应用概率统计　第三十一卷
第三期　2015年6月

Chinese Journal of Applied Probability
and Statistics Vol.31 No.3 Jun. 2015

帕累托索赔分布中风险参数的经验贝叶斯估计 *

温利民[1,2★]　张　美[1]　程子红[1]　章　溢[1,3]

([1]江西师范大学数学与信息科学学院, 南昌, 330022; [2]江西财经大学信息管理学院, 南昌, 330013)

([3]江西师范大学计算机信息工程学院, 南昌, 330022)

摘　要

本文建立了贝叶斯模型, 讨论了帕累托索赔额分布中参数的估计问题, 得到了风险参数的极大似然估计、贝叶斯估计和信度估计, 并证明了这些估计的强相合性. 在均方误差的意义下比较了这些估计的好坏, 并通过数值模拟对均方误差进行了验证, 结果表明, 贝叶斯估计比其他估计具有较小的均方误差. 最后, 给出了结构参数的估计并证明了经验贝叶斯估计和经验贝叶斯信度估计的渐近最优性.

关键词: 帕累托分布, 经验贝叶斯估计, 渐近最优, 信度估计.

学科分类号: O211.9.

§1. 引　　言

帕累托(Pareto)分布是意大利经济学家Vilfredo Pareto在研究经济统计资料时发现的, 因此后来称之为帕累托分布. 随着数学和相关学科的发展, 帕累托分布不仅仅应用到经济收入模型中, 也应用生物科学、可靠性理论等其他模型中. 相关研究包括He等(2014), Tudor (2014), Fahidy (2011), Dixit和Nooghabi (2011)等文献. 由于帕累托分布具有递减的失效率函数, 故常常用来描述个人收入、某种药理过程后病人的存活时间、股票价格波动、保险风险、商业失效等模型. 在目前的各个领域内都有重要的应用, Harris (1968)和Arnold (1983)对帕累托分布进行了详细的介绍. 帕累托分布或与其相近的分布被经济学家Steindl (1965)和Hagstroem (1960)用来解释一些经济现象.

在非寿险精算领域, 常用帕累托分布来描述再保险或具有免赔额保险的索赔分布. 关于帕累托分布在精算中的运用得到了广泛的研究, 可参考Ramsay (2003), Albrecher和Kortschak (2009), Brazauskas和Kleefeld (2009)等. 一般地, 设有n种保单, 设第i份保单有m_i年的索赔记录. 记$X_{i,j}$为第i种保单在第j年的索赔, 则$i=1,2,\ldots,n$, $j=1,2,\ldots,m_i$. 记$X_i=(X_{i1},X_{i2},\ldots,X_{i,m_i})'$. 若假设$X_{i,j}$来自帕累托分布的样本, 具有相同的自留额$x_0$, 其

*国家自然科学基金(71361015)、中国博士后基金面上资助项目(2013M540534)、中国博士后基金特别资助项目(2014T70615)、江西省博士后择优项目(2013KY53)和江西省自然科学基金(20142BAB201013)资助.

★通讯作者, E-mail: wlmjxnu@163.com.

本文2013年1月21日收到, 2014年12月18日收到修改稿.

doi: 10.3969/j.issn.1001-4268.2015.03.001

分布函数为

$$F_{X_{ij}}(x) = 1 - \left(\frac{x}{x_0}\right)^{-\theta_i}, \qquad x > x_0. \tag{1.1}$$

精算师关心的是每份保单的风险参数θ_i的估计, 进而可以对该分布的一些特征进行相应的统计推断, 例如估计保单的保费、需要提取的责任准备金等.

本文在Gamma先验分布下研究帕累托索赔分布参数的估计及其性质, 比较各种估计的优劣, 并证明经验贝叶斯估计和经验贝叶斯信度估计的渐近最优性. 风险参数θ_i的先验分布取为Gamma分布主要基于下面几个方面的原因. 首先, 由于风险参数θ_i本身取值的连续性和非负性, 恰适用于Gamma分布; Gamma分布中包含形状参数α和尺度参数β, 是一个比较大的分布指数族分布类, 当取不同的α和β值时可退化为指数分布、卡方分布等多种常用的分布, 因此是概率统计中先验分布的较佳选择, 类似的研究可参考Gómez-Déniz等(2006), Al-Saleh和Agarwal (2007), Griffin和Brown (2010), Meczarski和Zieliński (1991)等文献; 注意到在帕累托分布中, 风险参数θ_i的信息为

$$I(\theta_i) = \mathsf{E}\Big[\Big(-\frac{\partial^2 l}{\partial \theta_i^2}\Big)\Big|\theta\Big] = \frac{m_i}{\theta_i^2},$$

则θ_i的Jeffrey无信息先验分布可取为$\pi_{\theta_i}(\theta) \propto [I(\theta)]^{1/2} = 1/\theta$. 显然, Jeffrey无信息先验分布是Gamma分布当$\alpha \to 0$后当$\beta \to 0$时的近似; 另外, 在本文的第二节可以看出, Gamma分布正是帕累托分布的共轭先验分布, 这使得风险参数估计有较好的统计性质. 关于先验分布的选择问题可参考茆诗松等(1998), Lehmann和Casella (2003), Walker (2004), Gelman等(1995)等.

本文后面的内容安排如下, 第二节给出相关的假设和定义, 并讨论帕累托分布中参数的各种估计, 第三节证明估计的强相合性, 并利用数值模拟的方法比较各种估计的均方误差, 第四节讨论结构参数的估计, 并证明经验贝叶斯估计的渐近最优性, 最后一节给出了文章的结论.

§2. 帕累托分布中风险参数的几个估计

帕累托分布是保险精算中的一种重要的分布, 由于其分布的特殊性, 常常用来刻画具有免赔额的保单风险的索赔额分布. 假设x_0为保单风险的免赔额, 在保险中, 一般x_0是已知的常数. 假设θ_i为第i个保单的风险参数, 在θ_i给定条件下, 该保单在第j年的索赔服从帕累托分布. 由于风险的非齐次性, 不同的保单的风险参数是不相同的. 每份保单的风险参数的取值将形成一个分布, 因此θ_i为随机变量, 而θ_i的取值形成的分布称为先验分布. 对θ_i的统计推断就落入了贝叶斯框架.

假设 2.1 设X_{ij}为第i种保单在第j年的索赔额, 且假设第i个风险参数θ_i给定的条件下, X_{ij}, $j = 1, 2, \ldots, m_i$是来自帕累托分布的样本, 具有分布函数(1.1), 其中θ_i为该保单的

风险参数, 而x_0是已知的保单免赔额.

假设 2.2 设θ_i为相互独立的随机变量且有共同的Gamma分布, 先验密度函数为

$$\pi(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}\mathrm{e}^{-\lambda\theta}, \qquad \theta > 0, \tag{2.1}$$

其中$\alpha > 2$, $\beta > 0$为结构参数.

为方便, 记$X_i = (X_{i1}, \ldots, X_{i,m_i})'$表示第$i$个保单合同的损失样本, 而$X = (X_1', \ldots, X_n')'$表示所有保单损失的样本. 在统计学中, 参数$\theta_i$的一个重要的估计是极大似然估计, 记为$\widehat{\theta}_i^{\,\mathrm{mle}}$. 由于给定$\theta_i$下, X_{ij}的密度函数为

$$f_{X_{ij}}(x|\theta_i) = \frac{\theta_i}{x_0}\left(\frac{x}{x_0}\right)^{-(\theta_i+1)}, \qquad x > x_0, \tag{2.2}$$

则似然估计$\widehat{\theta}_i^{\,\mathrm{mle}}$为最大化下面的似然函数:

$$L(x;\theta_1,\ldots,\theta_n) = \prod_{i=1}^{n}\prod_{j=1}^{m_i}[f_{X_{ij}}(x_{ij}|\theta_i)] = \prod_{i=1}^{n}\left[\left(\frac{\theta_i}{x_0}\right)^{m_i}\cdot\prod_{j=1}^{m_i}\left(\frac{x_{ij}}{x_0}\right)^{-(\theta_i+1)}\right] \tag{2.3}$$

的解. 由极大似然估计的一般求解方法(可参考Lehmann和Casella, 2003), 容易得到

$$\widehat{\theta}_i^{\,\mathrm{mle}} = \frac{m_i}{\sum\limits_{j=1}^{m_i}\ln\left(\dfrac{X_{ij}}{x_0}\right)}. \tag{2.4}$$

注意到, θ_i的极大似然估计$\widehat{\theta}_i^{\,\mathrm{mle}}$仅仅与第$i$个保单的损失$X_i$有关, 而与其他保单的损失样本无关. 其主要原因是假设了各个保单的索赔相互独立导致的. 另一方面, 根据贝叶斯定理(见Gómez-Déniz等, 2006), 在平方损失函数下, θ_i的最优估计为θ_i的后验均值:

$$\widehat{\theta}_i^{\,B} = \mathsf{E}(\theta_i|X_1, X_2, \ldots, X_n). \tag{2.5}$$

命题 2.1 若风险满足假设2.1和假设2.2, 则风险参数θ_i的贝叶斯估计为

$$\widehat{\theta}_i^{\,B} = q_i\widehat{\theta}_i^{\,\mathrm{mle}} + (1-q_i)\frac{\alpha}{\beta}, \tag{2.6}$$

其中

$$q_i = \frac{\sum\limits_{j=1}^{m_i}\ln\left(\dfrac{X_{ij}}{x_0}\right)}{\sum\limits_{j=1}^{m_i}\ln\left(\dfrac{X_{ij}}{x_0}\right)+\beta}. \tag{2.7}$$

证明: 根据贝叶斯估计的一般求法, 容易得到θ_i的贝叶斯估计为

$$\widehat{\theta}_i^{\,B} = \frac{m_i+\alpha}{\sum\limits_{j=1}^{m_i}\ln\left(\dfrac{X_{ij}}{x_0}\right)+\beta}. \tag{2.8}$$

由极大似然估计(2.4), 则θ_i的贝叶斯估计$\widehat{\theta_i}^B$可以写成下式:

$$\begin{aligned}\widehat{\theta_i}^B &= \frac{\sum\limits_{j=1}^{m_i}\ln\left(\dfrac{X_{ij}}{x_0}\right)}{\sum\limits_{j=1}^{m_i}\ln\left(\dfrac{X_{ij}}{x_0}\right)+\beta}\frac{m_i}{\sum\limits_{j=1}^{m_i}\ln\left(\dfrac{X_{ij}}{x_0}\right)}+\frac{\beta}{\sum\limits_{j=1}^{m_i}\ln\left(\dfrac{X_{ij}}{x_0}\right)+\beta}\frac{\alpha}{\beta}\\ &= q_i\widehat{\theta_i}^{\text{mle}}+(1-q_i)\frac{\alpha}{\beta}. \qquad \Box\end{aligned}$$

注记 1 根据上面的命题2.1, 注意到$\mathsf{E}(\theta_i)=\alpha/\beta$, 则风险参数$\theta_i$的贝叶斯估计$\widehat{\theta_i}$可以看成$\theta_i$的极大似然估计$\widehat{\theta_i}^{\text{mle}}$和先验均值$\mathsf{E}(\theta_i)$的加权平均值, 且其权重满足$0<q_i<1$. 这与信度理论中的信度保费估计有非常类似的结论. 但注意到, 这里权重q_i是依赖于观测值X_{ij}, $j=1,2,\ldots,m_i$, 因此是随机变量; 而信度估计中的信度因子是不依赖于样本的非随机变量, 因此两者有显著的差别.

令$Y_{ij}=\ln(X_{ij}/x_0)$, 则在θ_i给定条件下, Y_{ij}相互独立且其条件密度为

$$f_{Y_{ij}}(y|\theta_i)=\theta_i\exp(-\theta_i y), \qquad y>0.$$

即Y_{ij}服从参数θ_i的指数分布. 令

$$\mu(\theta_i)=\mathsf{E}(Y_{ij}|\theta_i)=\frac{1}{\theta_i}, \qquad \sigma^2(\theta_i)=\mathsf{Var}\,(Y_{ij}|\theta_i)=\frac{1}{\theta_i^2}. \tag{2.9}$$

并引入记号$\mu=\mathsf{E}[\mu(\theta_i)]=\beta/(\alpha-1)$以及

$$\tau^2=\mathsf{Var}\,(\mu(\theta_i))=\frac{\beta^2}{(\alpha-1)^2(\alpha-2)}, \qquad \sigma^2=\mathsf{E}[\sigma^2(\theta_i)]=\frac{\beta^2}{(\alpha-1)(\alpha-2)}. \tag{2.10}$$

根据信度理论的思想(见Bühlmann和Gisler, 2005; 郑丹等, 2012; 方婧等, 2012), 若把$\mu(\theta_i)$的估计限定在样本$\{Y_{ij}, i=1,2,\ldots,n, j=1,2,\ldots,m_i\}$的非齐次线性组合中, 即求解最优化问题

$$\min_{b_0,b_{sj}\in R}\mathsf{E}\Big[\Big(\mu(\theta_i)-b_0-\sum_{s=1}^{n}\sum_{j=1}^{m_s}b_{sj}Y_{sj}\Big)^2\Big]. \tag{2.11}$$

则得到著名的Bühlmann非齐次信度估计:

$$\widehat{\mu(\theta_i)}^c=Z_i\overline{Y}_i+(1-Z_i)\mu, \tag{2.12}$$

其中

$$\overline{Y}_i=\frac{1}{m_i}\sum_{j=1}^{m_i}Y_{ij}, \qquad Z_i=\frac{m_i\tau^2}{m_i\tau^2+\sigma^2}=\frac{m_i}{m_i+\alpha-1}. \tag{2.13}$$

由于$\mu(\theta_i)=1/\theta_i$, 由方程$1/\theta_i=Z_i\overline{Y}_i+(1-Z_i)\mu$得到$\theta_i$的另一个估计:

$$\widehat{\theta_i}^c=\frac{1}{Z_i\overline{Y}_i+(1-Z_i)\mu}=\frac{m_i+\alpha-1}{\sum\limits_{j=1}^{m_i}Y_{ij}+\beta}=\frac{m_i+\alpha-1}{\sum\limits_{j=1}^{m_i}\ln\left(\dfrac{X_{ij}}{x_0}\right)+\beta}. \tag{2.14}$$

因为估计$\widehat{\theta}_i^{\,c}$是由非齐次信度估计$\widehat{\mu(\theta_i)}^{\,c}$得来的, 因此也称之为非齐次信度估计.

同理, 若把$\mu(\theta_i)$的估计限定在样本的齐次线性组合中, 且要求估计的无偏性, 即求解下面的最小化问题:

$$\begin{cases}\min\limits_{b_{sj}\in R}\mathsf{E}\Big[\Big(\mu(\theta_i)-\sum\limits_{s=1}^{n}\sum\limits_{j=1}^{m_s}b_{sj}Y_{sj}\Big)^2\Big]\\ \mathsf{E}[\mu(\theta_i)]=\mathsf{E}\Big[\sum\limits_{s=1}^{n}\sum\limits_{j=1}^{m_s}b_{sj}Y_{sj}\Big]\end{cases}. \tag{2.15}$$

则得到Bühlmann齐次信度估计:

$$\widehat{\mu(\theta_i)}^{\,\text{hom}}=Z_i\overline{Y}_i+(1-Z_i)\widehat{\mu}, \tag{2.16}$$

其中

$$\widehat{\mu}=\frac{\sum\limits_{i=1}^{n}Z_i\overline{Y}_i}{\sum\limits_{i=1}^{n}Z_i}=\frac{\sum\limits_{i=1}^{n}\dfrac{m_i\overline{Y}_i}{m_i+\alpha-1}}{\sum\limits_{i=1}^{n}\dfrac{m_i}{m_i+\alpha-1}}. \tag{2.17}$$

因此, 由方程$1/\theta_i=Z_i\overline{Y}_i+(1-Z_i)\widehat{\mu}$又得到$\theta_i$的一个估计:

$$\widehat{\theta}_i^{\,\text{hom}}=\frac{(m_i+\alpha-1)\sum\limits_{s=1}^{n}\dfrac{m_s}{m_s+\alpha-1}}{m_i\Big(\sum\limits_{s=1}^{n}\dfrac{m_s}{m_s+\alpha-1}\Big)\overline{Y}_i+(\alpha-1)\sum\limits_{s=1}^{n}\dfrac{m_s\overline{Y}_s}{m_s\alpha-1}}. \tag{2.18}$$

称之为齐次信度估计.

注记 2　在模型假设2.1-2.2中, 有

$$\mathsf{E}(X_{ij}|\theta_i)=\frac{\theta_i}{\theta_i-1}x_0,\qquad \mathsf{Var}\,(X_{ij}|\theta_i)=\frac{x_0^2\theta_i}{(\theta_i-1)^2(\theta_i-2)}.$$

如果对$\mathsf{E}(X_{ij}|\theta_i)$运用信度理论, 则也能得到某种形式下的信度估计. 但注意到此时$\mathsf{E}[\theta_i x_0/(\theta_i-1)]$, $\mathsf{Var}\,[(\theta_i/(\theta_i-1))x_0]$以及$\mathsf{E}[x_0\theta_i/((\theta_i-1)^2(\theta_i-2))]$在假设2.2下是不存在的, 因此利用这种方法求$\theta_i$的信度估计是没有意义的.

§3.　估计的比较

前面一节得到帕累托分布中风险参数θ_i的四个不同估计: 极大似然估计$\widehat{\theta}_i^{\,\text{mle}}$, 贝叶斯估计$\widehat{\theta}_i^{\,B}$, 非齐次信度估计$\widehat{\theta}_i^{\,c}$以及齐次信度估计$\widehat{\theta}_i^{\,\text{hom}}$. 因此有必要比较这些估计的好坏.

首先, 这四个估计显然都不是无偏估计. 下面我们来验证相合性.

命题 3.1　当$m_i\to\infty$时, 极大似然估计$\widehat{\theta}_i^{\,\text{mle}}$, 贝叶斯估计$\widehat{\theta}_i^{\,B}$, 非齐次信度估计$\widehat{\theta}_i^{\,c}$以及齐次信度估计$\widehat{\theta}_i^{\,\text{hom}}$都是风险参数$\theta_i$的强相合估计.

证明: 由于Y_{ij}, $j=1,2,\ldots$在θ_i给定条件下相互独立且同分布, 根据强大数定律, 有

$$\frac{1}{m_i}\sum_{j=1}^{m_i} Y_{ij}=\frac{1}{m_i}\sum_{j=1}^{m_i}\ln\Big(\frac{X_{ij}}{x_0}\Big)\to \mathsf{E}(Y_{ij}|\theta_i)=\frac{1}{\theta_i},\qquad \text{a.s.,}\tag{3.1}$$

根据几乎处处收敛的性质, 有$\widehat{\theta}_i^{\,\text{mle}}=1/\overline{Y}_i\to\theta_i$, a.s.以及

$$\widehat{\theta}_i^{\,B}=\frac{m_i+\alpha}{\sum\limits_{j=1}^{m_i}\ln\Big(\dfrac{X_{ij}}{x_0}\Big)+\beta}=\frac{1+\dfrac{\alpha}{m_i}}{\dfrac{1}{m_i}\sum\limits_{j=1}^{m_i}Y_{ij}+\dfrac{\beta}{m_i}}\to\frac{1+0}{\dfrac{1}{\theta_i}+0}=\theta_i,\qquad \text{a.s..}\tag{3.2}$$

另一方面, 当$m_i\to\infty$时, 显然$Z_i\to1$. 再次运用强大数定律, 有

$$\widehat{\theta}_i^{\,c}=\frac{1}{Z_i\overline{Y}_i+(1-Z_i)\mu}\to\frac{1}{\dfrac{1}{\theta_i}+0}=\theta_i,\qquad \text{a.s.}\tag{3.3}$$

以及

$$\widehat{\theta}_i^{\,\text{hom}}=\frac{1}{Z_i\overline{Y}_i+(1-Z_i)\widehat{\mu}}\to\frac{1}{\dfrac{1}{\theta_i}+0}=\theta_i,\qquad \text{a.s.,}$$

则证明了命题. □

因此, 从统计大样本意义上说这四个估计都是较好的估计. 但是, 在保险实际中, 一般样本容量m_i并不足够大, 这时为了比较这四个估计的差别, 我们采用均方误差作为衡量的标准. 为了计算的简便, 我们下面假设$m_1=m_2=\cdots=m_n=m$, 这时信度因子(2.13)为

$$Z_i=\frac{m}{m+\alpha-1}\overset{\text{denoted by}}{=}Z.$$

注意到, 这些估计的均方误差都没有显示表达式. 但是, 对给定的某个函数g, 在一定条件下, 根据泰勒公式, 近似的有

$$\mathsf{E}[(g(\widehat{\theta})-g(\theta))^2]\approx[g'(\theta)]^2\mathsf{E}[(\widehat{\theta}-\theta)^2].\tag{3.4}$$

因此我们可以近似的比较$g(\widehat{\theta}_i)$的均方误差, 来反映$\widehat{\theta}_i$的均方误差的大小.

取$g(x)=1/x$, 我们来分别计算$g(\widehat{\theta}_i^{\,\text{mle}})$, $g(\widehat{\theta}_i^{\,B})$, $g(\widehat{\theta}_i^{\,c})$以及$g(\widehat{\theta}_i^{\,\text{hom}})$的均方误差. 根据双重条件期望公式, 有

$$\text{MSE}_g(\text{mle})=\mathsf{E}\Big\{\mathsf{E}\Big[\Big(\frac{1}{m}\sum_{j=1}^{m}Y_{ij}-\frac{1}{\theta_i}\Big)^2\Big|\theta_i\Big]\Big\}=\mathsf{E}\Big(\frac{1}{m\theta_i^2}\Big)=\frac{\beta^2}{m(\alpha-1)(\alpha-2)}.$$

其次,

$$\text{MSE}_g(B)=\mathsf{E}[(g(\widehat{\theta}_i^{\,B})-g(\theta_i))^2]=\frac{(m+\alpha+2)\beta^2}{(m+\alpha)^2(\alpha-1)(\alpha-2)}.$$

另外, 根据信度理论非齐次信度估计与齐次信度估计的某些结果(参考Al-Saleh和Agarwal, 2007), 容易得到

$$\mathrm{MSE}_g(c) = \mathsf{E}[(g(\widehat{\theta}_i^{\,c}) - g(\theta_i))^2] = \frac{\tau^2\sigma^2}{m\tau^2+\sigma^2} = \frac{\beta^2}{(m+\alpha-1)(\alpha-1)(\alpha-2)}.$$

以及

$$\mathrm{MSE}_g(\mathrm{hom}) = \mathsf{E}[(g(\widehat{\theta}_i^{\,\mathrm{hom}}) - g(\theta_i))^2] = \frac{\beta^2(mn+\alpha-1)}{mn(m+\alpha-1)(\alpha-1)(\alpha-2)}.$$

令$M = \beta^2/[(\alpha-1)(\alpha-2)]$, 则有

$$\mathrm{MSE}_g(\mathrm{mle}) = \frac{M}{m}, \qquad \mathrm{MSE}_g(B) = \frac{(m+\alpha+2)M}{(m+\alpha)^2},$$

以及

$$\mathrm{MSE}_g(c) = \frac{M}{m+\alpha-1}, \qquad \mathrm{MSE}_g(\mathrm{hom}) = \frac{M(mn+\alpha-1)}{mn(m+\alpha-1)}.$$

因此, 得到下面的结论.

命题 3.2 在模型假设2.1-2.2成立时, 四个估计的均方误差有下面的大小关系:

(1) $\mathrm{MSE}_g(c) \le \mathrm{MSE}_g(B) \le \mathrm{MSE}_g(\mathrm{mle})$.

(2) $\mathrm{MSE}_g(c) \le \mathrm{MSE}_g(\mathrm{hom}) \le \mathrm{MSE}_g(\mathrm{mle})$.

(3) 均方误差$\mathrm{MSE}_g(B)$与$\mathrm{MSE}_g(\mathrm{hom})$的大小与$m, n, \alpha$的取值有关.

证明: 根据一些的数学计算, 容易验证(1)和(2). 对于(3), 令$k = (\alpha-1)/n$. 在保险实际中一般有$n > \alpha - 1$, 即$0 < k < 1$, 经过一些数学计算可得

$$\frac{1}{M}[\mathrm{MSE}_g(\mathrm{hom}) - \mathrm{MSE}_g(B)] = \frac{(1-k)m^2 + (\alpha-2-2\alpha k)m - k\alpha^2}{m(m+\alpha)^2(m+\alpha-1)}.$$

上式分子中看成m的一元二次函数, 其图像的开口向上, 则其判别式

$$\Delta = (\alpha-2-2\alpha k)^2 + (1-k)k\alpha^2 > 0.$$

设该一元二次方程的两个根分别为m_*与m^*, 则当$m_* < m < m^*$时候, $\mathrm{MSE}_g(B) - \mathrm{MSE}_g(\mathrm{hom}) > 0$, 反之, 若$m > m^*$或者$m < m_*$, 则$\mathrm{MSE}_g(B) - \mathrm{MSE}_g(\mathrm{hom}) < 0$. □

虽然命题3.2给出了$g(\theta_i) = 1/\theta_i$的几个估计的均方误差比较. 但如果需要参数θ_i的估计的均方误差的比较, 则没有显示的结果. 为了得到θ_i估计的均方误差的大小比较, 我们采用数值模拟的方法. 在模拟中, 取$\alpha = 4$, $\beta = 3$, $x_0 = 1$, $n = 10$. 且不失一般性, 设$m_1 = m_2 = \cdots = m_{10} = m$. 首先产生$\theta_i$的样本, 然后产生$X_{ij}$的样本, 并对各个估计计算均方误差. 在10000次重复模拟下计算这些估计的均方误差的平均值, 我们分别得到了θ_i和$g(\theta_i) = 1/\theta_i$的极大似然估计, 贝叶斯估计, 非齐次信度估计和齐次信度估计的均方误差的模拟结果, 如下表1和表2.

表1　当$m=8$时风险参数θ_i的均方误差

i	1	2	3	4	5	6	7	8	9	10
MS_1	0.5671	0.5010	0.5278	0.5394	0.5780	0.5404	0.5518	0.5257	0.5110	0.4849
MS_2	0.1715	0.1719	0.1732	0.1772	0.1756	0.1683	0.1697	0.1682	0.1674	0.1724
MS_3	0.1879	0.1858	0.1874	0.1919	0.1914	0.1837	0.1831	0.1813	0.1809	0.1874
MS_4	0.1991	0.1944	0.1975	0.2049	0.2042	0.1947	0.1957	0.1937	0.1937	0.1985

表2　当$m=8$时参数$g(\theta_i)=1/\theta_i$的均方误差

i	1	2	3	4	5	6	7	8	9	10
gMS_1	0.1875	0.1917	0.1665	0.1855	0.1911	0.1794	0.1985	0.1784	0.1942	0.1627
gMS_2	0.1410	0.1566	0.1341	0.1406	0.1495	0.1474	0.1524	0.1341	0.1332	0.1261
gMS_3	0.1328	0.1450	0.1240	0.1323	0.1400	0.1373	0.1422	0.1261	0.1274	0.1174
gMS_4	0.1377	0.1498	0.1281	0.1385	0.1442	0.1415	0.1478	0.1315	0.1328	0.1215

在上面的表1和表2中, MS_i, $i=1,2,3,4$和gMS_i, $i=1,2,3,4$分别表示θ_i与$g(\theta_i)=1/\theta_i$的极大似然估计, 贝叶斯估计, 非齐次信度估计和齐次信度估计的均方误差. 从模拟结果可以看出, 在θ_i的估计中, 均方误差的排序为$\mathrm{MS}_2<\mathrm{MS}_3<\mathrm{MS}_4<\mathrm{MS}_1$, 即此时最好的估计为贝叶斯估计. 显然, 由于$\theta_i$的贝叶斯估计$\widehat{\theta}_i^B$是所有样本的可测函数中使均方误差达到最小的估计. 而表2中$g(\theta_i)=1/\theta_i$的估计中均方误差的排序为$\mathrm{gMS}_3<\mathrm{gMS}_4<\mathrm{gMS}_2<\mathrm{gMS}_1$, 即达到最小的估计为信度估计$g(\widehat{\theta}_i^c)$, 此时主要原因是对$\theta_i$作函数变换以后就失去了其均方误差最小的性质. 表2的结果与我们证明的命题3.2是一致的. 然而, 这两个表的模拟结果显示了一个比较意外的统计现象, 就是式(3.4)只是近似的在大多数情况下成立, 而在某些情况下则未必成立. 表1和表2中贝叶斯估计和信度估计的均方误差比较就是这种另外情况. 因为在我们这个模型中, 经过倒数变换以后, 均方误差的最小性就发生了变化. 为了验证m对均方误差最优性的影响, 取$m=2$以及$m=80$, 重新做模拟得到下面的表:

表3　当$m=2$时参数$g(\theta_i)=1/\theta_i$的均方误差

i	1	2	3	4	5	6	7	8	9	10
gMS_1	0.7435	0.6726	0.7561	0.7824	0.7545	0.7323	0.7206	0.7615	0.7472	0.7639
gMS_2	0.3493	0.2714	0.3445	0.3050	0.3047	0.3328	0.3099	0.3260	0.3850	0.3171
gMS_3	0.3152	0.2448	0.3073	0.2781	0.2761	0.3007	0.2804	0.2929	0.3404	0.2892
gMS_4	0.3633	0.2968	0.3494	0.3326	0.3259	0.3485	0.3321	0.3409	0.3746	0.3371

表4　当$m=80$时参数$g(\theta_i)=1/\theta_i$的均方误差

i	1	2	3	4	5	6	7	8	9	10
gMS_1	0.0193	0.0192	0.0194	0.0202	0.0187	0.0171	0.0173	0.0200	0.0185	0.0184
gMS_2	0.0187	0.0186	0.0200	0.0195	0.0189	0.0167	0.0170	0.0213	0.0179	0.0179
gMS_3	0.0185	0.0184	0.0194	0.0193	0.0185	0.0165	0.0168	0.0206	0.0177	0.0177
gMS_4	0.0186	0.0185	0.0194	0.0194	0.0185	0.0166	0.0169	0.0205	0.0178	0.0178

从上面的表3可以看出, 当$m=2$时, 有$\mathrm{gMS}_3<\mathrm{gMS}_2<\mathrm{gMS}_4<\mathrm{gMS}_1$. 因此再次验证了命题3.2, 即当$m$很小时, $\mathrm{gMS}_2<\mathrm{gMS}_4$. 另外, 在表4中当$m=80$时四个均方误差几乎相同, 即随着$m$的增大这种均方误差的最优性将逐渐消失.

§4. 经验贝叶斯估计及渐近最优性

在上面一节中, 我们给出了风险参数θ_i的四个估计, 并用数值模拟的方法比较了各个估计的均方误差, 结论表明, 贝叶斯估计是四个估计中最好的一个, 其次是非齐次信度估计和齐次信度估计. 然而, 在实际运用中, 不管是贝叶斯估计还是信度估计, 都包含了结构参数α和β. 这两个结构参数在贝叶斯统计中也称为超参数. 在贝叶斯统计中, 确定α和β的估计有多种方法, 常用的方法有主观概率的分位点法, 极大后验分布法, 矩估计法等. 可参考Lehmann和Casella (2003), Gelman等(1995)等. 显然, 主观概率法需要根据专家观点确定分位数, 带有一定的主观性, 而在本模型中, 风险参数的后验分布为Gamma分布, 利用极大后验分布的方法也无法得到结构参数估计的显示表达式. 下面我们采用矩方法来结构参数估计α和β. 矩估计方法是经验贝叶斯统计中的重要方法, 特别是在保险精算中被广泛使用, 相关的研究可参考Bühlmann和Gisler (2005).

本节同样假设$m_i=m$, $i=1,2,\ldots,n$, 且$Y_{ij}=\ln(X_{ij}/x_0)$, 但很容易推广到更一般的情况.

由于$(\overline{Y}_i,\theta_i)$相互独立, 可以看成是该保单风险的一个样本. 其样本均值和样本方差分别为

$$\widehat{\mu_Y}=\frac{1}{n}\sum_{i=1}^{n}\overline{Y}_i,\qquad \widehat{\sigma_Y^2}=\frac{1}{n-1}\sum_{i=1}^{n}(\overline{Y}_i-\widehat{\mu_Y})^2. \tag{4.1}$$

根据双重期望公式, 容易得到

$$\mathsf{E}(\overline{Y}_i)=\frac{\beta}{\alpha-1},\qquad \mathsf{Var}\,(\overline{Y}_i)=\frac{\beta^2(m+\alpha-1)}{m(\alpha-1)^2(\alpha-2)}.$$

若令

$$\begin{cases}\widehat{\mu_Y}=\dfrac{\beta}{\alpha-1},\\[2mm] \widehat{\sigma_Y^2}=\dfrac{\beta^2(m+\alpha-1)}{m(\alpha-1)^2(\alpha-2)},\end{cases} \tag{4.2}$$

则可解得

$$\widehat{\alpha}=\frac{(m-1)\widehat{\mu_Y}^2+2m\widehat{\sigma_Y^2}}{m\widehat{\sigma_Y^2}-\widehat{\mu_Y}^2},\qquad \widehat{\beta}=\frac{\widehat{\mu_Y}(m\widehat{\mu_Y}^2+m\widehat{\sigma_Y^2})}{m\widehat{\sigma_Y^2}-\widehat{\mu_Y}^2}, \tag{4.3}$$

即得到α和β的矩估计.

根据强大数定律, 当m固定且$n\to\infty$时, 可以验证估计$\widehat{\alpha}$与$\widehat{\beta}$均满足一致强相合性, 即有$\widehat{\alpha}\to\alpha$, a.s.以及$\widehat{\beta}\to\beta$, a.s.. 且容易证明这两个估计是平方收敛的, 即$\mathsf{E}[(\widehat{\alpha}-\alpha)^2]\to 0$, 以及$\mathsf{E}[(\widehat{\beta}-\beta)^2]\to 0$.

将$\widehat{\alpha}$与$\widehat{\beta}$代入上节的贝叶斯估计与非齐次信度估计, 得到

$$\widetilde{\theta}_i^{\,B}=\frac{m+\widehat{\alpha}}{\sum\limits_{j=1}^{m}Y_{ij}+\widehat{\beta}},\qquad \widetilde{\theta}_i^{\,c}=\frac{m+\widehat{\alpha}-1}{\sum\limits_{j=1}^{m}Y_{ij}+\widehat{\beta}}. \tag{4.4}$$

注意到估计$\widetilde{\theta}_i^{\,B}$与$\widetilde{\theta}_i^{\,c}$再也不依赖于任何未知参数, 可以在实际中直接使用. 在贝叶斯统计学中, 称这种方法为经验贝叶斯方法, 而$\widetilde{\theta}_i^{\,B}$与$\widetilde{\theta}_i^{\,c}$称为$\theta_i$的经验贝叶斯估计和经验贝叶斯信度估计.

对于经验贝叶斯估计和经验贝叶斯信度估计, 我们最关心的是将结构参数的估计代入后其均方损失是否有较大影响? 这种影响随着样本容量n的增大是否能一致趋于零? 这就是所谓的经验贝叶斯最优性. 关于经验贝叶斯估计的文献可参考Robbins (1955, 1964), Mashayekhi (2002)等.

定义 4.1 设$\widetilde{\theta}$和$\widehat{\theta}$分别是参数θ的经验贝叶斯估计和贝叶斯估计, 若满足

$$\lim_{n\to\infty}|\mathsf{E}[(\widetilde{\theta}-\theta)^2]-\mathsf{E}[(\widehat{\theta}-\theta)^2]|=0, \tag{4.5}$$

则称经验贝叶斯估计$\widetilde{\theta}$是渐近最优的.

在本文中θ_i的经验贝叶斯估计$\widetilde{\theta}_i^{\,B}$和经验贝叶斯信度估计$\widetilde{\theta}_i^{\,c}$, 可以得到下面的结论.

定理 4.1 若存在$\delta>0$, 使得$\widehat{\theta}_i^{\,\mathrm{mle}}<\delta$, 则当$n\to\infty$时, 经验贝叶斯估计$\widetilde{\theta}_i^{\,B}$和经验贝叶斯信度估计$\widetilde{\theta}_i^{\,c}$都是渐近最优估计, 即有

$$\lim_{n\to\infty}\max_{1\le i\le m}|\mathsf{E}[(\widetilde{\theta}_i^{\,B}-\theta_i)^2]-\mathsf{E}[(\widehat{\theta}_i^{\,B}-\theta_i)^2]|=0 \tag{4.6}$$

以及

$$\lim_{n\to\infty}\max_{1\le i\le m}|\mathsf{E}[(\widetilde{\theta}_i^{\,c}-\theta_i)^2]-\mathsf{E}[(\widehat{\theta}_i^{\,c}-\theta_i)^2]|=0. \tag{4.7}$$

证明: 由于$\widetilde{\theta}_i^{\,B}$与$\widetilde{\theta}_i^{\,c}$的渐近最优性证明基本相同, 我们只证明(4.6)式. 记

$$Y_{i\cdot}=\sum_{j=1}^{m}\ln\Big(\frac{X_{ij}}{x_0}\Big).$$

首先, 根据平方和的分解并结合期望不等式有

$$\begin{aligned}|\mathsf{E}[(\widetilde{\theta}_i^{\,B}-\theta_i)^2]-\mathsf{E}[(\widehat{\theta}_i^{\,B}-\theta_i)^2]|&=|\mathsf{E}[(\widetilde{\theta}_i^{\,B}-\widehat{\theta}_i^{\,B})^2]+2\mathsf{E}[(\widetilde{\theta}_i^{\,B}-\widehat{\theta}_i^{\,B})(\widehat{\theta}_i^{\,B}-\theta_i)]|\\&\le\mathsf{E}[(\widetilde{\theta}_i^{\,B}-\widehat{\theta}_i^{\,B})^2]+2\sqrt{\mathsf{E}[(\widetilde{\theta}_i^{\,B}-\widehat{\theta}_i^{\,B})^2]\mathsf{E}[(\widehat{\theta}_i^{\,B}-\theta_i)^2]}.\end{aligned}$$

又因为

$$\mathsf{E}[(\widehat{\theta}_i^B-\theta_i)^2]=\mathsf{E}\Big[\Big(\frac{m+\alpha-\theta_i(Y_{i\cdot}+\beta)}{Y_{i\cdot}+\beta}\Big)^2\Big]\leq\frac{1}{\beta^2}\mathsf{E}[(m+\alpha-\theta_i(Y_{i\cdot}+\beta))^2]=\frac{m+\alpha}{\beta^2}\leq\infty.$$

因此只需证明 $\lim\limits_{n\to\infty}\max\limits_{1\leq i\leq m}\mathsf{E}[(\widetilde{\theta}_i^B-\widehat{\theta}_i^B)^2]=0$即可. 由于$\widehat{\theta}_i^{\rm mle}=m/Y_{i\cdot}$, 根据定理的条件有$\widehat{\theta}_i^{\rm mle}<\delta$, 即$Y_{i\cdot}>m/\delta$. 因此可以得到

$$\begin{aligned}\mathsf{E}[(\widetilde{\theta}_i^B-\widehat{\theta}_i^B)^2]&=\mathsf{E}\Big[\Big(\frac{m+\widehat{\alpha}}{Y_{i\cdot}+\widehat{\beta}}-\frac{m+\alpha}{Y_{i\cdot}+\beta}\Big)^2\Big]\\&=\mathsf{E}\Big[\frac{(\widehat{\alpha}-\alpha)^2}{(Y_{i\cdot}+\widehat{\beta})^2}\Big]+\mathsf{E}\Big[\Big(\frac{(m+\alpha)(\widehat{\beta}-\beta)}{(Y_{i\cdot}+\widehat{\beta})(Y_{i\cdot}+\beta)}\Big)^2\Big]\\&\leq\frac{\delta^4}{m^4}\mathsf{E}[(\widehat{\alpha}-\alpha)^2]+\frac{\delta^2(m+\alpha)^2}{m^2\beta^2}\mathsf{E}[(\widehat{\beta}-\beta)^2]\\&\to 0.\end{aligned}$$

因此(4.6)式成立, 同理可证(4.7)式. 则完成了定理的证明. □

为了验证在代入结构参数估计后对估计的均方误差的影响, 即研究贝叶斯估计、信度估计以及相应的经验贝叶斯估计、经验贝叶斯信度估计的均方误差情况, 我们在第3节相同的环境下进行数值模拟. 取$\alpha=4$, $\beta=3$, $x_0=1$, $n=10$, $m=20$. 但在经验贝叶斯估计和经验贝叶斯信度估计中采用$\widehat{\alpha},\widehat{\beta}$, 得到下面的结果.

表5 当$m=8$时贝叶斯估计及经验贝叶斯估计的均方误差

i	1	2	3	4	5	6	7	8	9	10
MSB	0.1727	0.1680	0.1689	0.1725	0.1703	0.1712	0.1721	0.1701	0.1709	0.1729
EMSB	0.2522	0.2098	0.2121	0.2175	0.2151	0.2088	0.2188	0.3723	0.2092	0.2163
MSC	0.1886	0.1820	0.1851	0.1883	0.1872	0.1825	0.1854	0.1848	0.1839	0.1874
EMSC	0.2715	0.2243	0.2266	0.2320	0.2308	0.2212	0.2329	0.3823	0.2232	0.2304

在上面的表5中, MSB和EMSB分别表示贝叶斯估计和对应的经验贝叶斯估计的均方误差, 而MSC和EMSC分别表示非齐次信度估计和对应的经验贝叶斯信度估计的均方误差. 显然, 由于结构参数的估计导致的均方误差的增大. 这两个经验贝叶斯估计仍然比极大似然估计的均方误差更小.

§5. 结论注记

本文考虑了具有免赔额保单风险的累托分布分布模型, 注意到在本文的讨论中, 根据保险的实际情况, 免赔额x_0假定是已知的, 但从讨论的过程中可以看出, 风险参数的估计与x_0是不相关的, 即x_0的值并不影响风险参数θ_i的估计. 文章在Gamma先验分布假设下, 讨

论了风险参数的极大似然估计, 贝叶斯估计, 非齐次信度估计以及齐次信度估计. 并讨论了这些估计的统计性质, 比较了这些估计的均方误差. 根据贝叶斯估计的定义, 贝叶斯估计是在所有可测函数中均方误差达到最小的估计, 这个结论在数值模拟的表1中得到了验证. 但有些意外的是, 对这些估计进行倒数变换以后, 贝叶斯估计的均方误差最优性就不存在了, 这时通过理论证明和数值模拟均表明非齐次信度估计的均方误差最小. 文章不仅证明了这四个估计的强相合性, 而且给出了结构参数的矩估计, 并得到贝叶斯估计和信度估计的经验贝叶斯估计, 最后对证明了经验贝叶斯估计的渐近最优性, 并对其均方误差进行了比较, 结论显示, 经验贝叶斯估计的均方误差仍然比极大似然估计的均方误差更小, 且可以在实际中直接使用.

参 考 文 献

[1] He, H., Zhou, N. and Zhang, R.M., On estimation for the Pareto distribution, *Statistical Methodology*, **21**(2014), 49–58.

[2] Tudor, C.A., Chaos expansion and asymptotic behavior of the Pareto distribution, *Statistics and Probability Letters*, **91**(2014), 62–68.

[3] Fahidy, T.Z., Applying Pareto distribution theory to electrolytic powder production, *Electrochemistry Communications*, **13(3)**(2011), 262–264.

[4] Dixit, U.J. and Nooghabi, M.J., Efficient estimation in the Pareto distribution with the presence of outliers, *Statistical Methodology*, **8(4)**(2011), 340–355.

[5] Harris, C.M., The Pareto distribution as a queue service discipline, *Operations Research*, **16(2)**(1968), 307–313.

[6] Arnold, B.C., *Pareto Distributions*, International Co-operative Publishing House, 1983.

[7] Steindl, J., *Random Processes and the Growth of Firms*, New York: Hafner Publishing Company, 1965.

[8] Hagstroem, K.G., Remarks on Pareto distributions, *Scandinavian Actuarial Journal*, **1960(1-2)** (1960), 59–71.

[9] Ramsay, C.M., A solution to the ruin problem for Pareto distributions, *Insurance: Mathematics and Economics*, **33(1)**(2003), 109–116.

[10] Albrecher, H. and Kortschak, D., On ruin probability and aggregate claim representations for Pareto claim size distributions, *Insurance: Mathematics and Economics*, **45(3)**(2009), 362–373.

[11] Brazauskas, V. and Kleefeld, A., Robust and efficient fitting of the generalized Pareto distribution with actuarial applications in view, *Insurance: Mathematics and Economics*, **45(3)**(2009), 424–435.

[12] Gómez-Déniz, E., Pérez-Sánchez, J.M. and Vázquez-Polo, F.J., On the use of posterior regret Γ-minimax actions to obtain credibility premiums, *Insurance: Mathematics and Economics*, **39(1)** (2006), 115–121.

[13] Al-Saleh, J.A. and Agarwal, S.K., Finite mixture of gamma distributions: A conjugate prior, *Computational Statistics and Data Analysis*, **51(9)**(2007), 4369–4378.

[14] Griffin, J.E. and Brown, P.J., Inference with normal-gamma prior distributions in regression problems, *Bayesian Analysis*, **5(1)**(2010), 171–188.

[15] Meczarski, M. and Zieliński, R., Stability of the Bayesian estimator of the Poisson mean under the inexactly specified gamma prior, *Statistics and Probability Letters*, **12(4)**(1991), 329–333.

[16] 茆诗松, 王静龙, 濮晓龙, 高等数理统计, 北京: 高等教育出版社, 1998.

[17] Lehmann, E.L. and Casella, G., *Theory of Point Estimation*, Springer-Verlag, 2003.

[18] Walker, S.G., Modern Bayesian asymptotics, *Statistical Science*, **19(1)**(2004), 111–117.

[19] Gelman, A., Carlin, J.B., Stern, H.S. and Bubin, D.B., *Bayesian Data Analysis*, New York: Chapman-Hall, 1995.

[20] Bühlmann, H. and Gisler, A., *A Course in Credibility Theory and its Applications*, Springer-Verlag, 2005.

[21] 郑丹, 章溢, 温利民, 具有时间变化效应的信度模型, 江西师范大学学报(自然科学版), **36(3)**(2012), 249–252.

[22] 方婧, 章溢, 温利民, 聚合风险模型下的信度估计, 江西师范大学学报(自然科学版), **36(6)**(2012), 607–611.

[23] Robbins, H., An empirical Bayes approach to statistics, In: *Proceedings of the Third Berkeley Symposium on Mathematics, Statistics and Probability, Vol. 1*, 1955, 157–164.

[24] Robbins, H., The empirical Bayes approach to statistical decision problems, *The Annals of Mathematical Statistics*, **35(1)**(1964), 1–20.

[25] Mashayekhi, M., On asymptotic optimality in empirical Bayes credibility, *Insurance: Mathematics and Economics*, **31(2)**(2002), 285–295.

The Empirical Bayes Estimation of Risk Parameters in Pareto Claim Distribution

Wen Limin[1,2]　Zhang Mei[1]　Cheng Zihong[1]　Zhang Yi[1,3]

([1] *School of Mathematics and Information Science, Jiangxi Normal University, Nanchang, 330022*)

([2] *School of Information Management, Jiangxi University of Finance and Economics, Nanchang, 330013*)

([3] *School of Computer and Information Engineering, Jiangxi Normal University, Nanchang, 330022*)

The Bayesian model is established in this paper, and the risk parameters of claim amounts in Pareto distribution are estimated. The maximum likelihood estimation, Bayesian estimation and credibility estimation are derived and the strong consistency of these estimates are proved. We also compared their mean square error both in theory and in numerical simulation. The results show that Bayesian estimation is better than other estimates in sense of mean square error. Finally, the structural parameters in Bayes estimation and credibility estimation are estimated and the corresponding empirical Bayes estimates are proved asymptotically optimal.

Keywords: Pareto distribution, empirical Bayes estimators, asymptotical optimality, credibility estimator.

AMS Subject Classification: 62G35.