2013 年 7 月　　重庆师范大学学报(自然科学版)　　Jul. 2013
第 30 卷 第 4 期　　Journal of Chongqing Normal University (Natural Science)　　Vol. 30 No. 4

DOI:10.11721/cqnuj20130411

基于记录值的 GE 分布参数的 Bayes 和经验 Bayes 估计*

王　琪[1]，黄文宜[2]

(1. 电子科技大学 中山学院，广东 中山 528402；2. 宜春学院 数学与计算机科学学院，江西 宜春 336000)

摘要:针对来自广义指数(GE)分布的记录值样本,研究了广义指数分布的参数的估计问题。首先给出了参数的最大似然估计;在参数的共轭先验分布为贝塔先验分布,损失函数为平方误差损失和 LINEX 损失函数情形下,导出了广义指数分布参数的 Bayes 和经验 Bayes 估计。最后进行了 Monte Carlo 数值模拟,对本文提出的几种参数估计进行比较,发现在合适的先验分布条件下 Bayes 和经验 Bayes 估计值更加接近参数真实值,因此在适当的先验分布下 Bayes 估计和经验 Bayes 估计较传统的最大似然估计好。

关键词:Bayes 和经验 Bayes 估计;平方误差损失;LINEX 损失;记录值

中图分类号:O213　　**文献标志码**:A　　**文章编号**:1672-6693(2013)04-0055-04

广义指数(GE)分布作为 Weibull 分布和伽玛分布的一个可选分布,自 Gupta 和 Kunda[1]提出以来,已经在在可靠性、寿命试验等领域得到了广泛的应用。关于该分布的统计推断理论引起了众多学者的兴趣并得到了深入的研究。文献[2]中讨论了广义指数分布参数的不同估计方法,并且比较了它们的优劣性;文献[3]讨论了广义指数分布参数的逆矩估计方法;文献[4]在熵损失函数下讨论了广义指数分布形状参数的 Bayes 估计和可容许性问题;文献[5]在平方误差损失、对数误差平方损失和广义熵损失函数下讨论了广义指数分布参数的 Bayes 估计问题。记录值自提出以来已被广泛应用到诸如天气、水文、地震、机械工程以及体育等诸多领域。关于记录值的统计推断理论引起很多学者的兴趣并得到了广泛的研究。如文献[6]基于记录值和截尾样本讨论了一些寿命分布的 Bayes 估计问题。文献[7]基于记录值样本在平方损失函数下讨论了指数分布参数的 Bayes 估计以及估计的可容许性问题。本文将在平方误差和 LINEX 损失函数下讨论广义指数分布的形状参数 θ 的 Bayes 估计和经验 Bayes 估计问题。

设随机变量 X 为服从参数为 θ 的广义指数分布,其概率密度函数为

$$f(x;\theta)=\theta e^{-x}(1-e^{-x})^{\theta-1} \tag{1}$$

其中 $\theta>0$ 为未知的形状参数。

定义 1[8]　设 $\{X_n,n\geqslant1\}$ 是一个独立同分布的随机变量序列,对于任意的 $n\geqslant1$,定义 $L(1)=1$, $L(n+1)=\min\{j:j>L(n),X_j<X_{L(n)}\}$,则 $\{X_{L(n)}\}$($\{L(n)\}$)称为序列 $\{X_n,n\geqslant1\}$ 的一个下记录值(记录时间)。

1 最大似然估计

设 $X_1,X_2,\cdots$ 为来自广义指数分布的独立同分布的随机样本,设观察到 n 个下记录值为 $X_{L(1)}=x_1,X_{L(2)}=x_2,\cdots,X_{L(n)}=x_n$,记 $R(x;\theta)=-\ln(1-F(x;\theta))$,$t$ 为 $T=-\ln(1-e^{-X_{L(n)}})$ 的观测值,由文献[8]有如下结果。

1) 给定观测值 $x=(x_1,x_2,\cdots,x_n)$ 的参数 θ 的似然函数为

$$l(\theta)=f(x_n;\theta)\prod_{i=1}^{n-1}\frac{f(x_i;\theta)}{F(x_i;\theta)}=\theta^n e^{-t\theta}\cdot(1-e^{x_n})^{-1}\prod_{i=1}^{n}e^{-x_i} \tag{2}$$

取对数后有 $\ln l(\theta)=n\ln\theta-t\theta-\ln(1-e^{x_n})-\sum_{i=1}^{n}x_i$,通过求解对数似然方程 $\frac{d\ln l(\theta)}{d\theta}=0$ 得 θ 的最大似然估计为

* **收稿日期**:2012-07-11　**修回日期**:2012-12-30　**网络出版时间**:2013-07-20 19:23

资助项目:江西省教育厅科技基金(No. GJJ11600)

作者简介:王琪,男,讲师,硕士,研究方向为 Bayes 统计分析,E-mail:wangqirhp@163.com

网络出版地址:http://www.cnki.net/kcms/detail/50.1165.N.20130720.1923.201304.55_008.html

$$\hat{\theta}_{\mathrm{ML}}=\frac{n}{T} \tag{3}$$

2) 第 n 个下记录值 $X_{L(n)}$ 的边缘概率密度函数为

$$f_n(x_n;\theta)=f(x_n;\theta)\frac{R^{n-1}(x_n;\theta)}{(n-1)!}=\theta \mathrm{e}^{-x_n}(1-\mathrm{e}^{-x_n})^{\theta-1}\frac{[-\theta\ln(1-\mathrm{e}^{-x_n})]^{n-1}}{(n-1)!} \tag{4}$$

从而由(4)式有 $T=-\ln(1-\mathrm{e}^{-X_{L(n)}})$ 的概率密度函数为 $f_T(t)=\frac{\theta^n}{\Gamma(n)}t^{n-1}\mathrm{e}^{-\theta t}$。

2 Bayes 和经验 Bayes 估计

下面将考虑在平方误差损失和 LINEX 损失函数下广义指数分布的参数的 Bayes 估计问题。

损失函数作为 Bayes 统计推断程序中是必不可少的。多数 Bayes 统计推断都是基于平方误差损失函数 $L(\hat{\theta},\theta)=(\hat{\theta}-\theta)^2$ 而展开讨论的,一般的平方误差损失函数具有对称性,对高估和低估赋予同样的惩罚。然而在估计可靠性和失效率函数时,高估带来的损失将比低估带来的损失更大,此时应用对称损失函数就不适合了,很多学者提出一些有用的非对称损失函数,其中 LINEX 损失函数为最常使用的非对称损失函数,其表达式为[9]

$$L(\Delta)=\mathrm{e}^{c\Delta}-c\Delta-1,c\neq 0 \tag{5}$$

其中 $\Delta=\hat{\theta}-\theta$,$\hat{\theta}$ 为参数 θ 的估计,c 为形状参数。

在 LINEX 损失(5)下,参数 θ 的 Bayes 估计为 $\hat{\theta}_{\mathrm{BL}}=-\frac{1}{c}\ln[E(\mathrm{e}^{-c\theta}|X)]$。

定理1　设 $X=(X_{L(1)},\cdots,X_{L(n)})$ 为来自广义指数分布的一个记录值样本,θ 的先验分布为伽玛分布 $\Gamma(\alpha,\beta)$,相应的概率密度函数为

$$\pi(\theta;\alpha,\beta)=\frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}\mathrm{e}^{-\beta\theta},\theta>0,\alpha,\beta>0 \tag{6}$$

则有:1) 在平方误差损失函数下,参数 θ 的 Bayes 估计为 $\hat{\theta}_{\mathrm{BS}}=\frac{n+\alpha}{\beta+T}$;2) 在 LINEX 损失函数下,参数 θ 的 Bayes 估计为 $\hat{\theta}_{\mathrm{BL}}=\frac{n+\alpha}{c}\ln\frac{T+\beta+c}{T+\beta}$。

证明　由(2)、(6)式及 Bayes 定理,参数 θ 的后验概率密度函数为

$$h(\theta|x)\propto l(\theta)\cdot\pi(\theta;\alpha,\beta)\propto\theta^n\mathrm{e}^{-\theta t}\theta^{\alpha-1}\mathrm{e}^{-\beta\theta}\propto\theta^{n+\alpha-1}\mathrm{e}^{-(\beta+t)\theta} \tag{7}$$

从而 θ 的后验分布为 $\Gamma(n+\alpha,\beta+t)$。

则有:1) 在平方误差损失函数下,参数 θ 的 Bayes 估计为其后验均值,从而 θ 的 Bayes 估计为 $\hat{\theta}_{\mathrm{BS}}=E(\theta|X)=\frac{n+\alpha}{\beta+T}$;2) 由(11)式有 $E(\mathrm{e}^{-c\theta}\mid X)=\int_0^{+\infty}\mathrm{e}^{-c\theta}\frac{(\beta+T)^{n+\alpha}}{\Gamma(n+\alpha)}\theta^{n+\alpha-1}\mathrm{e}^{-(\beta+T)\theta}\mathrm{d}\theta=\frac{(\beta+T)^{n+\alpha}}{\Gamma(n+\alpha)}\cdot\frac{\Gamma(n+\alpha)}{(\beta+T+c)^{n+\alpha}}=\left(\frac{T+\beta}{T+\beta+c}\right)^{n+\alpha}$。　　证毕

于是在 LINEX 损失函数下,参数 θ 的 Bayes 估计为 $\hat{\theta}_{BL}=-\frac{1}{c}\ln E(\mathrm{e}^{-c\theta}|X)=\frac{n+\alpha}{c}\ln\frac{T+\beta+c}{T+\beta},c\neq 0$ 。

当超参数 α 和 β 未知时,Bayes 估计就无法使用,可以利用过去样本和经验 Bayes 估计方法得到 α 和 β 的估计值。然后将得到的估计值代入到 Bayes 估计中就可以得到参数 θ 的经验 Bayes 估计。更多的关于经验 Bayes 方法的讨论可以参考文献[10-14]。

对于随机变量 X,假定 $X_{N+1,L(1)},X_{N+1,L(2)},\cdots,X_{N+1,L(n)}$ 为当前样本,要估计的参数 θ 的真值为 θ_{N+1}。$X_{l,L(1)},X_{l,L(2)},\cdots,X_{l,L(n)},l=1,2,\cdots,N$ 为过去 N 组样本。每个样本都服从广义指数分布。基于过去第 $l,l=1,2,\cdots,N$,组样本,利用(3)式得到对应的参数 θ 的 MLE 估计 θ_l 为

$$\hat{\theta}_l\equiv Y_l=\frac{n}{T_l} \tag{8}$$

其中 $T_l=-\ln(1-\mathrm{e}^{-X_{l,L(n)}})$。

对于给定的 θ_l，T_l 的条件概率密度为伽玛概率密度函数，于是 Y_l 服从如下概率密度函数

$$h(y_l\mid\theta_l)=\frac{(n\theta_l)^n}{\Gamma(n)}\frac{1}{y_l^{n+1}}\mathrm{e}^{-n\theta_l/y_l} \tag{9}$$

从而 $Y_l,l=1,2,\cdots,N$ 的边缘概率密度函数为 $h(y_l)=\int_0^\infty f(y_l\mid\theta_l)g(\theta_l;\alpha,\beta)\mathrm{d}\theta_l$，其中 $g(\theta_l;\alpha,\beta)$ 为 θ_l 的伽玛先验概率密度函数。

由(7)、(8)式得 $Y_l,l=1,2,\cdots,N$ 的概率密度函数为 $h(y_l)=\frac{\beta n^n}{B(\alpha,n)}\frac{(\beta y_l)^{\alpha-1}}{[n+\beta y_l]^{n+\alpha}},y_l>0$ 。利用矩估计法和(9)式，α 和 β 的矩估计量为

$$\hat{\alpha}=\frac{s_1^2}{s_2-s_1^2},\hat{\beta}=\frac{s_1}{s_2-s_1^2} \tag{10}$$

其中

$$s_1=\frac{n-1}{Nn}\sum_{l=1}^N\hat{\theta}_l,s_2=\frac{(n-1)(n-2)}{Nn^2}\sum_{l=1}^N\hat{\theta}_l^2 \tag{11}$$

因而，在平方误差损失和 LINEX 损失函数下参数 θ 的经验 Bayes 估计分别为 $\hat{\theta}_{\mathrm{EBS}}=\frac{n+\hat{\alpha}}{\hat{\beta}+T_{N+1,L(n)}}$ 和 $\hat{\theta}_{\mathrm{EBL}}=\frac{n+\hat{\alpha}}{c}\ln\frac{T_{N+1,L(n)}+\hat{\beta}+c}{T_{N+1,L(n)}+\hat{\beta}}$。其中 $T_{N+1,L(n)}=-\ln(1-\exp(-X_{N+1,L(n)}))$，$\hat{\alpha}$，$\hat{\beta}$为 α，β 的估计值并由(10)式给出。

3 数值模拟

利用 Monte Carlo 数值模拟来对最大似然估计、Bayes 估计和经验 Bayes 估计进行比较。步骤如下：1）给定 $\alpha=3.0$ 和 $\beta=1.0$，利用先验概率密度函数(6)式产生 $\theta=2.3043$；2）基于生成的参数 θ 值，利用 Monte Carlo 数值模拟一组来自广义指数分布(1)式的容量为 $n=8$ 的记录值样本，作为当前样本：0.176 9、0.380 7、0.541 3、0.911 4、1.218 1、2.869 9、3.807 9、5.047 0；3）对于给定的 $\alpha=3.0$，$\beta=1.0$，参数 θ 的过去值 $\theta_l,l=1,2,\cdots,N$ 的估计值 $\hat{\theta}_l\equiv Y_l$ 由(9)式的概率密度生成的样本观察值给出，这组样本($N=20$)为：0.263 2、0.318 1、0.177 7、0.317 4、0.274 3、0.264 6、0.252 7、0.236 1、0.493 6、0.205 7、0.187 9、0.308 4、0.586 3、0.541 7、0.800 0、0.175 0、0.692 4、1.799 5、0.230 4、0.634 7；4）先验参数 α 和 β 的估计值由步骤 3)和(10)、(11)式得出：$\hat{\alpha}=4.0083$，$\hat{\beta}=1.3176$；5）θ 的最大似然估计、Bayes 估计以及经验 Bayes 估计值由表 1 给出。

表 1　参数 θ 的各类估计值

最大似然估计	Bayes 估计			经验 Bayes 估计		
$\hat{\theta}_{\mathrm{ML}}$	$\hat{\theta}_{\mathrm{BS}}$	$\hat{\theta}_{\mathrm{BL}}(c=1.0)$	$\hat{\theta}_{\mathrm{BL}}(c=2.0)$	$\hat{\theta}_{\mathrm{EBS}}$	$\hat{\theta}_{\mathrm{EBL}}(c=1.0)$	$\hat{\theta}_{\mathrm{EBL}}(c=1.0)$
2.348 7	2.496 5	2.249 9	2.058 4	2.542 1	2.305 8	2.119 7

通过大量的数值模拟实验，得到如下结论：当采取合适的先验分布，也即当对参数的了解越多时，Bayes 和经验 Bayes 估计值越接近真实值，此时优于传统的最大似然估计。

致谢：诚挚感谢审稿专家给出的宝贵建议。

参考文献：

[1] Gupta R D，Kundu D. Generalized exponential distributions[J]. Australian and New Zealand Journal of Statistics，1999，41(2)：173-188.

[2] Gupta R D，Kundu D. Generalized exponential distributions：different method of estimations[J]. Journal of Statistical Computation and Simulation，2001，69(4)：315-338.

[3] 唐玉娜，施瑞，王炳兴. 广义指数分布的统计推断[J]. 统计与决策，2008(17)：18-19.
Tang Y N，Shi R，Wang B X. Statistical inference of generalized exponential distribution[J]. Statistics and Decision，2008(17)：18-19.

[4] 王国富. 熵损失函数下两参数广义指数分布形状参数的

Bayes 估计[J]. 统计与决策,2010(1):154-155.
Wang G F. Bayes estimation of shape parameter of two parameter generalized exponential distribution under entropy loss function[J]. Statistics and Decision,2010(1):154-155.
[5] Dey S. Bayesian Estimation of the shape parameter of the generalized exponential distribution under different loss functions[J]. Pak J Stat Oper Res,2010,4(2):163-174.
[6] Ahmadi J,Doostparast M. Bayesian estimation and prediction for some life distributions based on record values[J]. Stat Pap,2006,47(3):373-392.
[7] Asgharzadeh A. On Bayesian estimation from exponential distribution based on records[J]. The Korean Statistical Society,2009,38(2):125-130.
[8] Arnold B C,Balakrishnan N,Nagaraja H N. Records[M]. New York:John Wiley & Sons,1998.
[9] Zellner A. Bayesian estimation and prediction using asymmetric loss functions[J]. J Amer Statist Assoc,1986,81(394):446-451.
[10] Jaheen Z F. Empirical Bayes analysis of record statistics based on LINEX and quadratic loss functions[J]. Computers & Mathematics with Applications,2004,47(6/7):947-954.
[11] Ren H P. Empirical Bayes estimation in exponential model based on record values under asymmetric loss[J]. Advances in Intelligent and Soft Computing,2012,123:659-667.
[12] 陈玲,韦来生. 连续型单参数指数族参数的经验 Bayes 检验[J]. 系统科学与数学,2009,29(8):1142-1152.
Chen L,Wei L S. Convergence rates of the empirical Bayes test problem for continuous one-parameter exponential family[J]. Journal of Systems Science and Mathematical Sciences,2009,29(8):1142-1152.
[13] 王琪,李玮. 一类新的损失函数下负二项分布模型的 Bayes 可靠性分析[J]. 江西师范大学学报:自然科学版,2011,35(4):384-387.
Wang Q,Li W. The Bayesian analysis for negative binomial distribution under a new loss function[J]. Journal of Jiangxi Normal University:Natural Science,2011,35(4):384-387.
[14] 徐宝,姜玉秋,滕飞. 一种加权对称损失函数下一类指数分布模型参数的估计[J]. 四川师范大学学报:自然科学版,2011,34(4):484-487.
Xu B,Jiang Y Q,Teng F. Estimator of the scale parameter in a subclass of the exponential model under a weighted symmetric loss fanction[J]. Journal of Sichuan Normal University:Natural Science,2011,34(4):484-487.

Bayes and Empirical Bayes Estimation of Parameter of Generalized Exponential Distribution Based on Record Values

WANG Qi[1], HUANG Wen-yi[2]
(1. Zhongshan Institute, University of Electronic Science and Technology of China, Zhongshan Guangdong 528402;
2. School of Mathematics and Computer Science, Yichun University, Yichun Jiangxi 336000, China)

Abstract: According to the record values from the generalized exponential distribution sample, parameter estimation problem of generalized exponential distribution is studied. First, the maximum likelihood parameter estimation is given; when the conjugate prior distribution of parameters for the Berta prior distribution, and the loss function is the squared error loss and LINEX loss functions, Bayes and empirical Bayes estimators of the parameter of the generalized exponential distribution are derived. Finally, a Monte Carlo numerical simulation is given to compare these estimators presented in this paper. We found in the prior distribution of the right conditions and experience of the Bayes estimates closer to the true parameters, therefore the maximum likelihood proper prior distribution of Bayes and empirical Bayes estimators are closer than the maximum likelihood estimator when the prior distribution is proper. Then Bayes and empirical Bayes estimators are better than the maximum likelihood estimator under proper prior distribution.

Key words: Bayes and empirical Bayes estimators; squared error loss; LINEX loss; record value

(责任编辑 黄 颖)