应用概率统计　第三十一卷
第二期　2015年4月

Chinese Journal of Applied Probability
and Statistics Vol.31 No.2 Apr. 2015



# 逐步增加首失效截尾样本下参数估计的优良性 *

刘荣玄　吴高翔　朱先阳

(井冈山大学数理学院, 吉安, 343009)

## 摘　要

在对称平方损失函数下, 利用逐步增加首失效截尾样本, 研究两参数Pareto分布族参数的一致最小方差无偏估计(UMVUE), Bayes估计和参数型经验Bayes (PEB)估计. 按照均方误差(MSE)准则, 比较UMVUE与PEB估计的优良性. 根据风险函数导出Bayes估计与PEB估计的渐近性, 并获得它们的收敛速度$o(n^{-1})$. 在相同的置信水平下, 研究参数分别在经典统计和Bayes统计中的区间估计, 并利用数值模拟说明Bayes区间估计的精度高于经典统计区间估计.

**关键词**: 首失效截尾样本, Pareto分布族, 参数估计, 优良性, 区间估计.

**学科分类号**: O212.8.

## §1.　引　　言

自1897年意大利经济学家Vilfredo Pareto (1848－1923)提出Pareto分布后(参见Arnold, 1983), 经过一个多世纪的发展, Pareto分布已广泛地应用于诸多领域, 诞生了许多有关Pareto分布的研究文献. 如: 谢天华和叶鹰(2006)在平方损失函数下研究Pareto分布形状参数的渐进最优与可容许的经验Bayes估计及其性质; 侯华蕾等(2009)在平方损失函数和LINEX损失函数下研究了Pareto分布形状参数和可靠性指标的Bayes估计; 宋立新等(2011)在与信息论中的熵函数有关的q-对称熵损失函数下研究了Pareto分布形状参数的最小风险同变估计, Bayes估计以及可容许与不可容许估计, 并给出了形状参数的一个最小最大估计. 在可靠性寿命试验中, 截尾寿命试验是一类应用广泛的试验, 针对现代工业产品寿命不断增长, 成本增大等特点, 研究者为了节约试验时间和成本, 将传统的定数截尾和定时截尾试验进行不断的改进, Wu和Kus (2009)提出了逐步增加首失效截尾寿命试验, 这种试验既可以了解产品的失效机理, 又能了解产品的退化情况, 还可以节约时间和成本. 目前利用这种试验获取样本, 研究分布函数参数估计的文献并不多见. 王亮等(2012)基于逐步增加首失效截尾样本, 在对称和非对称损失函数下研究了Burr分布可靠性指标的Bayes估计. 本文采用这种试验获取样本, 在对称平方损失函数下, 研究两参数Pareto分布参数的UMVUE, PEB估计和区间估计, 并证明其优良性和渐近性, 通过数值模拟, 比较各类区间估计的精度.

*国家自然科学基金资助项目(11161024)和江西省“十二五”规划项目(12YB049)资助.
本文2013年11月18日收到, 2014年10月14日收到修改稿.
doi: 10.3969/j.issn.1001-4268.2015.02.003

设两参数Pareto分布的分布函数和概率密度函数分别为

$$F(t)=1-\theta^{\alpha}t^{-\alpha},\quad f(t)=\alpha\theta^{\alpha}t^{-(\alpha+1)},\qquad \alpha>0,\,\theta>0,\,t>\theta, \tag{1.1}$$

其中$\alpha$为形状参数, $\theta$为尺度参数, 显然密度函数(1.1)是单调递减的.

逐步增加首失效截尾样本试验是指: 假设有$m$个独立的实验组, 每个试验组有$s$个相互独立的样品, 将这$m$个实验组中的所有样品同时投入试验, 当第一个失效样品出现时, 记录失效时间$T_{1:r:m:s}$, 其中$r$为事先规定的实验终止时失效样品的个数, 并移去含失效样品组在内的$R_1+1$个试验组; 当第二个失效样品出现时, 记录失效时间$T_{2:r:m:s}$, 又移去含失效样品组在内的$R_2+1$个实验组; 依此类推, 直至第$r$个失效样品出现为止, 记录失效时间$T_{r:r:m:s}$, 并将剩余的含失效样品组在内的$R_r+1\,(R_r=m-r-R_1-R_2-\cdots-R_{r-1})$个实验组移去. 显然, 当$s=1$, $r=m$, $R=(R_1,R_2,\ldots,R_r)=(0,0,\ldots,0)$时, 该试验为完全样本试验; 当$s=1$, $R=(0,0,\ldots,0,m-r)$时, 试验为定数截尾试验; 当$s=1$, $R=(R_1,R_2,\ldots,R_r)$时, 实验为逐次截尾实验.

为书写方便, 我们记$T_{i:r:m:s}=T_i$, $i=1,2,\ldots,r$, 令$T=(T_1,T_2,\ldots,T_r)$表示为来自Pareto分布(1.1)的逐步增加首失效截尾样本.

**引理 1.1** 假设$T_1\le T_2\le\cdots\le T_r$是一列来自总体$T\sim\mathrm{Pareto}(\theta,\alpha)$的容量为$ms$的逐步增加首失效截尾样本, 则$\ln T_1\le\ln T_2\le\cdots\le\ln T_r$可看成是来自双参数指数分布$\mathrm{Exp}(\ln\theta,\alpha)$的容量为$ms$的逐步增加首失效截尾样本. 令$Z=\sum\limits_{i=1}^{r}(R_i+1)s\ln T_i-ms\ln T_1$, 则$Z\sim\Gamma(r-1,\alpha)$.

**证明:** 设$T\sim\mathrm{Pareto}(\theta,\alpha)$, 则显然有$\ln T\sim\mathrm{Exp}(\ln\theta,\alpha)$. 所以, 如果$T_1,T_2,\ldots,T_r$是来自$T$的逐步增加首失效截尾样本, 那么$\ln T_1,\ln T_2,\ldots,\ln T_r$是来自$\ln T$的逐步增加首失效截尾样本.

根据指数分布的经典变换

$$\begin{cases} Y_1=ms[\ln T_1-\ln\theta],\\ Y_2=(m-R_1-1)s[\ln T_2-\ln T_1],\\ \quad\vdots\\ Y_r=(m-R_1-\cdots-R_{r-1}-r+1)s[\ln T_r-\ln T_{r-1}],\end{cases} \tag{1.2}$$

易知, $Y_1,Y_2,\ldots,Y_r$独立同分布于$\mathrm{Exp}(\alpha)$.

由于指数分布$\mathrm{Exp}(\alpha)$是一个伽玛分布, 即$Y_i\sim\Gamma(1,\alpha)$, $i=1,2,\ldots,r$, 根据伽玛分布的有限可加性可知

$$\sum_{i=2}^{r}Y_i=\sum_{i=1}^{r}Y_i-Y_1=\sum_{i=1}^{r}(R_i+1)s\ln T_i-ms\ln T_1=Z\sim\Gamma(r-1,\alpha).\qquad\square \tag{1.3}$$

## §2. 参数估计

### 2.1 参数的UMVUE

假设$T_1 \le T_2 \le \cdots \le T_r$是一列来自概率密度为(1.1)式的Pareto$(\theta,\alpha)$分布的逐步增加首失效截尾样本, 其似然函数为

$$L(\theta,\alpha)=cs^r\prod_{i=1}^{r}f(t_i)[1-F(t_i)]^{s(R_i+1)-1}=cs^r\alpha^r\theta^{s\alpha\sum\limits_{i=1}^{r}(R_i+1)}\prod_{i=1}^{r}t_i^{-s\alpha(R_i+1)-1}, \tag{2.1}$$

其中$c=m(m-R_1-1)(m-R_1-R_2-2)\cdots(m-R_1-R_2-\cdots-R_{r-1}-r+1)$. 对数似然函数为

$$\ln L(\theta,\alpha)=\ln c+r\ln s+r\ln\alpha+s\alpha\sum_{i=1}^{r}(R_i+1)\ln\theta-\sum_{i=1}^{r}[s\alpha(R_i+1)+1]\ln t_i,$$

参数$\alpha,\theta$的极大似然估计为

$$\widehat{\alpha}_{\rm M}=(Z/r)^{-1},\qquad \widehat{\theta}_{\rm M}=T_1,$$

其中$Z=\sum\limits_{i=1}^{r}(R_i+1)s\ln T_i-ms\ln T_1$.

**引理 2.1** 假设$T_1 \le T_2 \le \cdots \le T_r$为Pareto$(\theta,\alpha)$分布的逐步增加首失效截尾样本, 估计量$\widehat{\alpha}_{\rm M}$是参数$\alpha$的极大似然估计, 则有 (a) $T_1\sim$ Pareto$(\theta,ms\alpha)$; (b) 估计量$\widehat{\alpha}_{\rm M}$服从逆伽玛分布, 即$\widehat{\alpha}_{\rm M}\sim\Gamma^{-1}(r-1,r\alpha)$; (c) $2r\alpha/\widehat{\alpha}_{\rm M}\sim\chi^2(2r-2)$; (d) $T_1$与$\widehat{\alpha}_{\rm M}$相互独立.

**证明:** (a) 因为$T_1$为最小次序统计量, 所以它的概率密度函数为

$$g(t_1)=msf(t_1)[1-F(t_1)]^{ms-1}=ms\alpha\theta^{ms\alpha}t_1^{-(ms\alpha+1)},$$

即$T_1\sim$ Pareto$(\theta,ms\alpha)$.

(b) 设$\widehat{\alpha}_{\rm M}$的分布函数为$F(x)$, 则由引理1.1知

$$F(x)=\mathsf{P}(\widehat{\alpha}_{\rm M}\le x)=\mathsf{P}\Big(\frac{r}{Z}\le x\Big)=1-\int_0^{r/x}\frac{\alpha^{r-1}}{\Gamma(r-1)}t^{r-2}{\rm e}^{-\alpha t}{\rm d}t,$$

于是有$F'(x)=\{(\alpha r)^{r-1}/[\Gamma(r-1)]\}x^{-[(r-1)+1]}{\rm e}^{-(\alpha r)/x}$, 所以$\widehat{\alpha}_{\rm M}\sim\Gamma^{-1}(r-1,\alpha r)$.

(c) 设随机变量$2r\alpha/\widehat{\alpha}_{\rm M}$的分布函数为

$$F(x)=\mathsf{P}\Big(\frac{2r\alpha}{\widehat{\alpha}_{\rm M}}\le x\Big)=1-\mathsf{P}\Big(\widehat{\alpha}_{\rm M}\le\frac{2r\alpha}{x}\Big)=1-\int_0^{2r\alpha/x}\frac{(r\alpha)^{r-1}}{\Gamma(r-1)}t^{-[(r-1)+1]}{\rm e}^{-r\alpha/t}{\rm d}t,$$

则有$F'(x)=\{1/[2^{r-1}\Gamma(r-1)]\}x^{(r-1)-1}{\rm e}^{-x/2}$, 从而$2r\alpha/\widehat{\alpha}_{\rm M}\sim\chi^2(2r-2)$.

(d) 根据指数分布经典变换(1.2), 显然$T_1$与$\widehat{\alpha}_{\rm M}$是相互独立. □

根据参数$\theta$和$\alpha$的极大似然估计$\widehat{\theta}_{\rm M}$和$\widehat{\alpha}_{\rm M}$, 我们来构造$\theta$和$\alpha$的UMVUE. 取

$$\widehat{\theta}_{\rm U}=\Big(1-\frac{Z}{ms(r-1)}\Big)\widehat{\theta}_{\rm M}, \tag{2.2}$$

$$\widehat{\alpha}_{\rm U}=\frac{r-2}{r}\widehat{\alpha}_{\rm M}, \tag{2.3}$$

则$\widehat{\theta}_{\rm U}$, $\widehat{\alpha}_{\rm U}$分别为$\theta$, $\alpha$的无偏估计, 且有$\mathsf{Var}\,(\widehat{\theta}_{\rm U})<\infty$, $\mathsf{Var}\,(\widehat{\alpha}_{\rm U})<\infty$.

事实上, 由引理2.1可知, $T_1$与$\widehat{\alpha}_{\rm M}$相互独立, 而$\widehat{\theta}_{\rm M}=T_1$, 从而$\widehat{\theta}_{\rm M}$与$\widehat{\alpha}_{\rm M}$相互独立, 即$\widehat{\theta}_{\rm M}$与$Z$相互独立, 于是有

$$\mathsf{E}(\widehat{\theta}_{\rm U})=\mathsf{E}\Big[\Big(1-\frac{Z}{ms(r-1)}\Big)\widehat{\theta}_{\rm M}\Big]=\Big[1-\frac{1}{ms(r-1)}\mathsf{E}(Z)\Big]\mathsf{E}(\widehat{\theta}_{\rm M})=\theta, \qquad ms>\frac{1}{\alpha}. \tag{2.4}$$

$$\mathsf{Var}\,(\widehat{\theta}_{\rm U})=\mathsf{E}(\widehat{\theta}_{\rm U}^2)-[\mathsf{E}(\widehat{\theta}_{\rm U})]^2=\Big[1-\frac{2}{ms\alpha}+\frac{1}{(ms\alpha)^2(r-1)}\Big]\Big[\frac{ms\alpha}{ms\alpha-2}\Big]\theta^2<\infty, \qquad ms>\frac{2}{\alpha}. \tag{2.5}$$

$$\mathsf{E}(\widehat{\alpha}_{\rm U})=\mathsf{E}\Big(\frac{r-2}{r}\widehat{\alpha}_{\rm M}\Big)=\frac{r-2}{r}\int_0^\infty\frac{(r\alpha)^{r-1}}{\Gamma(r-1)}t^{-r+1}{\rm e}^{-r\alpha/t}{\rm d}t=\alpha, \qquad r>2. \tag{2.6}$$

$$\mathsf{Var}\,(\widehat{\alpha}_{\rm U})=\mathsf{Var}\,\Big(\frac{r-2}{r}\widehat{\alpha}_{\rm M}\Big)=\Big(\frac{r-2}{r}\Big)^2\mathsf{Var}\,(\widehat{\alpha}_{\rm M})=\frac{1}{r-3}\alpha^2<\infty, \qquad r>3. \tag{2.7}$$

现在证明, $(\widehat{\theta}_{\rm U},\widehat{\alpha}_{\rm U})$是$(\theta,\alpha)$的UMVUE.

事实上, 根据逐步增加首失效截尾样本$T_1\le T_2\le\cdots\le T_r$, 作如下变换:

$$\begin{cases} Y_1=ms\ln T_1,\\ Y_2=(m-R_1-1)s[\ln T_2-\ln T_1],\\ \quad\vdots\\ Y_r=(m-R_1-\cdots-R_{r-1}-r+1)s[\ln T_r-\ln T_{r-1}]. \end{cases} \tag{2.8}$$

由指数分布的经典变换可知, 随机变量$Y_1,Y_2,\ldots,Y_r$相互独立, 且$Y_1\sim{\rm Exp}(ms\ln\theta,\alpha)$, $Y_2$, $Y_3,\ldots,Y_r$具有相同分布${\rm Exp}(\alpha)$, 因此多维随机变量$(y_1,y_2,\ldots,y_r)$的联合概率密度为

$$f(y_1,y_2,\ldots,y_r,\ln\theta,\alpha)=\alpha^r\exp\{-\alpha(y_1-ms\ln\theta)\}\exp\Big\{-\alpha\sum_{i=2}^r y_i\Big\},$$

根据因子分解定理可知, $\Big(y_1,\sum\limits_{i=2}^r y_i\Big)$是多维随机变量$(\ln\theta,\alpha)$的充分统计量, 即$(T_1,Z)$是$(\theta,\alpha)$的充分统计量.

由于Pareto分布族$\{{\rm Pareto}(\theta,\alpha):\forall\,\theta>0,\alpha>0\}$是完备的, 因此$(T_1,Z)$还是$(\theta,\alpha)$的完备统计量, 从而可知$(T_1,Z)$是$(\theta,\alpha)$的充分完备统计量. 再由(2.2), (2.3)式可知, 估计量$\widehat{\theta}_{\rm U}$, $\widehat{\alpha}_{\rm U}$是随机向量$(T_1,Z)$的函数, 由(2.4), (2.5), (2.6), (2.7)可知$\widehat{\theta}_{\rm U}$, $\widehat{\alpha}_{\rm U}$分别为参数$\theta,\alpha$的无偏估计, 且它们的方差有界, 根据Lehmann-Scheffe定理可知, $(\widehat{\theta}_{\rm U},\widehat{\alpha}_{\rm U})$是$(\theta,\alpha)$的UMVUE.

### 2.2 参数$\alpha$的Bayes估计和PEB估计

设参数$\theta$为离散型的, 其先验分布为

$$\mathsf{P}(\theta=\theta_i)=\lambda_i,\qquad i=1,2,\ldots,N, \tag{2.9}$$

其中$N$为给定的正整数, $\theta_i,\lambda_i$为已知常数, 且$\sum\limits_{i=1}^{N}\lambda_i=1$.

设参数$\alpha$在$\theta=\theta_i$的条件下的共轭先验分布为

$$\pi(\alpha|\theta_i)=\frac{\rho^\beta}{\Gamma(\beta)}\alpha^{\beta-1}\mathrm{e}^{-\rho\alpha},\qquad \alpha>0,\ i=1,2,\ldots,N, \tag{2.10}$$

其中$\beta,\rho$为超参数, 且有$\rho>0,\ \beta>0$.

根据引理1.1及式(2.9)和(2.10)可知, 参数$\alpha$的后验条件概率密度为

$$h(\alpha|z,\theta_i)\propto f(z|\alpha,\theta_i)\pi(\alpha|\theta_i)\mathsf{P}(\theta=\theta_i)\propto\alpha^{(r+\beta)-2}\mathrm{e}^{-(z+\rho)\alpha},$$

所以, $\alpha|z,\theta_i\sim\Gamma(r+\beta-1,z+\rho)$.

根据条件分布可知二维随机向量$(\theta,\alpha)$的先验概率密度为

$$\pi^*(\theta_i,\alpha)=\pi(\alpha|\theta_i)\mathsf{P}(\theta=\theta_i)=\frac{\rho^\beta}{\Gamma(\beta)}\alpha^{\beta-1}\mathrm{e}^{-\rho\alpha}\lambda_i,\qquad i=1,2,\ldots,N, \tag{2.11}$$

于是可得$(\theta,\alpha)$的联合后验概率密度为

$$h^*(\theta_i,\alpha|z)=\frac{f(z|\alpha,\theta_i)\pi(\alpha|\theta_i)\mathsf{P}(\theta=\theta_i)}{\displaystyle\int_0^\infty\sum_{i=1}^{N}f(z|\alpha,\theta_i)\pi(\alpha|\theta_i)\mathsf{P}(\theta=\theta_i)\mathrm{d}\alpha}=\lambda_i\frac{(z+\rho)^{r+\beta-1}}{\Gamma(r+\beta-1)}\alpha^{r+\beta-2}\mathrm{e}^{-(z+\rho)\alpha}.$$

从而可知随机变量$\alpha$的后验分布为

$$f_{\alpha|z}(\alpha|z)=\sum_{i=1}^{N}h(\theta_i,\alpha|z)=\frac{(z+\rho)^{r+\beta-1}}{\Gamma(r+\beta-1)}\alpha^{r+\beta-2}\mathrm{e}^{-(z+\rho)\alpha}. \tag{2.12}$$

假设对称平方损失函数为

$$L(d,\alpha)=(d-\alpha)^2,$$

其中$d=d(z)$为$Z$的样本函数, 是参数$\alpha$的判决函数, 则参数$\alpha$的Bayes估计为

$$\widehat{\alpha}_{\mathrm{B}}=\mathsf{E}(\alpha|z)=\int_0^\infty\alpha f_{\alpha|z}(\alpha|z)\mathrm{d}z=\frac{r+\beta-1}{z+\rho}. \tag{2.13}$$

如果超参数存在未知情况, 我们可用历史样本, 采用点估计法对参数进行估计, 进而得到参数$\alpha$的PEB估计.

现假设超参数$\beta$已知, $\rho$未知, $(Z_1,\alpha_1),(Z_2,\alpha_2),\ldots,(Z_n,\alpha_n),(Z_{n+1}=Z,\alpha)$为一串独立同分布的随机向量, $Z_1,Z_2,\ldots,Z_n$是可观测的有共同边缘概率密度的历史样本, $Z_{n+1}$为当前样本, $\alpha_1,\alpha_2,\ldots,\alpha_n$和$\alpha$为不可观测的, 但有相同的先验分布.

由引理1.1和式(2.11)式可知

$$\mathsf{E}(Z)=\int_0^\infty\int_0^\infty\sum_{i=1}^N zf(z|\theta_i,\alpha)\pi(\theta_i,\alpha)\mathrm{d}\alpha\mathrm{d}z=\frac{r-1}{\beta-1}\rho.$$

设$\overline{Z}$为样本均值, 即$\overline{Z}=n^{-1}\sum\limits_{i=1}^n Z_i$, 用$\overline{Z}$代替$\mathsf{E}(Z)$, 于是得到参数$\rho$的估计为

$$\widehat{\rho}=\frac{\beta-1}{r-1}\overline{Z}. \tag{2.14}$$

因为

$$\mathsf{E}_n(\widehat{\rho})=\mathsf{E}_n\Big(\frac{\beta-1}{r-1}\overline{Z}\Big)=\frac{\beta-1}{r-1}\mathsf{E}_n(\overline{Z})=\rho,$$

其中$\mathsf{E}_n$表示对随机向量$(Z_1,Z_2,\ldots,Z_n)$求数学期望, 所以估计量$\widehat{\rho}$还是超参数$\rho$的无偏估计.

将$\widehat{\rho}$替换(2.13)式中的$\rho$, 得到参数$\alpha$的PEB估计为

$$\widehat{\alpha}_{\mathrm{E}}=\frac{(r+\beta-1)(r-1)}{(r-1)z+(\beta-1)\overline{z}}. \tag{2.15}$$

如果超参数$\beta,\rho$均未知, 同样可以采用矩估计方法得到参数$\beta$与$\rho$的估计$\widehat{\beta}$与$\widehat{\rho}$, 再用它们替换(2.13)式中的$\beta,\rho$, 就可得到参数$\alpha$的PEB估计.

## §3. 估计的优良性

**定理 3.1** 设$\widehat{\alpha}_{\mathrm{U}}$和$\widehat{\alpha}_{\mathrm{E}}$分别为参数$\alpha$的UMVUE和PEB估计, 当$r>4$, $\beta>2$时, 则有

$$\lim_{n\to\infty}[\mathrm{MSE}(\widehat{\alpha}_{\mathrm{E}})-\mathrm{MSE}(\widehat{\alpha}_{\mathrm{U}})]<0.$$

**证明:** 因为

$$\begin{aligned}
&\mathrm{MSE}(\widehat{\alpha}_{\mathrm{E}})-\mathrm{MSE}(\widehat{\alpha}_{\mathrm{U}})\\
=&\ \mathsf{E}_*[(\widehat{\alpha}_{\mathrm{E}}-\alpha)^2-(\widehat{\alpha}_{\mathrm{U}}-\alpha)^2]\\
=&\ \mathsf{E}_*[\widehat{\alpha}_{\mathrm{E}}^2-2\alpha\widehat{\alpha}_{\mathrm{E}}+2\alpha\widehat{\alpha}_{\mathrm{U}}-\widehat{\alpha}_{\mathrm{U}}^2]\\
=&\ \mathsf{E}_m[\mathsf{E}_\alpha(\widehat{\alpha}_{\mathrm{E}}^2-2\alpha\widehat{\alpha}_{\mathrm{E}}+2\alpha\widehat{\alpha}_{\mathrm{U}}-\widehat{\alpha}_{\mathrm{U}}^2)|(z_1,z_2,\ldots,z_n,z)]\\
=&\ -[\mathsf{E}_m(\widehat{\alpha}_{\mathrm{B}}-\widehat{\alpha}_{\mathrm{U}})^2-\mathsf{E}_m(\widehat{\alpha}_{\mathrm{E}}-\widehat{\alpha}_{\mathrm{B}})^2]\\
=&\ (r+\beta-1)^2\mathsf{E}_m\Big(\frac{\rho-\widehat{\rho}}{(z+\widehat{\rho})(z+\rho)}\Big)^2-\mathsf{E}_m\Big(\frac{r+\beta-1}{z+\rho}-\frac{r-2}{z}\Big)^2.
\end{aligned}$$

由于$Z_1, Z_2, \ldots, Z_n$和$(z,\alpha)$相互独立, 可知$z$与$\widehat{\rho}$独立, 故有

$$\mathrm{MSE}(\widehat{\alpha}_{\mathrm{E}}) - \mathrm{MSE}(\widehat{\alpha}_{\mathrm{U}}) \le (r+\beta-1)^2 \mathsf{E}_m(\rho-\widehat{\rho})^2 \mathsf{E}_m\Big(\frac{1}{z^4}\Big) - \mathsf{E}_m\Big(\frac{r+\beta-1}{z+\rho} - \frac{r-2}{z}\Big)^2,$$

其中$\mathsf{E}_\alpha, \mathsf{E}_*$和$\mathsf{E}_m$分别表示对随机变量$\alpha$, 随机向量$(Z_1, Z_2, \ldots, Z_n, (Z,\alpha))$和$(Z_1, Z_2, \ldots, Z_n, Z)$求数学期望.

又因为

$$\begin{aligned}\mathsf{E}_m[(\widehat{\rho}-\rho)^2] &= \mathsf{Var}\,(\widehat{\rho}) = \Big(\frac{\beta-1}{r-1}\Big)^2 \mathsf{Var}\,(\overline{Z}) \\ &= \Big(\frac{\beta-1}{r-1}\Big)^2 \frac{1}{n^2} \sum_{i=1}^n \mathsf{Var}\,(Z_i) = \Big(\frac{\beta-1}{r-1}\Big)^2 \frac{\mathsf{Var}\,(Z)}{n},\end{aligned}$$

由(2.11)可知

$$\begin{aligned}\mathsf{E}\Big(\frac{1}{\alpha}\Big) &= \int_0^\infty \sum_{i=1}^N \frac{1}{\alpha}\pi^*(\theta_i,\alpha)\mathrm{d}\alpha = \frac{\rho}{\beta-1}, \\ \mathsf{E}\Big(\frac{1}{\alpha^2}\Big) &= \int_0^\infty \sum_{i=1}^N \frac{1}{\alpha^2}\pi^*(\theta_i,\alpha)\mathrm{d}\alpha = \frac{\rho^2}{(\beta-1)(\beta-2)}.\end{aligned}$$

由引理1.1可知$Z|\alpha \sim \Gamma(r-1,\alpha)$, 所以

$$\begin{aligned}\mathsf{E}(Z) &= \mathsf{E}_\alpha[\mathsf{E}_Z(Z|\alpha)] = \mathsf{E}_\alpha\Big(\frac{r-1}{\alpha}\Big) = \frac{r-1}{\beta-1}\rho, \\ \mathsf{E}(Z^2) &= \mathsf{E}_\alpha[\mathsf{E}_Z(Z^2|\alpha)] = \mathsf{E}_\alpha\Big(\frac{r(r-1)}{\alpha^2}\Big) = \frac{r(r-1)\rho^2}{(\beta-1)(\beta-2)}, \\ \mathsf{Var}\,(Z) &= \mathsf{E}(Z^2) - [\mathsf{E}(Z)]^2 = \frac{(r-1)(r+\beta-2)}{(\beta-1)^2(\beta-2)}\rho^2.\end{aligned}$$

从而可知

$$\lim_{n\to\infty} \mathsf{E}_m(\rho-\widehat{\rho})^2 = 0.$$

因为

$$\mathsf{E}_m\Big(\frac{1}{z^4}\Big) = \mathsf{E}_\alpha\Big[\mathsf{E}_Z\Big(\frac{1}{z^4}\Big|\alpha\Big)\Big] = \frac{(\beta+1)(\beta+2)(\beta+3)}{(r-2)(r-3)(r-4)\rho^4} > 0.$$

显然

$$\mathsf{E}_m\Big(\frac{r+\beta-1}{z+\rho} - \frac{r-2}{z}\Big)^2 > 0.$$

所以当$r, \beta, \rho$分别选取某数时, 下列各式

$$\mathsf{E}_m\Big(\frac{r+\beta-1}{z+\rho} - \frac{r-2}{z}\Big)^2, \qquad (r+\beta-1)^2, \qquad \mathsf{E}_m\Big(\frac{1}{z^4}\Big)$$

分别为某个确定的正数, 取$\varepsilon$, 满足

$$0 < \varepsilon < \mathsf{E}_m\Big(\frac{r+\beta-1}{z+\rho} - \frac{r-2}{z}\Big)^2 \Big/ (r+\beta-1)^2 \mathsf{E}_m(z^{-4}),$$

则一定存在一个确定的正数$N$, 当样本容量为$N+1$或大于$N+1\,(n>N)$时, 有

$$\mathsf{E}_m(\rho-\widehat{\rho})^2<\mathsf{E}_m\Big(\frac{r+\beta-1}{z+\rho}-\frac{r-2}{z}\Big)^2\Big/(r+\beta-1)^2\mathsf{E}_m(z^{-4}),$$

所以有$\lim\limits_{n\to\infty}[\mathrm{MSE}(\widehat{\alpha}_{\mathrm{E}})-\mathrm{MSE}(\widehat{\alpha}_{\mathrm{U}})]<0.$ □

## §4. 估计的渐近性

**引理 4.1** 设$Z_1,Z_2,\ldots,Z_n$为来自总体$Z$的历史样本, $\overline{Z}$为样本均值, $\widehat{\rho}$由(2.14)式给出, 则有

$$\mathsf{E}_n[(\widehat{\rho}-\rho)^2]=o(n^{-1}).$$

证明见上述.

**定理 4.1** 假设$R_{\mathrm{E}},R_{\mathrm{B}}$分别为参数$\alpha$的PEB估计和Bayes估计的Bayes风险, 当$r>5$, $\beta>2$, $\rho>0$时, 则有

$$R_{\mathrm{E}}-R_{\mathrm{B}}=o(n^{-1}).$$

**证明:** 由(2.13), (2.15)及王立春和韦来生(2002, §3, 引理3.1)可知, 因为

$$\begin{aligned}&\mathsf{E}_*[(\widehat{\alpha}_{\mathrm{E}}-\widehat{\alpha}_{\mathrm{B}})(\widehat{\alpha}_{\mathrm{B}}-\alpha)]\\=&\ \mathsf{E}_{(Z_1,Z_2,\ldots,Z_n,Z)}\{\mathsf{E}_\alpha[(\widehat{\alpha}_{\mathrm{E}}\widehat{\alpha}_{\mathrm{B}}+\alpha\widehat{\alpha}_{\mathrm{B}}-\widehat{\alpha}_{\mathrm{B}}^2-\alpha\widehat{\alpha}_{\mathrm{E}})|Z_1,Z_2,\ldots,Z_n,Z]\}\\=&\ \mathsf{E}_{(Z_1,Z_2,\ldots,Z_n,Z)}[\widehat{\alpha}_{\mathrm{E}}\widehat{\alpha}_{\mathrm{B}}+\widehat{\alpha}_{\mathrm{B}}\mathsf{E}(\alpha|Z)-\widehat{\alpha}_{\mathrm{B}}^2-\widehat{\alpha}_{\mathrm{E}}\mathsf{E}(\alpha|Z)]=0,\end{aligned}$$

所以

$$\begin{aligned}R_{\mathrm{E}}-R_{\mathrm{B}}&=\mathsf{E}_*[(\widehat{\alpha}_{\mathrm{E}}-\alpha)^2]-\mathsf{E}_*[(\widehat{\alpha}_{\mathrm{B}}-\alpha)^2]\\&=\mathsf{E}_*[(\widehat{\alpha}_{\mathrm{E}}-\widehat{\alpha}_{\mathrm{B}})^2]\\&=(r+\beta-1)^2\mathsf{E}_*\Big[\Big(\frac{1}{z+\widehat{\rho}}-\frac{1}{z+\rho}\Big)^2\Big]\\&=(r+\beta-1)^2\mathsf{E}_*\Big[\Big(\frac{\rho-\widehat{\rho}}{(z+\widehat{\rho})(z+\rho)}\Big)^2\Big]\\&\le(r+\beta-1)^2\mathsf{E}_*\Big(\frac{1}{z^4}\Big)\mathsf{E}_*(\widehat{\rho}-\rho)^2\\&\le(r+\beta-1)^2\frac{(\beta+3)(\beta+2)(\beta+1)\beta}{(r-2)(r-3)(r-4)(r-5)}\frac{1}{\rho^4}\mathsf{E}_*(\widehat{\rho}-\rho)^2\\&=o(n^{-1}).\end{aligned}$$

由此可知, 当$n$较大时参数$\alpha$的Bayes风险可以近似代替PEB估计的Bayes风险. □

关于参数$\theta$的Bayes估计, PEB估计, 及其优良性和渐近性可以类似得到.

## §5.　区间估计

在经典统计中, 由引理2.1可知随机变量$2\alpha Z \sim \chi^2(2r-2)$, 假设参数$\alpha$的置信水平为$1-\sigma$, 则

$$\mathsf{P}\big(\chi^2_{1-\sigma/2}(2r-2) < 2\alpha Z < \chi^2_{\sigma/2}(2r-2)\big) = 1-\sigma,$$

于是可得参数$\alpha$的置信水平为$1-\sigma$的置信区间为

$$\Big(\frac{\chi^2_{1-\sigma/2}(2r-2)}{2Z}, \frac{\chi^2_{\sigma/2}(2r-2)}{2Z}\Big).$$

在Bayes统计中, 参数$\alpha$为随机变量, 由式(2.12)可知, 它的后验分布为$\alpha|Z \sim \Gamma(r+\beta-1, z+\rho)$, 令$W = 2(Z+\rho)\alpha$, 则由李翔和韦来生(2011)可知, 在$Z$已知的条件下, 随机变量$W$的条件概率密度为

$$f_{W|Z}(w|z) = \frac{1}{2^{(r+\beta-1)}\Gamma(r+\beta-1)} w^{(r+\beta-1)-1} \mathrm{e}^{-w/2},$$

即$W|Z \sim \chi^2(2(r+\beta-1))$. 于是可得参数$\alpha$的可信水平为$1-\sigma$的可信区间为

$$\Big(\frac{\chi^2_{1-\sigma/2}(2(r+\beta-1))}{2(z+\rho)}, \frac{\chi^2_{\sigma/2}(2(r+\beta-1))}{2(z+\rho)}\Big). \tag{5.1}$$

如果参数$\alpha$的先验分布中的超参数$\beta$已知, $\rho$未知(其它情况类似), $\rho$的估计量为(2.14)式, 于是可得$\alpha$的可信水平近似为$1-\sigma$的PEB可信区间为

$$\Big(\frac{\chi^2_{1-\sigma/2}(2(r+\beta-1))}{2(z+\widehat{\rho})}, \frac{\chi^2_{\sigma/2}(2(r+\beta-1))}{2(z+\widehat{\rho})}\Big). \tag{5.2}$$

它的近似程度从下列数值模拟获知.

## §6.　随机模拟

由(2.14)式可知, $\widehat{\rho}$是$\rho$的矩估计, 根据矩估计的性质, 显然, 当参数$\alpha$的Bayes区间估计的可信水平为$1-\sigma$时, PEB可信区间的可信水平以$o(n^{-1})$的收敛速度收敛于$1-\sigma$. 我们取$n=50$, $r=10$, 进行数值模拟, 来说明PEB可信区间(5.1)覆盖参数$\alpha$的水平.

假设Pareto分布的形状参数$\alpha=2$, 超参数$\beta=4$, 可信水平$1-\sigma=0.95$. 在计算机上采用随机模拟的方法生成Pareto分布的, 容量$n=50$, 失效数$r=20$的逐步增加首失效截尾样本, 按照引理1.1, 得到随机变量$Z$的一个样品$Z_1$, 依此做法反复50次, 就可得到$Z$的, 容量为50的样本, $Z_1, Z_2, \ldots, Z_{50}$, 将样本观察值代入$\alpha$的PEB可信区间(5.1), 得到$\alpha$的一个可信区间. 在计算机上重复上述实验100次, 得到$Z$的容量为50的100个样本和$\alpha$的100个可信区间, 在计算机上将参数$\alpha$与其PEB可信区间的关系模拟图形如下:

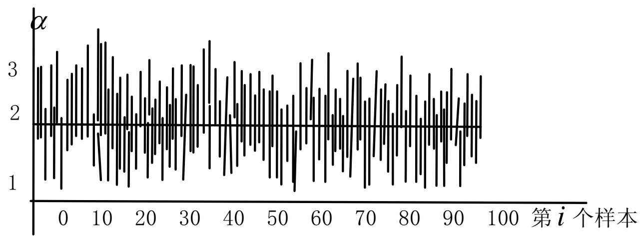


图1 PEB可信区间, 其中线条"|"表示区间

从图1中统计它们之间的关系, 100个区间中有93个区间包含参数$\alpha$. 模拟结果显示, $\alpha$的PEB可信区间的覆盖水平达到93%, 这与$\alpha$的Bayes区间估计的可信水平0.95接近. 随着$n, r$的增大, 接近0.95的程度会越高. 从而可通过它们的区间长度来比较区间估计的精度.

上述几类区间估计都选取相同的可信水平0.95, 在进行数值模拟时, 比较它们的可信区间长度, 长度越短, 区间估计越优. 表1为数值模拟结果(表中的数字表示经典统计的可信区间长度减去PEB可信区间的长度的差).

表1 数值模拟表

| $n$ | $\beta$ | 区间差 | | | | |
|---|---|---|---|---|---|---|
| | | $r=10$ | $r=15$ | $r=20$ | $r=25$ | $r=30$ |
| 50 | 2 | 0.2265 | 0.0235 | 0.0079 | 0.0029 | 0.0007 |
| | 4 | 0.0968 | 0.0141 | 0.0052 | 0.0019 | 0.0005 |
| | 8 | 0.0467 | 0.0090 | 0.0038 | 0.0014 | 0.0004 |
| 100 | 2 | 0.2258 | 0.0213 | 0.0066 | 0.0027 | 0.0006 |
| | 4 | 0.0908 | 0.0126 | 0.0050 | 0.0016 | 0.0004 |
| | 8 | 0.0469 | 0.0090 | 0.0041 | 0.0014 | 0.0004 |
| 200 | 2 | 0.2165 | 0.0212 | 0.0077 | 0.0028 | 0.0007 |
| | 4 | 0.0918 | 0.0135 | 0.0040 | 0.0019 | 0.0005 |
| | 8 | 0.0458 | 0.0090 | 0.0040 | 0.0015 | 0.0004 |

从表1可以看出, 随着历史样本容量$n$的增大, 区间估计差变小, 但变化不大: 实验终止时失效样品的个数$r$对区间估计的影响较大, 随着$r$的增大, 经典统计和PEB得到的置信区间的长度越来越接近. 从表1中同样可以发现超参数$\beta$的变化对区间估计有较大的影响.

由此可知, PEB的区间估计要优于经典统计的区间估计.

## 参 考 文 献

[1] Arnold, B.C., *Pareto Distribution*, International Cooperative Publishing House, Fairland, Maryland, 1983.

[2] 谢天华, 叶鹰, Pareto分布参数的渐进最优与可容许的经验Bayes估计, 应用数学, **19(增)**(2006), 237–240.

[3] 侯华蕾, 师义民, 李豪亮, 双边定数截尾下Pareto分布的可靠性, 数理统计与管理, **28(5)**(2009), 826–830.

[4] 宋立新, 王明秋, 王晓光, q-对称熵损失函数下Pareto分布参数估计, 大连理工大学学报, **51(4)**(2011), 616–620.

[5] Wu, S.J. and Kus, C., On estimation based on progressive first-failure-censored sampling, *Computational Statistics and Data Analysis*, **53(10)**(2009), 3659–3670.

[6] 王亮, 师义民, 孙天宇, 常萍, 两参数Pareto分布逐步首失效样本的Bayes估计, 系统工程理论与实践, **32(11)**(2012), 2498–2503.

[7] 王立春, 韦来生, 刻度指数族参数的经验Bayes估计的收敛速度, 数学年刊, **23A(5)**(2002), 555–564.

[8] 李翔, 韦来生, 指数分布定数截尾数据下刻度参数的经验Bayes估计, 中国科学院研究生院学报, **28(2)**(2011), 147–154.



# Optimal Property of Parametric Estimation under Progressively First-Failure-Censored Samples

LIU RONGXUAN　　WU GAOXIANG　　ZHU XIANYANG

(*School of Mathematics and Physics, Jinggangshan University, Ji'an, 343009*)

Under the symmetric loss functions, by means of progressively first-failure-censored samples, the paper studies the uniformly minimum variance unbiased estimation (UMVUE), Bayes estimation and parametric empirical Bayes estimation (PEB) based on two-parameter Pareto distribution. According to the code of mean squared error (MSE), the gradualism of parametric Bayes and PEB estimation is investigated by applying risk function and by comparing the optimal property between UMVUE and PEB estimations to obtain their convergence rate. Based on the same confidence level, parametric interval estimation in classical and Bayes statistics is analyzed. A conclusion can be made that the precision of interval estimation in Bayes statics is higher than that in classical statics by means of numerical simulation.

**Keywords:** First-failure-censored samples, Pareto distribution, parameters estimation, optimal property, interval estimation.

**AMS Subject Classification:** 62C12, 62F05.