\,|\,x^*] &= \frac{(A+b)^{k+a}}{\Gamma(k+a)}\int_0^{+\infty}\theta^{k+a-1}\mathrm{e}^{-\theta(A+b+2\ln\alpha-2\ln x)}\mathrm{d}\theta\\ &= \frac{(A+b)^{k+a}}{(A+b+2\ln\alpha-2\ln x)^{k+a}},\end{aligned}$$

因此在加权平方损失函数下, 可靠度函数 $R(x)$ 的 Bayes 估计为

$$\widehat{R}_{\mathrm{B3}}(x) = \frac{\mathsf{E}[R^{-1}(x)\,|\,x^*]}{\mathsf{E}[R^{-2}(x)\,|\,x^*]} = \frac{(A+b+2\ln\alpha-2\ln x)^{k+a}}{(A+b+\ln\alpha-\ln x)^{k+a}}.$$

定理得证. $\square$

2) α 和 θ 的 Bayes 估计

取参数 α,θ 的联合无信息先验分布为

$$\pi(\alpha,\theta) = \frac{1}{\alpha\theta},\qquad \alpha,\theta>0.$$

根据 Bayes 公式可得参数 α, θ 的联合后验分布为

$$\pi(\alpha,\theta\,|\,x^*) = \frac{n(B-n\ln x_{1:n})^{k-1}}{\Gamma(k-1)}\theta^{k-1}\alpha^{n\theta-1}\mathrm{e}^{-B\theta},\qquad 0<\alpha\leqslant x_{1:n},\ 0<\theta<+\infty, \tag{8}$$

其中 $B=\sum\limits_{i=1}^{k}\ln x_{i:n}+(n-k)\ln t$.

可以得到 θ 的后验分布为 Gamma 分布, 概率密度函数为

$$\pi_1(\theta\,|\,x^*) = \frac{(B-n\ln x_{1:n})^{k-1}}{\Gamma(k-1)}\theta^{k-2}\mathrm{e}^{-\theta(B-n\ln x_{1:n})},\qquad 0<\theta<+\infty. \tag{9}$$

α 的后验分布为

$$\pi_2(\alpha\,|\,x^*) = n(k-1)(B-n\ln x_{1:n})^{k-1}\alpha^{-1}(B-n\ln\alpha)^{-k},\qquad 0<\alpha<x_{1:n}. \tag{10}$$

因此在平方损失函数下, 参数 θ 的 Bayes 估计为

$$\widehat{\theta}_{\mathrm{B}} = \int_0^{+\infty}\theta\pi_1(\theta\,|\,x^*)\mathrm{d}\theta = \frac{k-1}{B-n\ln x_{1:n}}.$$

在平方损失函数下, 参数 α 的 Bayes 估计为

$$\begin{aligned}\widehat{\alpha}_{\mathrm{B}} &= \int_0^{+\infty}\alpha\pi_2(\alpha\,|\,x^*)\mathrm{d}\alpha\\ &= n(k-1)(B-n\ln x_{1:n})^{k-1}\int_0^{x_{1:n}}(B-n\ln\alpha)^{-k}\mathrm{d}\alpha.\end{aligned} \tag{11}$$

由于 (11) 式中的定积分不易求出, 因此 α 的 Bayes 估计没有显式表达式, 但是可以进行数值计算求出近似值. 下面介绍一种采用随机模拟求参数 α 的 Bayes 估计的方法.

给定 θ 的值时, 参数 α 的条件后验分布为

$$\pi(\alpha\,|\,\theta,x^*)=\frac{n\theta}{x_{1:n}^{n\theta}}\alpha^{n\theta-1},\qquad 0<\alpha\leqslant x_{1:n},\tag{12}$$

在平方损失函数下, α 的 Bayes 估计为

$$\widetilde{\alpha}_{\rm B}=\int_0^{x_{1:n}}\alpha\pi(\alpha\,|\,\theta,x^*){\rm d}\alpha=\frac{n\theta}{n\theta+1}x_{1:n}.\tag{13}$$

由于 θ 是未知的, 由 (13) 式并不能求出具体值, 可以借助 θ 的后验分布 (9) 式模拟计算出近似值, 具体步骤如下:

步骤 1: 产生 N 个服从 Gamma 分布 (9) 式的随机数, 记为 $\theta^{(1)},\theta^{(2)},\cdots,\theta^{(N)}$;

步骤 2: 把 $\theta^{(1)},\theta^{(2)},\cdots,\theta^{(N)}$ 代入 (13) 式, 分别计算得到

$$\frac{n\theta^{(i)}}{n\theta^{(i)}+1}x_{1:n},\qquad i=1,2,\cdots,N;$$

步骤 3: 计算出 α 的 Bayes 估计为

$$\widetilde{\alpha}_{\rm B}\approx\frac{1}{N}\sum_{i=1}^{N}\frac{n\theta^{(i)}}{n\theta^{(i)}+1}x_{1:n}.$$

根据参数的后验分布 (9)–(10) 式还可以分别求出 θ 和 α 的可信水平为 $100(1-\gamma)\%$ 的 Bayes 可信区间为 $(\widehat{\theta}_{\rm L},\,\widehat{\theta}_{\rm U})$, $(\widehat{\alpha}_{\rm L},\,x_{1:n})$, 其中 $\widehat{\theta}_{\rm L}$, $\widehat{\theta}_{\rm U}$ 以及 $\widehat{\alpha}_{\rm L}$ 满足

$$\int_0^{\widehat{\theta}_{\rm L}}\pi_1(\theta\,|\,x^*){\rm d}\theta=\int_{\widehat{\theta}_{\rm U}}^{+\infty}\pi_1(\theta\,|\,x^*){\rm d}\theta=\gamma/2,\qquad\int_0^{\widehat{\alpha}_{\rm L}}\pi_2(\alpha\,|\,x^*){\rm d}\alpha=\gamma.$$

对于大多数多参数分布总体而言, 要求出参数或者参数的函数的 Bayes 估计, 通常要计算较复杂的多重积分, 在很多情况下都没有显式表达式, 可以选择利用 Gibbs 抽样法、Lindley 方法和 Tierney-Kadane 方法来计算.

§5.　随机模拟

设 $r_1,r_2,\cdots,r_n$ 是来自均匀分布 U(0,1) 的相互独立随机数, 根据两参数 Pareto 分布的分布函数, 给定参数 α,θ 的值, 利用反函数法, 则 $x_i=\alpha r_i^{-1/\theta}$, $i=1,2,\cdots,n$, 即为来自两参数 Pareto 分布的随机数. 给定 t_1,t_2,m 的值, 则可以得到双定时混合截尾样本. 超参数的值取 $(a,b)=(2,1)$, 且取 $(c_1,c_2)=(3,2)$, 利用所得到的样本进行统计推断. 以上过程重复 10,000 次, 可以得到参数的各种点估计的均值、平均相对误差 (在括号中) 以及参

数 θ 的置信水平为 95% 的置信区间的平均置信下限和上限, 其中相对误差的计算公式为 $\mathrm{RE}(\widehat{\beta})=|\widehat{\beta}-\beta|/\beta$. 模拟计算结果如表 1–4 所示.

表 1　α 已知时, θ 的估计

n	(α,θ)	t_1	t_2	m	$\widehat{\theta}_{\mathrm{M}}$	$\widehat{\theta}_{\mathrm{B1}}$	$\widehat{\theta}_{\mathrm{B2}}$	$\widehat{\theta}_{\mathrm{B3}}$	θ_{L}	θ_{U}
20	(5,2)	8	9	11	2.0897	2.0665	1.9272	1.7879	0.9470	3.2323
					(0.2242)	(0.1918)	(0.1917)	(0.2077)		
	(8,3)	11	14	12	3.1865	2.9412	2.7580	2.5748	1.5136	4.8593
					(0.2162)	(0.1680)	(0.1780)	(0.2025)		
30	(5,2)	8	9	18	2.0610	2.0493	1.9543	1.8592	1.1464	2.9757
					(0.1840)	(0.1660)	(0.1638)	(0.1712)		
	(8,3)	11	14	19	3.1399	2.9812	2.8559	2.7306	1.8155	4.4643
					(0.1797)	(0.1476)	(0.1483)	(0.1583)		
50	(5,2)	8	9	30	2.0457	2.0406	1.9814	1.9222	1.3410	2.7505
					(0.1408)	(0.1323)	(0.1301)	(0.1327)		
	(8,3)	11	14	31	3.0915	2.9977	2.9181	2.8384	2.0731	4.1099
					(0.1348)	(0.1189)	(0.1188)	(0.1232)		

表 2　α 已知时, θ 的 E-Bayes 估计

n	(α,θ)	t_1	t_2	m	$\widehat{\theta}_{\mathrm{EB1}}$	$\widehat{\theta}_{\mathrm{EB2}}$	$\widehat{\theta}_{\mathrm{EB3}}$
20	(5,2)	8	9	11	2.0111	1.8708	1.7305
					(0.1918)	(0.1983)	(0.2203)
	(8,3)	11	14	12	2.8862	2.7007	2.5151
					(0.1729)	(0.1874)	(0.2159)
30	(5,2)	8	9	18	2.0083	1.9130	1.8176
					(0.1647)	(0.1667)	(0.1782)
	(8,3)	11	14	19	2.9358	2.8097	2.6836
					(0.1484)	(0.1525)	(0.1656)
50	(5,2)	8	9	30	2.0135	1.9542	1.8949
					(0.1310)	(0.1312)	(0.1358)
	(8,3)	11	14	31	2.9647	2.8849	2.8050
					(0.1189)	(0.1206)	(0.1268)

表 3　α 已知时, 可靠度函数 $R(x)$ 的估计 ($x=9$)

n	(α,θ)	t_1	t_2	m	$\widehat{R}(x)$	$\widehat{R}_{\mathrm{B1}}(x)$	$\widehat{R}_{\mathrm{B2}}(x)$	$\widehat{R}_{\mathrm{B3}}(x)$
20	(5,2)	8	9	11	0.2950	0.3219	0.2934	0.2631
					(0.2566)	(0.2934)	(0.2263)	(0.2613)
	(8,3)	11	14	12	0.7096	0.7113	0.7059	0.7003
					(0.0843)	(0.0581)	(0.0587)	(0.0600)
30	(5,2)	8	9	18	0.2993	0.3176	0.2979	0.2774
					(0.2118)	(0.1852)	(0.1938)	(0.2176)
	(8,3)	11	14	19	0.7072	0.7079	0.7042	0.7005
					(0.0671)	(0.0511)	(0.0520)	(0.0532)
50	(5,2)	8	9	30	0.3023	0.3136	0.3012	0.2885
					(0.1623)	(0.1485)	(0.1536)	(0.1659)
	(8,3)	11	14	31	0.7031	0.7044	0.7020	0.6996
					(0.0496)	(0.0415)	(0.0421)	(0.0428)

表 4　α 和 θ 都未知时的估计

n	(α,θ)	t_1	t_2	m	$\widehat{\alpha}$	$\widehat{\theta}$	$\widehat{\alpha}_{\mathrm{B}}$	$\widehat{\theta}_{\mathrm{B}}$
20	(5,2)	8	9	11	5.1326	2.2649	5.0075	2.0866
					(0.0265)	(0.2598)	(0.0175)	(0.2338)
	(8,3)	11	14	12	8.1542	3.4593	8.0205	3.2133
					(0.0193)	(0.2578)	(0.0136)	(0.2244)
30	(5,2)	8	9	18	5.0885	2.1599	5.0052	2.0430
					(0.0177)	(0.2055)	(0.0121)	(0.1961)
	(8,3)	11	14	19	8.1137	3.2475	8.0245	3.0740
					(0.0142)	(0.2054)	(0.0102)	(0.1952)
50	(5,2)	8	9	30	5.0585	2.0950	5.0084	2.0267
					(0.0117)	(0.1537)	(0.0080)	(0.1490)
	(8,3)	11	14	31	8.0683	3.1450	8.0149	3.0435
					(0.0085)	(0.1534)	(0.0066)	(0.1489)

从表 1–4 中可以看出: α 和 θ 的 Bayes 估计都要优于相应的极大似然估计, 体现了 Bayes 估计的优点; 在不同的参数真值之下, 随着 n 的增大, 参数及可靠度的估计的平均相

对误差逐渐减小, 参数 θ 的置信区间的平均长度逐渐变小, 体现了参数估计的大样本性质.

§6. 例　　子

下面分析文献 [16] 中第 3.1 节中的例子, 其中的失效数据服从 II 型 Pareto 分布, 经过变换后可以得到来自两参数 Pareto 分布 (1) 的失效数据, 按从小到大的顺序排列如下:

$$0.5009, 0.5040, 0.5142, 0.5221, 0.5261, 0.5418, 0.5473, 0.5834, 0.6091, 0.6252,$$
$$0.6404, 0.6498, 0.6750, 0.7031, 0.7099, 0.7168, 0.7918, 0.8465, 0.9035, 1.1143.$$

已知 $\alpha=0.5$, 当取 $c_1=3$, $c_2=1$, $a=0.8$, $b=0.5$ 时利用前面所得到的结论, 计算出形状参数 θ 的极大似然估计、E-Bayes 估计以及可靠度的估计, 所得结果列于表 5 中.

表 5　形状参数 θ 及可靠度的估计 ($x=0.6$)

t_1	t_2	m	$\widehat{\theta}_{\mathrm{M}}$	$\widehat{\theta}_{\mathrm{EB1}}$	$\widehat{\theta}_{\mathrm{EB2}}$	$\widehat{\theta}_{\mathrm{EB3}}$	$\widehat{R}_{\mathrm{B1}}(x)$	$\widehat{R}_{\mathrm{B2}}(x)$	$\widehat{R}_{\mathrm{B3}}(x)$
0.54	0.58	5	3.8871	3.7385	3.1634	2.5882	0.5691	0.5355	0.4967
		7	3.1361	3.1467	2.7765	2.4063	0.6042	0.5835	0.5606
0.55	0.59	6	4.5407	4.2497	3.7498	3.2498	0.5131	0.4821	0.4471
		8	3.2746	3.2596	2.9165	2.5734	0.5892	0.5696	0.5481
0.61	0.63	7	3.1672	3.1660	2.8645	2.5630	0.5941	0.5770	0.5585
		10	3.1360	3.1368	2.8641	2.5913	0.5939	0.5784	0.5618
0.64	0.66	9	2.9884	3.0069	2.7454	2.4840	0.6064	0.5919	0.5762
		11	3.3257	3.3024	3.0578	2.8132	0.5736	0.5593	0.5441
0.68	0.78	12	3.3857	3.3561	3.1247	2.8932	0.5667	0.5531	0.5386
		15	3.5416	3.4993	3.2993	3.0994	0.5490	0.5370	0.5243

§7. 小　　结

两参数 Pareto 分布作为可靠性工程中的一种重要的分布, 近年来受到许多统计学者的广泛关注. 本文提出了一类新的截尾寿命试验方案, 并在此类截尾数据下讨论了参数及可靠度函数的估计问题. 利用频率方法计算出极大似然估计值, 根据 Bayes 方法, 在不同损失函数和先验分布下得到了参数和可靠度函数的估计值, 并把估计值与真值进行比较, 计算出平均相对误差. 随机模拟结果表明参数的 Bayes 估计要优于极大似然估计, 文中所采用

的先验分布是合理的. 在本文的研究基础之上, 可以提出另一种新的截尾试验方案 — 双定数混合截尾试验, 将在另一篇论文中进行讨论.

参 考 文 献

[1] HASSAN H M, EL-SHAARAWI A H. Likelihood inference for pollutant loading under type I censoring [J]. *Environ Monit Assess*, 2020, **192(4)**: 225 (16 pages).

[2] HAN M. E-Bayesian estimation and its E-posterior risk of the exponential distribution parameter based on complete and type I censored samples [J]. *Comm Statist Theory Methods*, 2020, **49(8)**: 1858–1872.

[3] MONDAL S, KUNDU D. On the joint Type-II progressive censoring scheme [J]. *Comm Statist Theory Methods*, 2020, **49(4)**: 958–976.

[4] NAGATSUKA H, BALAKRISHNAN N. A consistent method of estimation for the three-parameter lognormal distribution based on Type-II right censored data [J]. *Comm Statist Theory Methods*, 2016, **45(19)**: 5693–5708.

[5] REZAPOUR M, ALAMATSAZ M H, BALAKRISHNAN N. Distribution of the number of observations greater than the ith dependent progressively Type-II censored order statistic and its use in goodness-of-fit testing [J]. *Comm Statist Theory Methods*, 2015, **44(12)**: 2517–2529.

[6] BASAK I, BALAKRISHNAN N. A note on the prediction of censored exponential lifetimes in a simple step-stress model with Type-II censoring [J]. *Calcutta Stat Assoc Bull*, 2018, **70(1)**: 57–73.

[7] BASAK I, BALAKRISHNAN N. Prediction of censored exponential lifetimes in a simple step-stress model under progressive Type II censoring [J]. *Comput Statist*, 2017, **32(4)**: 1665–1687.

[8] ZHU X J, BALAKRISHNAN N, FENG C Z, et al. Exact likelihood-ratio tests for joint type-II censored exponential data [J]. *Statistics*, 2020, **54(3)**: 636–648.

[9] KUNDU D, JOARDER A. Analysis of Type-II progressively hybrid censored data [J]. *Comput Statist Data Anal*, 2006, **50(10)**: 2509–2528.

[10] CHILDS A, CHANDRASEKAR B, BALAKRISHNAN N. Exact likelihood inference for an exponential parameter under progressive hybrid censoring schemes [M] // VONTA F, NIKULIN M, LIMNIOS N, et al. (eds) *Statistical Models and Methods for Biomedical and Technical Systems*. Boston, MA: Birkhäuser, 2008: 319–330.

[11] SHAFAY A R, BALAKRISHNAN N. Estimation and prediction for an exponential form distribution based on combined Type-II censored samples [J]. *Bull Malays Math Sci Soc*, 2019, **42(4)**: 1535–1553.

[12] LIN C T, HSU Y Y, LEE S Y, et al. Inference on constant stress accelerated life tests for log-location-scale lifetime distributions with type-I hybrid censoring [J]. *J Stat Comput Simul*, 2019, **89(4)**: 720–749.

[13] BALAKRISHNAN N, SHAFAY A R. One- and two-sample Bayesian prediction intervals based on Type-II hybrid censored data [J]. *Comm Statist Theory Methods*, 2012, **41(9)**: 1511–1531.

[14] 卫超, 师义民. 逐步 II 型混合截尾下 Pareto 分布的统计分析 [J]. *火力与指挥控制*, 2014, **39(4)**: 27–29+33.

[15] 韩明. 证券投资预测的 E-Bayes 方法 [J]. *运筹与管理*, 2005, **14(5)**: 98–102.

[16] NIGM A M, AL-HUSSAINI E K, JAHEEN Z F. Bayesian one-sample prediction of future observations under Pareto distribution [J]. *Statistics*, 2003, **37(6)**: 527–536.

Statistical Inference of Two-Parameter Pareto Distribution under Double Type-I Hybrid Censoring Scheme

LONG Bing

(*School of Mathematics and Physics, Jingchu University of Technology, Jingmen, 448000, China*)

ZHANG Zhongzhan

(*Faculty of Science, Beijing University of Technology, Beijing, 100124, China*)

Abstract: Based on a new lifetime testing scheme proposed in this paper, that is, double type-I hybrid censoring scheme, the maximum likelihood estimates of the parameters are obtained for two-parameter Pareto distribution. Also, the asymptotic confidence interval of θ is obtained from Fisher information. When α is known, the Bayesian and E-Bayesian estimates of θ, as well as the Bayesian estimates of reliability function under different loss functions are obtained using the Gamma prior distribution. When both α and θ are unknown, the joint noninformation prior distribution is taken, and the Bayesian estimates of α and θ are calculated under the squared error loss function. Using Monte Carlo method to simulate the double type-I hybrid censored samples, the estimates of the unknown parameters and reliability function are obtained, with the increase of the sample size, the relative errors and the lengths of the confidence interval decrease gradually. Finally, a numerical example is analyzed.

Keywords: two-parameter Pareto distribution; double type-I hybrid censoring; prior distribution; Bayesian estimation; E-Bayesian estimation

2020 Mathematics Subject Classification: 62F10; 62F15