第43卷 第2期 《新疆师范大学学报》(自然科学版) Vol.43,No.2
2024年6月 Journal of Xinjiang Normal University Jun. 2024
(Natural Sciences Edition)

逐步增加Ⅱ型截尾下Pareto分布形状参数的Bayes估计

赵孟茹,周菊玲*
(新疆师范大学 数学科学学院,新疆 乌鲁木齐 830017)

摘　要:基于逐步增加Ⅱ型截尾样本,首先得出Pareto分布形状参数的极大似然估计,考虑两个损失函数和形状参数的两个先验分布,得出该分布形状参数的4个Bayes估计。由数值模拟结果发现,上述四个Bayes估计值的均方误差均小于极大似然估计值,其中,当损失函数为二次损失函数,形状参数的先验分布为共轭先验分布时的Bayes估计的均方误差较小,估计效果更理想,且实例分析与数值模拟结果相符。其次在二次损失函数下,针对形状参数先验分布选取共轭先验分布,给出Pareto分布形状参数的多层Bayes估计和E-Bayes估计。

关键词:逐步增加Ⅱ型截尾;Pareto分布;二次损失;Q对称熵损失;Bayes估计

中图分类号:O212.1　　**文献标识码**:A　　**文章编号**:1008-9659(2024)02-0001-09

Pareto分布是一种幂律概率分布,通常被应用于描述地球物理和保险精算等可观察的现象以及反映产品质量情况等问题,由此学者也对该分布的相关知识产生了研究兴趣。Frederico等人在现有估计方法的基础上为Pareto分布的参数开发出一类新的广义概率加权矩估计,讨论了新估计量的渐近正态性,并通过模拟数据集及真实数据集验证其更适合小样本情况[1];Abravesh等人采用林德利近似法和Monte Carlo积分法等四种Bayes估计方法,研究了Ⅱ型截尾样本下Pareto分布的应力-强度可靠性[2]。Han研究了不同损失函数下Pareto分布参数的E-Bayes估计及其预期均方误差(E-MSE),并利用MCMC方法分别给出形状参数和尺度参数的Bayes置信区间[3]。龙兵等人在双定数混合截尾与双边定时截尾下,得到Pareto分布尺度参数的极大似然估计与形状参数的极大似然估计、置信区间和Bayes估计以及采用EM算法解决形状参数的估计问题[4-5];基于Pareto分布的定数截尾样本,周巧娟研究了对称熵损失函数下形状参数的条件Γ - minimax估计问题,并结合随机模拟验证其具有良好的稳健性[6];刘芹等人讨论了Pareto分布形状参数在三种损失函数下的Bayes估计,并由实例说明在Mlinex损失函数下的估计效果更佳[7];徐圣楠等人讨论了缺失数据下混合Pareto分布形状参数估计量的相合性与渐近正态性[8];在加权对称熵损失函数下,吴月丹等人分别得到了广义Pareto分布在不同先验分布下的Bayes估计与Mini max估计[9]。与上述几类缺失数据样本不同,逐步增加Ⅱ型截尾试验过程中可以随机移除一定数量的试验样本,此类缺失数据下也有学者进行不同方向的深入探究。在逐步增加Ⅱ型截尾下,李凤等人分别讨论了Pareto分布在两种不同损失函数下形状参数和可靠性指标的Bayes估计[10];李琼等人采用Markov chain Monte Carlo方法分别探究了Pareto分布形状参数和尺度参数的Bayes估计[11];邵媛媛等人分别讨论了复合瑞利分布、Lomax分布与广义指数分布各参数和可靠度函数的估计问题[12]。文章在此前龙兵等人和杨君慧等人对Pareto分布研究[13-14]的基础上,讨论在逐步增加Ⅱ型截尾试验下,形状参数的极大似然估计和几种Bayes估计问题。

[收稿日期]2023-07-17　　[修回日期]2023-09-04
[基金项目]国家自然科学基金项目(11801488);新疆师范大学校级科研平台招标课题(XJNUSYS2019B05)。
[作者简介]赵孟茹(1998-),女,硕士研究生,主要从事数理统计方面研究,E-mail:1550079161@qq.com.
*[通讯作者]周菊玲(1968-),女,教授,主要从事数理统计方面研究,E-mail:326815649@qq.com.

1 预备知识

定义1[4] 设X为随机变量,若X服从Pareto分布,其分布函数为

$$F(x;\theta,a)=1-\left(ax^{-1}\right)^{\theta},x>0 \tag{1}$$

其密度函数为

$$f(x;a,\theta)=\theta a^{\theta}x^{-(\theta+1)},x>0 \tag{2}$$

其中,a为尺度参数,θ为形状参数,且$a>0,\theta>0$.

逐步增加Ⅱ型截尾模型简述为:假设现有n个服从Pareto分布的样品同一时间进行试验,设$m<n$,$R_j(j=1,2,\cdots,m)$为给定的正整数。当出现第一个失效样品时,将失效时间记为X_1,并在其余的$n-1$个试验样品中随机移除R_1个样品不再参与试验。当出现第二个失效样品时,将失效时间记为X_2,并在其余的$n-1-R_1-1$个试验样品中随机移除R_2个样品不再参与试验。依此继续进行试验,直到出现第m个失效样品时,将失效时间记为X_m,并将其余的$R_m=n-m-\sum_{j=1}^{m-1}R_j$个样品全部移除。特别地,当$R_j=0$时即为全样本的情况。

引理1[10] Pareto分布形状参数的共轭先验分布为伽玛分布。

证明 设$X=\left(X_1,X_2,\cdots\cdots,X_n\right)$是服从Pareto分布的一组样本,其尺度参数已知,取形状参数的先验分布为$\Gamma\left(\alpha,\beta\right)$,有

$$\pi(\theta)=\frac{\beta^{\alpha}}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}$$

此样本的联合密度函数为

$$f(x|\theta)=\prod_{j=1}^{n}f\left(x_j\right)=\theta^{n}a^{\theta n}\prod_{j=1}^{n}x_j^{-(\theta+1)}$$

θ的后验密度函数为

$$\pi(\theta|x)=\frac{\pi(\theta)f(x|\theta)}{\int_0^{+\infty}\pi(\theta)f(x|\theta)d\theta}\propto\frac{\left(\beta-n\ln a+\sum_{j=1}^{n}\ln x_j\right)^{n+\alpha}}{\Gamma(n+\alpha)}\theta^{n+\alpha-1}e^{-\theta\left(\beta-n\ln a+\sum_{j=1}^{n}\ln x_j\right)}$$

由伽玛分布的定义知

$$\theta\Big|x\sim\Gamma\left(n+\alpha,\beta-n\ln a+\sum_{j=1}^{n}\ln x_j\right)$$

即伽玛分布$\Gamma\left(\alpha,\beta\right)$为形状参数的共轭先验分布。

文章考虑以下两种先验分布:

(1)无信息先验分布:$\pi_1(\theta)=\dfrac{1}{\theta}$;

(2)共轭先验分布:$\pi_2(\theta)=\dfrac{\beta^{\alpha}}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}$.

2 形状参数的极大似然估计和Bayes估计

2.1 形状参数的极大似然估计

逐步增加Ⅱ型截尾样本下的Pareto分布的似然函数为

$$L(a,\theta|x)=M\prod_{j=1}^{m}f\left(x_j\right)\left[1-F\left(x_j\right)\right]^{R_j}=M\prod_{j=1}^{m}\left[\theta a^{\theta}x_j^{-(\theta+1)}\right]\left\{1-\left[1-\left(ax_j^{-1}\right)^{\theta}\right]\right\}^{R_j}$$

$$=M\prod_{j=1}^{m}\frac{\theta}{a}\left(\frac{x_j}{a}\right)^{-\left[\theta\left(1+R_j\right)+1\right]}=M\left(\frac{\theta}{a}\right)^{m}\prod_{j=1}^{m}\left(\frac{x_j}{a}\right)^{-\left[\theta\left(1+R_j\right)+1\right]}$$

其中，$M=n\left(n-1-R_1\right)\left(n-2-R_1-R_2\right)\cdots\cdots\left(n-\sum_{j=1}^{m-1}\left(R_j+1\right)\right)$，且与任何参数无关。其对数似然函数为

$$\ln L\left(a,\theta|x\right)=\ln M+m\ln\theta-m\ln a-\sum_{j=1}^{m}\left[\theta\left(1+R_j\right)+1\right]\ln\left(\frac{x_j}{a}\right)$$

则得到形状参数的极大似然估计

$$\hat{\theta}_M=\frac{m}{T}$$

其中
$$T=\sum_{j=1}^{m}\left(1+R_j\right)\ln\left(\frac{x_j}{a}\right)$$

2.2 二次损失函数下形状参数的Bayes估计

定义2[15] 二次损失函数

$$S_1\left(\theta,\delta\right)=\frac{\left(\delta-\theta\right)^2}{\theta^2} \tag{3}$$

引理2[15] 在二次损失函数下，对Pareto分布形状参数取任意先验分布，θ的Bayes估计为

$$\hat{\delta}_1=\frac{E\left(\theta^{-1}|X\right)}{E\left(\theta^{-2}|X\right)} \tag{4}$$

其中，$\hat{\delta}_1$为θ的估计值，且解唯一。

定理1 基于逐步增加Ⅱ型截尾试验，在损失函数（式(3)）下，设$X=\left(X_1,X_2,\cdots\cdots,X_n\right)$为一组服从Pareto分布的独立同分布样本，当形状参数的先验分布为无信息先验分布$\pi_1(\theta)$时，其Bayes估计为

$$\hat{\delta}_{11}=\frac{m-2}{T}$$

证明 由逐步增加Ⅱ型截尾样本下的Pareto分布的联合概率密度可得

$$L\left(a,\theta|x\right)=M\left(\frac{\theta}{a}\right)^m\prod_{j=1}^{m}\left(\frac{x_j}{a}\right)^{-\left[\theta\left(1+R_j\right)+1\right]}\propto\theta^m e^{-\theta T}$$

当形状参数的先验分布为$\pi_1(\theta)$时，其后验分布为

$$h_1\left(a,\theta|x\right)=\frac{\pi_1(\theta)L\left(a,\theta|x\right)}{\int_0^{+\infty}\pi_1(\theta)L\left(a,\theta|x\right)d\theta}=\frac{T^m}{\Gamma(m)}\theta^{m-1}e^{-\theta T}$$

$$E\left(\theta^{-1}|X\right)=\int_0^{+\infty}\theta^{-1}h_1\left(a,\theta|x\right)d\theta=\int_0^{+\infty}\theta^{-1}\frac{T^m}{\Gamma(m)}\theta^{m-1}e^{-\theta T}d\theta=\frac{T^m}{\Gamma(m)}\int_0^{+\infty}\theta^{m-2}e^{-\theta T}d\theta=\frac{T}{m-1}$$

$$E\left(\theta^{-2}|X\right)=\int_0^{+\infty}\theta^{-2}h_1\left(a,\theta|x\right)d\theta=\int_0^{+\infty}\theta^{-2}\frac{T^m}{\Gamma(m)}\theta^{m-1}e^{-\theta T}d\theta=\frac{T^m}{\Gamma(m)}\int_0^{+\infty}\theta^{m-3}e^{-\theta T}d\theta=\frac{T^2}{(m-1)(m-2)}$$

由式(4)可得，此条件下形状参数的Bayes估计为

$$\hat{\delta}_{11}=\frac{E\left(\theta^{-1}|X\right)}{E\left(\theta^{-2}|X\right)}=\frac{m-2}{T}$$

定理2 基于逐步增加Ⅱ型截尾试验，在损失函数（式(3)）下，设$X=\left(X_1,X_2,\cdots\cdots,X_n\right)$为一组服从Pareto分布的独立同分布样本，当形状参数的先验分布为共轭先验分布$\pi_2(\theta)$时，其Bayes估计为

$$\hat{\delta}_{12}=\frac{\alpha+m-2}{\beta+T}$$

证明 当形状参数的先验分布为$\pi_2(\theta)$时，其后验分布为

$$h_2\left(a,\theta|x\right)=\frac{\pi_2(\theta)L\left(a,\theta|x\right)}{\int_0^{+\infty}\pi_2(\theta)L\left(a,\theta|x\right)d\theta}=\frac{\left(\beta+T\right)^{\alpha+m}}{\Gamma\left(\alpha+m\right)}\theta^{\alpha+m-1}e^{-\theta(\beta+T)}$$

$$E\left(\theta^{-1}|X\right)=\int_0^{+\infty}\theta^{-1}h_2\left(a,\theta|x\right)d\theta=\int_0^{+\infty}\theta^{-1}\frac{\left(\beta+T\right)^{\alpha+m}}{\Gamma\left(\alpha+m\right)}\theta^{\alpha+m-1}e^{-\theta(\beta+T)}d\theta$$

$$=\frac{\left(\beta+T\right)^{\alpha+m}}{\Gamma\left(\alpha+m\right)}\int_0^{+\infty}\theta^{\alpha+m-2}e^{-\theta(\beta+T)}d\theta=\frac{\beta+T}{\alpha+m-1}$$

$$E\left(\theta^{-2}|X\right)=\int_0^{+\infty}\theta^{-2}h_2\left(a,\theta|x\right)d\theta=\int_0^{+\infty}\theta^{-2}\frac{\left(\beta+T\right)^{\alpha+m}}{\Gamma\left(\alpha+m\right)}\theta^{\alpha+m-1}e^{-\theta(\beta+T)}d\theta$$

$$=\frac{\left(\beta+T\right)^{\alpha+m}}{\Gamma\left(\alpha+m\right)}\int_0^{+\infty}\theta^{\alpha+m-3}e^{-\theta(\beta+T)}d\theta=\frac{\left(\beta+T\right)^2}{(\alpha+m-1)(\alpha+m-2)}$$

由式(4)可得,此条件下形状参数的Bayes估计为

$$\hat{\delta}_{12}=\frac{E\left(\theta^{-1}|X\right)}{E\left(\theta^{-2}|X\right)}=\frac{\alpha+m-2}{\beta+T}$$

2.3 Q对称熵损失函数下形状参数的Bayes估计

定义3[16] Q对称熵损失函数

$$S_2(\theta,\delta)=\left(\frac{\theta}{\delta}\right)^q+\left(\frac{\delta}{\theta}\right)^q-2 \tag{5}$$

引理3[16] 在Q对称熵损失函数下,对Pareto分布形状参数取任意先验分布,形状参数的Bayes估计为

$$\hat{\delta}_2=\left(\frac{E\left(\theta^q|X\right)}{E\left(\theta^{-q}|X\right)}\right)^{\frac{1}{2q}} \tag{6}$$

其中,$\hat{\delta}_2$为θ的估计值,且解唯一,此处考虑$q>0$时的情况。

定理3 基于逐步增加Ⅱ型截尾试验,在损失函数(式(5))下,设$X=\left(X_1,X_2,\cdots\cdots,X_n\right)$为一组服从Pareto分布的独立同分布样本,当形状参数的先验分布为无信息先验分布$\pi_1(\theta)$时,其Bayes估计为

$$\hat{\delta}_{21}=\left[\frac{\Gamma\left(m+q\right)}{\Gamma\left(m-q\right)}\right]^{\frac{1}{2q}}\frac{1}{T}$$

证明 由定理1知,$h_1\left(a,\theta|x\right)=\dfrac{T^m}{\Gamma(m)}\theta^{m-1}e^{-\theta T}$.

$$E\left(\theta^q|X\right)=\int_0^{+\infty}\theta^q h_1\left(a,\theta|x\right)d\theta=\int_0^{+\infty}\theta^q\frac{T^m}{\Gamma(m)}\theta^{m-1}e^{-\theta T}d\theta$$

$$=\frac{T^m}{\Gamma(m)}\int_0^{+\infty}\theta^{m+q-1}e^{-\theta T}d\theta=\frac{\Gamma\left(m+q\right)}{\Gamma(m)T^q}$$

$$E\left(\theta^{-q}|X\right)=\int_0^{+\infty}\theta^{-q}h_1\left(a,\theta|x\right)d\theta=\int_0^{+\infty}\theta^{-q}\frac{T^m}{\Gamma(m)}\theta^{m-1}e^{-\theta T}d\theta$$

$$=\frac{T^m}{\Gamma(m)}\int_0^{+\infty}\theta^{m-q-1}e^{-\theta T}d\theta=\frac{\Gamma\left(m-q\right)}{\Gamma(m)T^{-q}}$$

由式(6)可得,此条件下形状参数的Bayes估计为

$$\hat{\delta}_{21}=\left[\frac{E\left(\theta^q|X\right)}{E\left(\theta^{-q}|X\right)}\right]^{\frac{1}{2q}}=\left[\frac{\Gamma\left(m+q\right)}{\Gamma\left(m-q\right)}\right]^{\frac{1}{2q}}\frac{1}{T}$$

定理4 基于逐步增加Ⅱ型截尾试验,在损失函数(式(5))下,设$X=\left(X_1,X_2,\cdots,X_n\right)$为一组服从Pareto分布的独立同分布样本,当形状参数的先验分布为共轭先验分布$\pi_2(\theta)$时,其Bayes估计为

$$\hat{\delta}_{22}=\left[\frac{\Gamma\left(\alpha+m+q\right)}{\Gamma\left(\alpha+m-q\right)}\right]^{\frac{1}{2q}}\frac{1}{\beta+T}$$

证明 由定理2可知，$h_2(a,\theta|x)=\dfrac{(\beta+T)^{\alpha+m}}{\Gamma(\alpha+m)}\theta^{\alpha+m-1}e^{-\theta(\beta+T)}$.

$$E(\theta^q|X)=\int_0^{+\infty}\theta^q h_2(a,\theta|x)d\theta=\int_0^{+\infty}\theta^q\frac{(\beta+T)^{\alpha+m}}{\Gamma(\alpha+m)}\theta^{\alpha+m-1}e^{-\theta(\beta+T)}d\theta$$

$$=\frac{(\beta+T)^{\alpha+m}}{\Gamma(\alpha+m)}\int_0^{+\infty}\theta^{\alpha+m+q-1}e^{-\theta(\beta+T)}d\theta=\frac{\Gamma(\alpha+m+q)}{\Gamma(\alpha+m)(\beta+T)^q}$$

$$E(\theta^{-q}|X)=\int_0^{+\infty}\theta^{-q}h_2(a,\theta|x)d\theta=\int_0^{+\infty}\theta^{-q}\frac{(\beta+T)^{\alpha+m}}{\Gamma(\alpha+m)}\theta^{\alpha+m-1}e^{-\theta(\beta+T)}d\theta$$

$$=\frac{(\beta+T)^{\alpha+m}}{\Gamma(\alpha+m)}\int_0^{+\infty}\theta^{\alpha+m-q-1}e^{-\theta(\beta+T)}d\theta=\frac{\Gamma(\alpha+m-q)}{\Gamma(\alpha+m)(\beta+T)^{-q}}$$

由式(6)可得，此条件下形状参数的Bayes估计为

$$\hat{\delta}_{22}=\left[\frac{E(\theta^q|X)}{E(\theta^{-q}|X)}\right]^{\frac{1}{2q}}=\left[\frac{\Gamma(\alpha+m+q)}{\Gamma(\alpha+m-q)}\right]^{\frac{1}{2q}}\frac{1}{\beta+T}$$

3 形状参数的多层Bayes估计和E-Bayes估计

基于逐步增加Ⅱ型截尾试验，在二次损失函数和共轭先验分布的条件下，讨论Pareto分布形状参数的多层Bayes估计和E-Bayes估计问题，并对其进行数值模拟。

3.1 形状参数的多层Bayes估计

从定理2和定理4所得结果可以看出，共轭先验分布$\Gamma(\alpha,\beta)$中的超参数α,β较难确定。假定两个超参数是相互独立的随机变量，分别取均匀分布$U(0,1),U(0,c)$为α,β的超先验分布时，其概率密度为

$$\pi(\alpha)=1,\pi(\beta)=\frac{1}{c} \tag{7}$$

其中，$0<\beta<c(c>0)$，考虑到稳健性，c应取较小的值。

定理5 基于逐步增加Ⅱ型截尾试验，在损失函数（式(3)）下，设$X=(X_1,X_2,\cdots\cdots,X_n)$为一组服从Pareto分布的独立同分布样本，当形状参数的先验分布为共轭先验分布$\pi_2(\theta)$时，超参数α,β的先验分布由式(7)给出，则此分布形状参数的多层Bayes估计为

$$\hat{\delta}_{HB}=\frac{\displaystyle\int_0^1\int_0^c\frac{\beta^\alpha\Gamma(\alpha+m-1)}{\Gamma(\alpha)(\beta+T)^{\alpha+m-1}}d\alpha d\beta}{\displaystyle\int_0^1\int_0^c\frac{\beta^\alpha\Gamma(\alpha+m-2)}{\Gamma(\alpha)(\beta+T)^{\alpha+m-2}}d\alpha d\beta}$$

证明 假定超参数α,β独立，由此得形状参数的多层先验密度为

$$g(\theta)=\int_0^1\int_0^c\pi_2(\theta)\pi(\alpha)\pi(\beta)d\alpha d\beta=\frac{1}{c}\int_0^1\int_0^c\frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}d\alpha d\beta$$

则形状参数的后验分布为

$$h_3(a,\theta|x)=\frac{g(\theta)L(a,\theta|x)}{\int_{-\infty}^{+\infty}g(\theta)L(a,\theta|x)d\theta}=\frac{\displaystyle\int_0^1\int_0^c\frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha+m-1}e^{-(\beta+T)\theta}d\alpha d\beta}{\displaystyle\int_0^1\int_0^c\frac{\beta^\alpha\Gamma(\alpha+m)}{\Gamma(\alpha)(\beta+T)^{\alpha+m}}d\alpha d\beta}$$

$$E(\theta^{-1}|X)=\int_{-\infty}^{+\infty}\theta^{-1}h_3(a,\theta|x)d\theta=\frac{\displaystyle\int_0^1\int_0^c\int_0^{+\infty}\frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha+m-1-1}e^{-(\beta+T)\theta}d\theta d\alpha d\beta}{\displaystyle\int_0^1\int_0^c\frac{\beta^\alpha\Gamma(\alpha+m)}{\Gamma(\alpha)(\beta+T)^{\alpha+m}}d\alpha d\beta}=\frac{\displaystyle\int_0^1\int_0^c\frac{\beta^\alpha\Gamma(\alpha+m-1)}{\Gamma(\alpha)(\beta+T)^{\alpha+m-1}}d\alpha d\beta}{\displaystyle\int_0^1\int_0^c\frac{\beta^\alpha\Gamma(\alpha+m)}{\Gamma(\alpha)(\beta+T)^{\alpha+m}}d\alpha d\beta}$$

$$E\left(\theta^{-2}\middle|X\right)=\int_{-\infty}^{+\infty}\theta^{-2}h_3\left(a,\theta\middle|x\right)d\theta=\frac{\int_0^1\int_0^c\int_0^{+\infty}\frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha+m-2-1}e^{-(\beta+T)\theta}d\theta d\alpha d\beta}{\int_0^1\int_0^c\frac{\beta^\alpha\Gamma(\alpha+m)}{\Gamma(\alpha)(\beta+T)^{\alpha+m}}d\alpha d\beta}$$

$$=\frac{\int_0^1\int_0^c\frac{\beta^\alpha\Gamma(\alpha+m-2)}{\Gamma(\alpha)(\beta+T)^{\alpha+m-2}}d\alpha d\beta}{\int_0^1\int_0^c\frac{\beta^\alpha\Gamma(\alpha+m)}{\Gamma(\alpha)(\beta+T)^{\alpha+m}}d\alpha d\beta}$$

由式(4)得形状参数的多层Bayes估计为

$$\hat{\delta}_{HB}=\frac{E\left(\theta^{-1}\middle|X\right)}{E\left(\theta^{-2}\middle|X\right)}=\frac{\int_0^1\int_0^c\frac{\beta^\alpha\Gamma(\alpha+m-1)}{\Gamma(\alpha)(\beta+T)^{\alpha+m-1}}d\alpha d\beta}{\int_0^1\int_0^c\frac{\beta^\alpha\Gamma(\alpha+m-2)}{\Gamma(\alpha)(\beta+T)^{\alpha+m-2}}d\alpha d\beta}$$

3.2 形状参数的E-Bayes估计

相较于形状参数的多层Bayes估计,求解E-Bayes估计过程中的积分计算较少,更为简便,在应用方面更具优势。

定义4[17] 如果 $\hat{\theta}_B(a,b)$ 连续,则称 $\hat{\theta}_{EB}=\iint\limits_D\hat{\theta}_B(a,b)\cdot p(a,b)dadb$ 为形状参数的E-Bayes估计,其中,$\iint\limits_D\hat{\theta}_B(a,b)\cdot p(a,b)dadb<\infty$,$D$ 为 a 与 b 取值范围构成的集合,$p(a,b)$ 为 a 与 b 的概率密度函数,$\hat{\theta}_B(a,b)$ 为形状参数的Bayes估计。

定理6 基于逐步增加Ⅱ型截尾试验,在损失函数(式(3))下,设 $X=\left(X_1,X_2,\cdots\cdots,X_n\right)$ 为一组服从Pareto分布的独立同分布样本,当形状参数的先验分布为共轭先验分布 $\pi_2(\theta)$ 时,超参数 α,β 的先验分布由式(7)给出,则此分布形状参数的E-Bayes估计为

$$\hat{\delta}_{EB}=\frac{(-3+2m)}{2c}\ln\left(1+\frac{c}{T}\right)$$

证明 假定超参数 α,β 独立,再由式(7)知

$$\pi(\alpha,\beta)=\pi(\alpha)\pi(\beta)=\frac{1}{c}$$

由定理2和定义4得E-Bayes估计为

$$\hat{\delta}_{EB}=\iint\limits_D\hat{\delta}_{12}\cdot\pi(\alpha,\beta)d\alpha d\beta=\int_0^1\int_0^c\frac{\alpha+m-2}{\beta+T}\frac{1}{c}d\alpha d\beta$$

$$=\frac{1}{c}\int_0^1(\alpha+m-2)\int_0^c\frac{1}{\beta+T}d\beta d\alpha=\frac{1}{c}\ln\left(1+\frac{c}{T}\right)\int_0^1(\alpha+m-2)d\alpha$$

$$=\frac{1}{c}\ln\left(1+\frac{c}{T}\right)\frac{-3+2m}{2}=\frac{-3+2m}{2c}\ln\left(1+\frac{c}{T}\right)$$

4 随机模拟

借助 R 软件随机生成一组服从Pareto分布的逐步增加Ⅱ型截尾样本,其操作步骤如下:

(1)产生 n 个服从均匀分布 $U(0,1)$ 的独立同分布样本 $U_1,U_2,\cdots,U_n$;

(2)考虑Pareto分布尺度参数 a 已知的情形,假设 $a=1,\theta=1.5$,此时令 $X_i=\left(1-U_i\right)^{-\frac{1}{\theta}},i=1,2,\cdots,n$,由此得到 $X_1,X_2,\cdots,X_n$,即 n 个服从 $a=1,\theta=1.5$ 的Pareto分布样本;

(3)对上述 n 个所得样本按从小到大排序,并给定试验中随机移离的个数 $R_j,1\leqslant j\leqslant m$,从而得到一组逐

步增加Ⅱ型截尾数据。

对于共轭先验分布及损失函数的参数,取$\alpha = 2, \beta = 1, q = 2$,则可根据极大似然估计与定理1~4中所得Bayes估计表达式得出不同条件下Pareto分布形状参数的估计值。考虑到样本的随机性,模拟时将上述过程重复1000次后求各估计值的平均偏差及平均均方误差,所得结果如表1、表2所示。

表1　n, m取不同值时,$\hat{\theta}_M, \hat{\delta}_{11}, \hat{\delta}_{12}$的模拟结果

n	m	$\hat{\theta}_M$		$\hat{\delta}_{11}$		$\hat{\delta}_{12}$	
		平均偏差	均方误差	平均偏差	均方误差	平均偏差	均方误差
10	5	0.361	1.045	−0.331	0.726	−0.187	0.264
	7	0.263	0.667	−0.256	0.358	−0.142	0.211
	9	0.195	0.463	−0.175	0.299	−0.108	0.183
20	10	0.166	0.366	−0.162	0.249	−0.090	0.161
	14	0.153	0.265	−0.102	0.168	−0.056	0.131
	18	0.107	0.178	−0.088	0.126	−0.054	0.102
40	20	0.098	0.144	−0.056	0.115	−0.029	0.100
	30	0.064	0.099	−0.055	0.071	−0.023	0.073
	35	0.055	0.077	−0.043	0.066	−0.018	0.061
60	30	0.051	0.092	−0.040	0.079	−0.019	0.066
	45	0.032	0.054	−0.031	0.048	−0.014	0.051
	55	0.028	0.043	−0.022	0.038	−0.009	0.041

表2　n, m取不同值时,$\hat{\delta}_{21}, \hat{\delta}_{22}$的模拟结果

n	m	$\hat{\delta}_{21}$		$\hat{\delta}_{22}$	
		平均偏差	均方误差	平均偏差	均方误差
10	5	0.113	0.808	0.161	0.362
	7	0.109	0.513	0.132	0.293
	9	0.090	0.375	0.118	0.248
20	10	0.058	0.291	0.104	0.219
	14	0.054	0.190	0.092	0.173
	18	0.051	0.172	0.083	0.143
40	20	0.055	0.126	0.090	0.130
	30	0.040	0.086	0.053	0.081
	35	0.024	0.068	0.034	0.065
60	30	0.022	0.088	0.034	0.077
	45	0.014	0.054	0.032	0.051
	55	0.012	0.046	0.026	0.043

对于共轭先验分布中参数的取值范围,取$\alpha = 2, \beta = 1, c = 2$,则可根据定理5和定理6得出相应条件下Pareto分布形状参数的多层Bayes估计值和E-Bayes估计值。同样,模拟时将上述过程重复1000次后求各估计值的平均偏差及平均均方误差,结果如表3所示。

表3 n,m 取不同值时,$\hat{\delta}_{HB}$,$\hat{\delta}_{EB}$ 的模拟结果

n	m	$\hat{\delta}_{HB}$		$\hat{\delta}_{EB}$	
		平均偏差	均方误差	平均偏差	均方误差
20	10	−0.232	0.229	−0.279	0.200
	14	−0.163	0.167	−0.199	0.145
	18	−0.132	0.120	−0.161	0.110
60	30	−0.074	0.068	−0.115	0.071
	45	−0.049	0.050	−0.075	0.048
	55	−0.036	0.038	−0.050	0.037

由表1和表2所示,与形状参数的极大似然估计值相比,4个Bayes估计值与真值的平均偏差和均方误差更小,说明形状参数的Bayes估计效果更优;在二次损失函数下,对比$\hat{\delta}_{11}$和$\hat{\delta}_{12}$的模拟数据可以看出,当形状参数的先验分布为共轭先验分布时所得估计值的均方误差较小,在Q对称熵损失函数下所得结论亦如此。在无信息先验分布下,对比$\hat{\delta}_{11}$和$\hat{\delta}_{21}$的数据可知,二次损失函数下形状参数的Bayes估计的均方误差较小,在共轭先验分布下所得结论也相同。从表3数据可以看出,随着n,m值增大,形状参数的多层Bayes估计和E-Bayes估计的平均偏差越来越小,两者均方误差值的差距也减小,估计效果与同等条件下的Bayes估计$\hat{\delta}_{12}$越来越接近,且估计效果呈现出优于$\hat{\delta}_{12}$的趋势。由此,基于逐步增加Ⅱ型截尾数据,在二次损失函数及共轭先验分布的条件下,Pareto分布形状参数的Bayes估计效果更佳。

5 实例分析

根据文献[5]中的实例,得到60个服从尺度参数$a = 618$,形状参数$\theta = 2.42$的Pareto分布的样本数据,并按照从小到大的顺序将其进行排列:

626,628,634,643,648,654,659,661,669,671,675,676,683,687,696,702,704,711,715,
727,728,729,735,739,739,740,744,753,761,766,780,791,806,830,830,839,903,928,
940,950,994,998,1048,1054,1105,1210,1247,1267,1274,1436,1507,1568,1628,1665,
1959,1989,2202,2285,2747,3350.

当$m = 30$时,给定$R_7 = R_{14} = R_{20} = R_{25} = 0$,$R_m = 5$,其余$R_j = 1$;当$m = 55$,给定$R_{10} = R_{19} = R_{26} = R_{36} = R_{42} = 1$,其余$R_j = 0$,即可得到两组不同的逐步增加Ⅱ型截尾数据,由文章给出的极大似然估计与四种Bayes估计的表达式,借助R软件对Pareto分布的形状参数进行估计。设参数$q = 2, \alpha = 2, \beta = 1$,模拟结果如表4所示。

表4 形状参数的估计结果

m	估计量	估计值	偏差	均方误差
30	$\hat{\theta}_M$	2.710	0.290	0.084
	$\hat{\delta}_{11}$	2.529	0.109	0.012
	$\hat{\delta}_{12}$	2.486	0.066	0.004
	$\hat{\delta}_{21}$	2.663	0.243	0.059
	$\hat{\delta}_{22}$	2.608	0.188	0.035
55	$\hat{\theta}_M$	2.521	0.101	0.010
	$\hat{\delta}_{11}$	2.430	0.010	9.548×10^{-5}
	$\hat{\delta}_{12}$	2.411	−0.009	8.222×10^{-5}
	$\hat{\delta}_{21}$	2.498	0.078	0.006
	$\hat{\delta}_{22}$	2.476	0.056	0.003

当 $m=30$ 时情况同上;当 $m=45$ 时,给定 $R_j=1, j=3,5,10,12,13,19,21,22,24,26,29,32,34,36,42$,其余 $R_j=0$,根据定理2所得Bayes估计以及多层Bayes估计和E-Bayes估计的表达式,借助R软件对Pareto分布的形状参数进行估计。设参数 $c=2, \alpha=2, \beta=1$,模拟结果如表5所示。

表5 形状参数的三种估计结果

m	$\hat{\delta}_{12}$		$\hat{\delta}_{HB}$		$\hat{\delta}_{EB}$	
	偏差	均方误差	偏差	均方误差	偏差	均方误差
30	0.259	0.067	0.239	0.057	0.132	0.017
45	0.137	0.019	0.117	0.014	0.054	0.003

由表4可知,在样本容量固定的条件下,形状参数4个Bayes估计的均方误差和偏差值均小于极大似然估计,估计效果较好;同时,在先验分布为共轭先验分布,形状参数在二次损失函数下所得估计值的偏差最小,与真值更接近。由表5可以看出,在二次损失函数和共轭先验分布的条件下,形状参数的E-Bayes估计的均方误差更小,具有更高的估计精度,实例分析与随机模拟结论相符。

6 结论

基于逐步增加Ⅱ型截尾实验下Pareto分布的样本数据,通过极大似然估计和Bayes估计两种方法分别得到形状参数的估计表达式。对比可知,考虑样本信息和形状参数先验信息的Bayes估计具有更佳的估计效果。在Bayes估计中,考虑不同的损失函数下选取不同的先验分布也会影响形状参数的估计效果,特别地,在二次损失函数下,形状参数选共轭先验分布时所得的Bayes估计值的均方误差与偏差更小,数值模拟结果更为理想,最后结合实例验证了模拟结论。

参考文献:

[1] FREDERICO C, AYANA M. A New Class of Generalized Probability-weighted Moment Estimators for the Pareto Distribution[J]. Mathematics, 2023, 11(05): 1076-1076.

[2] ABRAVESH A, GANJI M, MOSTAFAIY B. Classical and Bayesian Estimation of Stress-strength Reliability in Type II Censored Pareto Distributions[J]. Communications in Statistics - Simulation and Computation, 2019, 48(08): 2333-2358.

[3] MING H. The E-Bayesian Estimation and Its E-MSE of Pareto Distribution Parameter under Different Loss Functions[J]. Journal of Statistical Computation and Simulation, 2020, 90(10): 1834-1848.

[4] 龙兵,张忠占.双定数混合截尾下两参数Pareto分布的统计分析[J].数学物理学报,2022,42(01):269-281.

[5] 龙兵.双边定时截尾下Pareto分布的参数估计[J].华中师范大学学报(自然科学版),2018,52(03):310-315.

[6] 周巧娟.定数截尾试验下Pareto分布参数的条件Γ-minimax估计[J].邵阳学院学报(自然科学版),2022,19(01):8-13.

[7] 刘芹,周菊玲.不同损失函数下Pareto分布参数的Bayes估计[J].数学的实践与认识,2022,52(06):160-165.

[8] 徐圣楠,张桂颖,刘洋萍.基于缺失数据下混合Pareto分布参数的估计[J].通化师范学院学报,2022,43(12):19-23.

[9] 吴月丹,徐宝.一种对称损失下两参数广义Pareto分布形状参数的Bayes分析[J].南昌大学学报(理科版),2023,47(01):21-27.

[10] 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.

[11] 李琼,武东.逐步增加Ⅱ型截尾下Pareto分布的贝叶斯分析[J].上海第二工业大学学报,2012,29(02):135-138.

[12] 邵媛媛,周菊玲,董翠玲.逐步增加Ⅱ型截尾下复合瑞利分布参数的Bayes估计[J].山东师范大学学报(自然科学版),2020,35(03):318-323.

[13] 龙兵,王芳,刁长新.逐步增加的Ⅱ型截尾下Lomax分布形状参数的估计[J].统计与决策,2017,(08):15-18.

[14] 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.

[15] 黄壹玲.不同损失函数下Poisson-Lomax分布的参数估计[D].乌鲁木齐:新疆师范大学,2020.

[16] 王秋平.Q对称熵损失下两参数Lomax分布形状参数的Bayes估计[J].佳木斯大学学报(自然科学版),2015,33(02):314-315.

[17] 姚惠,吴现荣.Linex损失下Lomax分布形状参数的几种Bayes估计[J].黔南民族师范学院学报,2012,32(06):113-116.

(下转第25页)