第31卷第3期 成都信息工程大学学报 Vol. 31 No. 3
2016年6月 JOURNAL OF CHENGDU UNIVERSITY OF INFORMATION TECHNOLOGY Jun. 2016

文章编号：2096-1618(2016)03-0323-04

逐次定数截尾下 Lomax 分布的贝叶斯分析

罗嘉成， 陈勇明
（成都信息工程大学应用数学学院，四川 成都 610225）

摘要：在工程试验中，受经费和时间等条件的限制，一般较难获得完全样本，在逐次定数截尾样本条件下，在对称损失和非对称损失函数下，选择先验分布为无信息先验分布，探讨 Lomax 分布的形状参数和可靠性指标的贝叶斯估计。利用蒙特卡罗模拟方法，比较极大似然估计和两种损失函数下贝叶斯估计的均方误差。模拟结果表明在非对称损失函数下的贝叶斯估计具有更高的精度。

关 键 词：贝叶斯统计；可靠性分析；Lomax 分布；逐次定数截尾；极大似然估计

中图分类号：O213.2 **文献标志码**：A

0 引言

Lomax 分布可视为由指数伽玛分布组合而成的混合分布，由于其分布形式类似于 Pareto 分布，文献[1]将其定义为 II 型 Pareto 分布。在医学试验和工程科学等方面，大多数产品都具有单调递增或单调递减的失效率，这是产品的明显寿命特征，因此该分布在工程试验中占有重要的地位，近年来较多学者关注该分布的统计推断问题。文献[2]利用 EM 算法探讨逐次定数截尾样本下 Lomax 分布的参数估计问题，文献[3]基于混合逐步定数截尾，利用对称和非对称损失函数探讨 Lomax 分布未知参数的贝叶斯估计，以上考虑的是删失样本，而文献[4-5]在完全样本下利用 EM 算法和 Newton-Raphson（牛顿拉弗森）数值方法对 Lomax 分布的参数进行估计，文献[6]在完全样本下利用几种不同的损失函数，对 Lomax 分布形状参数的贝叶斯估计的不同表达式进行研究。

在工程试验中，受诸多试验条件如经费、时间的限制，试验者获得的失效样本很有限，获得的样本通常是截尾样本。定时和定数截尾方案由于简单容易操作而备受工程试验者欢迎，但是这两种截尾方案也有不是那么完美，它们不允许在试验过程中移走受试样品而缺乏灵活性。为进一步节省时间和成本，最近几十年，作为普通截尾的一种推广，逐次截尾样本在工程试验中更常用，在获得逐次截尾样本后可以通过常用的统计方法获得样品的各项可靠性指标，文献[7-11]讨论逐次定数截尾下一些分布参数的估计问题。相对应的，逐次截尾可分为逐次定时和逐次定数截尾，文献[12]对这两种截尾方案的统计分析方法作了详细的介绍。为更符合实际工程应用的需要，在逐次定数截尾模型下，对产品寿命服从 Lomax 分布形状参数和可靠性指标进行估计，利用贝叶斯方法得出估计值，最后利用蒙特卡罗模拟计算方法对估计值的均方误差进行比较分析。

1 模型及假设

称产品的寿命 X 服从 Lomax 分布，若其分布函数、密度函数分别表示如下：

$$F(x)=1-(1+x/\lambda)^{-\theta}$$

$$f(x)=\theta/\lambda(1+x/\lambda)^{-(\theta+1)} \tag{1}$$

易得可靠度和失效率函数[13]分别为

$$R(t)=(1+t/\lambda)^{-\theta},H(t)=\theta/(\lambda+t) \tag{2}$$

由文献[12]知，逐次定数截尾的数学模型如下：假设 n 个样品同时进行试验，并且样品的失效时间均可被检测到，设 $m<n,R_1,R_2,\cdots,R_m$ 是试验前预先设定的非负整数，且满足条件 $R_1+R_2+\cdots+R_{m-1}+R_m+m=n$。记第一个失效样品出现的失效时刻为 $X_{1:m:n}$，与此同时从余下 $n-1$ 个未失效样品中随机移走 R_1 个样品并继续试验；记第二个失效样品出现的失效时刻为 $X_{2:m:n}$，并在余下 $n-2-R_1$ 个样品中随机移走 R_2 个样品并继续试验；以同样的方法继续进行试验直至第 m 个失效样品出现，记失效时刻为 $X_{m:m:n}$ 并移走余下的未失效样本，即 $R_m=n-m-R_1-R_2-\cdots-R_{m-1}$ 个样品，此时停止试验。从以上逐次定数截尾样本的数学模型易知，当 $R_1=R_2=\cdots=R_m=0$ 时，该样本退化为前面介绍的定数截尾样本；当 $R_1=R_2=\cdots=R_{m-1}=0$ 时，相对应的是完全试验样本。

收稿日期：2016-01-19
基金项目：国家社会科学基金资助项目(13BZZ055)；四川省科学技术厅应用基础计划基金资助项目(2015JY0022)

2 参数的极大似然估计

设 $X=(X_{1:m:n},X_{2:m:n},\cdots,X_{m:m:n})$ 是来自上述逐次定数截尾模型的样本,为书写方便,简记 $X_i\equiv X_{i:m:n}$,$i=1,2,\cdots,m$, 则 Lomax 分布下的极大似然函数为

$$\begin{aligned}L&=A\prod_{i=1}^{m}f(x_i)\ [1-F(x_i)]^{R_i}\\&=A\prod_{i=1}^{m}\frac{\theta}{\lambda}(1+x_i/\lambda)^{-(\theta+R_i\theta+1)}\end{aligned}\tag{3}$$

其中:$A=n(n-1-R_1)(n-2-R_1-R_2)\cdots(n-m-R_1-R_2\cdots R_m+1)$ 是与未知参数无关的常数因子。因此上述函数的对数似然函数表示如下:

$$l=\ln L=K+m\ln\theta-m\ln\lambda-\sum_{i=1}^{m}(\theta+R_i\theta+1)\ln(1+x_i/\lambda)\tag{4}$$

且 $\theta>0,X_i\geq\lambda>0,K$ 为正常数。由式(4)可得参数 θ 的极大似然估计的计算公式为

$$\frac{\partial l}{\partial\theta}=\frac{m}{\theta}-\sum_{i=1}^{m}(1+R_i)\ln(1+x_i/\lambda)\tag{5}$$

令 $\dfrac{\partial l}{\partial\theta}=0$,则

$$\hat{\theta}_M=\frac{m}{\sum_{i=1}^{m}(1+R_i)\ln(1+x_i/\lambda)}\tag{6}$$

因此,可靠度和失效率函数的极大似然估计分别如下:

$$\hat{R}(t)=(1+t/\lambda)^{-\hat{\theta}},\ \hat{H}(t)=\hat{\theta}/(\lambda+t)\tag{7}$$

3 无信息先验分布下 θ 和可靠性指标的贝叶斯估计

在对产品无任何经验信息可查时,选取参数的无信息先验分布是常用的手段,因为在参数空间上对参数的选择没有任何偏倚,即:$\pi_1(\theta)=1/\theta$,由贝叶斯公式易得 θ 的后验密度[14]:

$$\begin{aligned}\pi_1(\theta|x,\lambda)&\propto L(x;\theta,\lambda)\pi(\theta)\\&=\frac{\theta^{m-1}\prod_{i=1}^{m}(1+x_i/\lambda)^{-(\theta+R_i\theta)}}{\int_0^{+\infty}\theta^{m-1}\prod_{i=1}^{m}(1+x_i/\lambda)^{-(\theta+R_i\theta)}\mathrm{d}\theta}\end{aligned}\tag{8}$$

令 $\sum_{i=1}^{m}(1+R_i)\ln(1+x_i/\lambda)=q$,则 $\prod_{i=1}^{m}(1+x_i/\lambda)^{-(\theta+R_i\theta)}=e^{-\theta q}$,故有

$$\pi_1(\theta|x,\lambda)=\frac{\theta^{m-1}e^{-\theta q}}{\int_0^{+\infty}\theta^{m-1}e^{-\theta q}\mathrm{d}\theta}=\frac{\theta^{m-1}e^{-\theta q}}{\Gamma(m)}\tag{9}$$

3.1 平方损失函数下参数及可靠性指标的贝叶斯估计

平方损失在贝叶斯统计中是常用的一种对称损失函数,由于其可使计算工作变得简便,因而很受统计学家们欢迎,其表达式为:$L(\theta,\delta)=(\theta-\delta)^2$,其中 δ 是参数 θ 的估计值。由贝叶斯理论易知参数 θ 的贝叶斯估计就是其后验均值。在无信息先验分布下,由(9)式可得形状参数 θ 的贝叶斯估计为

$$\hat{\theta}_{BS}=\int_0^{+\infty}\frac{\theta^m e^{-\theta q}q^m}{\Gamma(m)}\mathrm{d}\theta=\frac{m}{q}\tag{10}$$

因此,可靠度函数 $R(t)$ 以及失效率函数 $H(t)$ 的贝叶斯估计分别为

$$\begin{aligned}\hat{R}_{BS}(t)&=\int_0^{+\infty}\frac{\theta^{m-1}e^{-\theta q}q^m\ (1+t/\lambda)^{-\theta}}{\Gamma(m)}\mathrm{d}\theta\\&=\int_0^{+\infty}\frac{\theta^{m-1}e^{-\theta q}q^m e^{-\theta\omega}}{\Gamma(m)}\mathrm{d}\theta=\left(\frac{q}{q+\omega}\right)^m\end{aligned}\tag{11}$$

$$\hat{H}_{BS}(t)=\frac{\hat{\theta}_{BS}}{\lambda+t}=\frac{m}{q(\lambda+t)}\tag{12}$$

其中 $\omega=\ln(1+\dfrac{t}{\lambda})$。

3.2 Linex 损失函数下参数及可靠性指标的贝叶斯估计

Linex 损失是非对称损失函数,其表达式为:$L(\theta,\delta)=e^{c(\delta-\theta)}-c(\delta-\theta)-1,c\neq0,\delta$ 是 θ 的估计值。由 Linex 损失函数的定义可知,当 $c>0$ 时,过高估计比过低估计带来更严重的损失;而当 $c<0$ 时则反之;当 $c\to0$ 时,Linex 函数退化为对称的平方损失函数。文中只考虑 $c>0$ 的情形,对于 $c<0$ 的情形类似。易知,该损失函数关于 δ 是严凸的,并且在 $\delta=\theta$ 处取得最小值。

因此,Linex 损失函数下形状参数 θ 的估计为

$$\begin{aligned}\hat{\theta}_{BL}&=-c^{-1}\ln[E_\theta(e^{-c\theta})]\\&=-c^{-1}\ln\left[\int_0^{+\infty}e^{-c\theta}\pi_1(\theta|x,\lambda)\mathrm{d}\theta\right]\\&=-c^{-1}m\ln\left(\frac{q}{q+c}\right)\end{aligned}\tag{13}$$

由此,易得可靠度函数 $R(t)$ 以及失效率函数 $H(t)$ 的表达式分别为

$$\begin{aligned}\hat{R}_{BL}(t)&=-c^{-1}\ln[E_R(e^{-cR})]\\&=-c^{-1}\ln\left\{\int_0^{+\infty}\left[\sum_{i=0}^{+\infty}\frac{(-c)^i}{i!}e^{-\theta\omega i}\pi_1(\theta|x,\lambda)\right]\mathrm{d}\theta\right\}\\&=-c^{-1}\ln\left[\sum_{i=0}^{+\infty}\frac{(-c)^i}{i!}\left(\frac{q}{q+\omega i}\right)^m\right]\end{aligned}\tag{14}$$

$$\hat{H}_{BL}(t)=\frac{\hat{\theta}_{BL}}{\lambda+t}=-\frac{m}{c(\lambda+t)}\ln\left(\frac{q}{q+c}\right)\tag{15}$$

4　随机模拟

首先产生 Lomax 分布在逐次定数截尾样本，利用文献[15]给出的算法，逐次定数截尾数据产生步骤如下：

（1）产生 m 个来自均匀分布 $U(0,1)$ 的独立随机样本 $W_1, W_2, \cdots, W_m$；

（2）在预先设定的逐次定数截尾策略 $R_1, R_2, \cdots, R_m$ 下，令 $V_i = W_i^{1/(i+R_m+R_{m-1}+\cdots+R_{m-i+1})}$，$i = 1,2,\cdots,m$；

（3）再令 $U_i = 1-(V_m V_{m-1}\cdots V_{m-i+1})$，$i = 1,2,\cdots,m$，则 $U_1, U_2, \cdots, U_m$ 是来自均匀分布 $U(0,1)$ 的逐次定数截尾数据；

（4）令 $X_i = F^{-1}(U_i)$，$i = 1,2,\cdots,m$，是来自于 Lomax 分布的逐次定数截尾试验数据，其中 $F^{-1}(\cdot)$ 是分布函数的反函数。

下面利用蒙特卡罗模拟方法对参数的估计精度进行比较，其中表 1 为不同截尾策略下的逐次定数截尾样本，表 2 和表 3 为对应截尾样本下参数及可靠性指标的贝叶斯估计的均方误差。

表 1　逐次定数截尾样本

截尾策略	方案 1	方案 2	方案 3
$(R_1, R_2, \cdots, R_m)$	(1,2,0,1,0,2,0,2,0,2)	(1,0,2,0,1,0,2,1,0,0,1,0,2,0,0)	(2,1,0,0,1,0,0,1,0,0,0,2,0,1,0,0,1,0,1,0)

表 2　参数和各指标的均方误差估计（$\lambda=2, t=0.5, c=0.1$）

样本量(n,m)	$\hat\theta_M$	$\hat\theta_{BS}$	$\hat\theta_{BL}$	$\hat R_M$	$\hat R_{BS}$	$\hat R_{BL}$	$\hat H_M$	$\hat H_{BS}$	$\hat H_{BL}$
$n=20, m=10$	0.8542	0.8314	0.1438	2.8640	1.7626	1.2534	2.3505	1.2746	0.8962
$n=25, m=15$	0.6296	0.5263	0.0846	1.3868	1.2739	1.0213	1.9264	1.1468	0.6389
$n=30, m=20$	0.2974	0.1249	0.0524	1.1170	1.1045	1.0134	1.3725	0.6923	0.3573

表 3　参数和各指标的均方误差估计（$\lambda=2.5, t=1.5, c=0.1$）

样本量(n,m)	$\hat\theta_M$	$\hat\theta_{BS}$	$\hat\theta_{BL}$	$\hat R_M$	$\hat R_{BS}$	$\hat R_{BL}$	$\hat H_M$	$\hat H_{BS}$	$\hat H_{BL}$
$n=20, m=10$	0.5381	0.4244	0.1689	2.4680	1.4626	0.5537	2.2505	1.6946	0.6862
$n=25, m=15$	0.3628	0.3463	0.0859	1.3758	1.2739	1.0265	1.6894	1.4368	0.8289
$n=30, m=20$	0.1964	0.1232	0.0579	2.3570	1.2545	1.0135	1.6545	0.6763	0.2473

5　结束语

下面对 Lomax 分布在逐次定数截尾样本下的结果作简要分析：

（1）由表 2 和表 3 知，在无信息先验分布下，由于 Linex 损失是非对称损失函数，考虑错误估计对参数估计带来的损失，因此无论样本量的大小，平方损失函数下的估计和极大似然估计的均方误差都较 Linex 损失函数下估计的均方误差要大。

（2）由表 2 和表 3 知，随着样本量 n 和截尾样本 R 的增加，各个估计的均方误差都相应减小，这也是符合一般统计结论的，特别对于对样本量依赖性更强的极大似然估计效果更明显。

（3）在试验中，可以根据样本容量的差异，选取不同的估计方法，使参数的估计达到满意的精度。样本容量不是很大时，为达到更高的参数精度，用贝叶斯估计比用极大似然估计更适合，这也是符合贝叶斯理论的，因为贝叶斯公式综合了先验信息和样本信息，从而使估计精度更高。

参考文献：

[1]　Abd Ellah A H. Bayesian one sample prediction bounds for the Lomax distribution[J]. Indian Journal of Pure & Applied Mathematics, 2003, 34(1): 101-110.

[2]　Helu A. Estimation on Lomax progressive censoring using the EM algorithm[J]. Journal of Statistical Computation & Simulation, 2013, 85(5): 1035-1052.

[3]　Ma Y, Shi Y. Inference for lomax distribution based on type-II progressively hybrid censored data[J]. Journal of Physical Sciences, 2013, (17): 33-41.

[4]　Yang M, Wei C D, Fan Q Z. Parameter estimation for Lomax distribution under type II censoring[J]. Advanced Materials Research, 2014, 912: 1663-1668.

[5]　Howlader H A, Hossain A M. Bayesian survival es-

timation of Pareto distribution of the second kind based on failure-censored data[J]. Computational statistics & data analysis, 2002, 38(3): 301-314.

[6] 姚惠,谢林.不同损失下Lomax分布形状参数的Bayes估计[J].数学杂志,2012,31(6):1131-1135.

[7] 李凤,师义民.逐步增加Ⅱ型截尾下Pareto分布的Bayes估计[J].数学的实践与认识,2012,42(13):137-142.

[8] 侯华蕾,师义民,李豪亮.双边定数截尾下Pareto分布的可靠性分析[J].数理统计与管理,2009,28(5):826-830.

[9] 杨君慧,师义民,曹弘毅.逐步增加II型截尾试验下广义指数分布的统计分析[J].统计与决策,2014,(16):28-30.

[10] 李凤,师义民.逐次定数截尾下Pareto分布参数的逆矩估计[J].统计与决策,2010,(24):156-157.

[11] 王炳兴. Burr Type XII分布的统计推断[J].数学物理学报,2008,28(6).

[12] Balakrishnan N, Aggarwala R. Progressive Censoring[M]. Boston: Birkhäuser,2000.

[13] Lawless J F. Statistical models and methods for lifetime data[M]. New York: Wiey , 1982.

[14] 茆诗松,汤银才,王玲玲.可靠性统计[M].北京:高等教育出版社,2008.

[15] Sandhu N B R A. A Simple Simulational Algorithm for Generating Progressive Type-II Censored Samples[J]. American Statistician, 1995, 49(2):229-230.

Bayesian Analysis for Lomax Distribution under Progressively Type-II Censored Data

LUO Jia-cheng, CHEN Yong-ming

(College of Applied Mathematics, Chengdu University of Information Technology, Chengdu 610225, China)

Abstract: In life testing, due to budget limits and times constrains, it is sometimes hard to get a complete sample. In this paper, based on progressively Type-II censored sample, choosing the Non-informative prior for the unknown parameter, the Bayes estimators of the shape parameter, failure rate and reliability function have been discussed for the Lomax distribution, under symmetric loss and asymmetric loss function respectively. Finally, a numerical example using Monte Carlo simulation is presented to compare the estimators. The conclusion shows that the Bayes estimators under Linex loss function are better.

Key words: Bayesian statistics; reliability analysis; Lomax distribution; progressively type-II censored samples; maximum likelihood estimation