第 44 卷 第 3 期 江西师范大学学报(自然科学版) Vol. 44 No. 3
2020 年 5 月 Journal of Jiangxi Normal University(Natural Science) May 2020

文章编号: 1000-5862(2020)03-0269-06

零膨胀泊松模型中风险参数的贝叶斯估计

张良超,周金亮,温利民*
(江西师范大学数学与信息科学学院,江西 南昌 330022)

摘要: 该文建立了贝叶斯模型,讨论了零膨胀泊松分布中参数的估计问题. 在平方损失函数、Linex 损失函数和 Stein 损失函数下得到了风险参数的贝叶斯估计. 进而,引进了信度理论,在平方损失函数下得到了风险参数的信度估计,证明了估计的相合性. 最后,通过数值模拟的方法对估计的收敛性进行了比较.
关键词: 零膨胀泊松模型; 贝叶斯估计; 参数估计
中图分类号: O 211.9 文献标志码: A DOI: 10.16357/j.cnki.issn1000-5862.2020.03.10

0 引言

零膨胀泊松分布(ZIP)是一种取值为非负的离散型混合分布,D. Lambert[1] 提出了零膨胀泊松模型,并对焊接瑕疵数据进行了拟合. 随着社会经济等各方面的发展,大量的计数数据广泛地存在于农业、医学、环境、经济学和保险精算等领域[2-5].

在非寿险精算中,对于索赔次数的拟合,泊松分布是一个较好的选择. 但实际的索赔次数数据往往具有零膨胀特点,即索赔次数在零点处存在较大的概率堆积. 此时若继续用泊松分布来拟合数据,则会导致推断错误. 零膨胀现象产生的原因很多,如保单持有者增强了风险防范意识,减少了索赔发生的概率; 保险公司在设计的保单中含有免赔额或无赔款折扣等条款,导致被保险人为了得到保费折扣对小的事故不报告给保险公司等. 为了拟合在数据中过多的零值,零膨胀模型是一种有效的方法,如 K. K. W. Yau 等[6] 在 D. Lambert[1] 的基础上提出了零膨胀负二项分布模型,M. Denuit 等[7] 和 R. Winkelmann[8] 分别对零膨胀现象进行了讨论和分析.

对模型参数的估计是统计学的重要课题之一,传统的估计方法有矩估计、极大似然估计与最小二乘估计等. 当样本较小时,传统的估计方法得到的估计误差都较大. 而基于贝叶斯框架下的后验估计也是参数估计的一种重要方法,该方法可以利用参数的先验信息,有效降低估计的误差. 文献[9]讨论了在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,结果表明贝叶斯估计比其他估计具有较小的均方误差. 文献[10]讨论了零膨胀对数级数分布中参数的估计问题,给出了参数的矩估计、极大似然估计和贝叶斯估计,结果表明在样本容量较小时贝叶斯估计的优势更明显.

在统计决策问题中,最重要的就是损失函数,它把决策行动与后果联系起来. 在非寿险精算中,当保险公司对未来的索赔次数进行估计时可能出现 2 种偏差. 对索赔次数估计过高,进而征收保费过高,投保人数量减少; 对索赔次数估计过低,进而征收保费不够,造成亏损. 平方损失函数使得过高和过低估计具有相等的惩罚程度. 然而在实际保险运用中,过低的征收保费可能导致公司偿付能力不足而破产,因此低保费应该比高保费有更大的惩罚,此时对应的损失函数应取为非对称损失函数. H. R. Varian 于 1975 年提出的 Linex 损失就是一种非对称损失函数. 文献[11]研究了 Linex 损失函数下的信度估计问题; 文献[12]考虑了 Linex 损失函数下的估计和预测问题,研究了该损失函数的性质和应用. W. James等[13] 引进了一种非对称凸损失函数,称为 Stein 损失函数. 而贝叶斯估计依赖于在参数给定下的样本分布和参数的先验分布. 这些分布的具体形式,特别是参数的先验分布在实际中是未知的. 针对这个问题,可利用 H. Bühlmann[14] 的理论,将参数的估计限定在样本的线性函数中,通过求解期望平方损失,得到参数的信度估计.

本文在先验分布为 Gamma 分布下研究零膨胀

收稿日期: 2019-12-27
基金项目: 国家自然科学基金(71761019)和江西省教育厅研究生创新基金(YC2019-S125)资助项目.
通信作者: 温利民(1979-),男,江西石城人,教授,博士,博士生导师,主要从事精算学的研究. E-mail: wlmjxnu@163.com

泊松模型中风险参数的贝叶斯估计,将借助信度理论提出零膨胀泊松模型中参数的信度估计. 在 3 种损失函数下讨论风险参数的贝叶斯估计,并通过数值模拟验证估计的相合性,在均方标准差的意义下比较贝叶斯估计的优良性.

1 零膨胀泊松模型的贝叶斯后验期望估计

设有 n 个相同类型保单,第 i 份保单有 T_i 年的索赔记录. 记 N_{ij} 为第 i 种保单在第 j 年的索赔次数,则 $i=1,2,\cdots,n$,$j=1,2,\cdots,T_i$. 若假设 N_{ij} 是来自零膨胀泊松分布的样本,其概率分布为

$$P(N_{ij}=n_{ij}\mid\theta_i,\omega)=I_{\{0\}}(n_{ij})\omega+(1-\omega)\mathrm{e}^{-\theta_i}\theta_i^{n_{ij}}/n_{ij}!,\tag{1}$$

其中 $\omega\in[0,1]$ 为零膨胀概率,而 $\theta_i\in(0,+\infty)$ 为第 i 个保单的风险参数;当 $x\in A$ 时,示性函数 $I_A(x)=1$,当 $x\notin A$ 时,$I_A(x)=0$. 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.

零膨胀泊松分布是拟合计数型数据的一种合理分布,常常用来刻画具有零膨胀特点的保单索赔次数分布. 该分布有 2 个参数:泊松分布的均值参数 θ_i 和零膨胀参数 ω,在零膨胀泊松模型中,通常 ω 取为固定值. 由(1)式可知,索赔次数取值为 0 的概率关于 ω 单调递增. 当 $\omega=0$ 时,退化为泊松分布. 而 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 年的索赔次数服从零膨胀泊松分布. 由于风险的非齐次性,不同保单的风险参数是不相同的,因此 θ_i 常常假设为随机变量,具有某个先验分布. 故对 θ_i 的估计就落入了贝叶斯框架.

假设 1 设 N_{ij} 为第 i 种保单在第 j 年的索赔次数,其中 $j=1,2,\cdots,T_i$. 且在第 i 个风险参数 θ_i 给定的条件下,N_{ij} 是来自零膨胀泊松分布的样本,具有概率分布(1),其中 θ_i 为该保单的风险参数,ω 为零膨胀参数.

假设 2 设 θ_i 为相互独立且服从共同的 Gamma 分布,先验密度函数为

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

其中 $\alpha>0$、$\beta>0$ 为结构参数.

为书写方便,记

$$T_0=\sum_{j=1}^{T_i}I_{(n_{ij}=0)},M_i=\prod_{j=1}^{T_i}n_{ij}!,n_{i\cdot}=\sum_{j=1}^{T_i}n_{ij},N=(N_{i1},N_{i2},\cdots,N_{iT_i}),n=(n_{i1},n_{i2},\cdots,n_{iT_i}).$$

定理 1 若风险满足假设 1 和假设 2,则 $N=(N_{i1},N_{i2},\cdots,N_{iT_i})$ 的边际分布为

$$P(N=n)=\sum_{k=0}^{T_0}C_{T_0}^k\omega^{T_0-k}(1-\omega)^{T_i-T_0+k}\Gamma(n_{i\cdot}+\alpha)\beta^\alpha/(\Gamma(\alpha)M_i(T_i-T_0+k+\beta)^{n_{i\cdot}+\alpha}).\tag{2}$$

证 由条件期望公式知 $P(N=n)=\int_0^{+\infty}P(N=n\mid\theta_i)\pi(\theta_i)\mathrm{d}\theta_i$,根据假设 1 得

$$P(N=n\mid\theta_i)=\prod_{j=1}^{T_i}(I_{\{0\}}(n_{ij})\omega+(1-\omega)\mathrm{e}^{-\theta_i}\theta_i^{n_{ij}}/n_{ij}!)=\sum_{k=0}^{T_0}C_{T_0}^k\omega^{T_0-k}(1-\omega)^{T_i-T_0+k}\mathrm{e}^{-\theta_i(T_i-T_0+k)}\theta_i^{n_{i\cdot}}/M_i,$$

所以

$$P(N=n)=\sum_{k=0}^{T_0}C_{T_0}^k\omega^{T_0-k}(1-\omega)^{T_i-T_0+k}\cdot\int_0^{+\infty}\frac{\beta^\alpha}{M_i\Gamma(\alpha)}\theta_i^{n_{i\cdot}+\alpha-1}\mathrm{e}^{-\theta_i(T_i-T_0+k)}\mathrm{d}\theta_i.$$

整理上式即为所求(2)式.

定理 2 若风险满足假设 1 和假设 2,则参数 θ_i 的后验期望估计为

$$E(\theta_i\mid N=n)=(n_{i\cdot}+\alpha)(\sum_{k=0}^{T_0}C_{T_0}^k\omega^{-k}(1-\omega)^k(T_i-T_0+k+\beta)^{-(n_{i\cdot}+\alpha+1)})/(\sum_{k=0}^{T_0}C_{T_0}^k\omega^{-k}(1-\omega)^k(T_i-T_0+k+\beta)^{-(n_{i\cdot}+\alpha)}).\tag{3}$$

证 由贝叶斯定理 $\pi_*(\theta_i\mid N=n)=\pi(\theta_i)P(N=n\mid\theta_i)/(P(N=n))$ 可知,风险参数 θ_i 的后验密度为

$$\pi_*(\theta_i\mid N=n)=(\sum_{k=0}^{T_0}C_{T_0}^k\omega^{-k}(1-\omega)^k\exp(-\theta_i(T_i-T_0+k+\beta))\theta_i^{n_{i\cdot}+\alpha-1})/(\sum_{k=0}^{T_0}C_{T_0}^k\omega^{-k}(1-\omega)^k\Gamma(n_{i\cdot}+\alpha)(T_i-T_0+k+\beta)^{-(n_{i\cdot}+\alpha)}),\tag{4}$$

因此参数 θ_i 的后验期望估计为

$$E(\theta_i\mid N=n)=\int_0^{+\infty}\theta_i\pi_*(\theta_i\mid N=n)\mathrm{d}\theta_i=(\sum_{k=0}^{T_0}C_{T_0}^k\omega^{-k}(1-\omega)^k\int_0^{+\infty}\exp(-\theta_i(T_i-T_0+k+\beta))\theta_i^{n_{i\cdot}+\alpha}\mathrm{d}\theta_i)/(\sum_{k=0}^{T_0}C_{T_0}^k\omega^{-k}(1-\omega)^k\Gamma(n_{i\cdot}+\alpha)(T_i-T_0+k+\beta)^{-(n_{i\cdot}+\alpha)}).$$

整理上式即为所求,定理 2 得证.

2　损失函数下风险参数的贝叶斯估计

本节在 Stein 损失、Linex 损失和平方损失下，给出当零膨胀泊松分布风险参数的先验分布为 Gamma 分布时的贝叶斯估计. 平方损失函数使得过高和过低估计具有相等的惩罚程度，Stein 损失函数使得过低估计比过高估计具有更大的惩罚，Linex 损失函数使得过高估计比过低估计具有更大的惩罚. 在保险实际中，保险公司可根据自身情况进行选择.

引理 1[15]　在平方损失函数 $L(\theta,\delta) = (\delta-\theta)^2$ 下，δ 为参数 θ 的估计，对任一先验分布 $\pi(\theta)$，θ 的贝叶斯估计为后验期望值.

在零膨胀泊松模型中，根据假设 1 和假设 2，容易得到如下定理.

定理 3　若风险满足假设 1 和假设 2，则在平方损失函数下，风险参数 θ_i 的贝叶斯估计为

$$\delta_{square} = (n_{i\cdot}+\alpha)\left(\sum_{k=0}^{T_0} C_{T_0}^k \omega^{-k}(1-\omega)^k (T_i - T_0 + k + \beta)^{-(n_{i\cdot}+\alpha+1)}\right) \Big/ \left(\sum_{k=0}^{T_0} C_{T_0}^k \omega^{-k}(1-\omega)^k (T_i - T_0 + k + \beta)^{-(n_{i\cdot}+\alpha)}\right)$$

定理 4　若风险满足假设 1 和假设 2，则在 Stein 损失函数 $L(\theta,\delta) = \delta/\theta - \ln(\delta/\theta) - 1$ 下，使期望损失达到最小，得到风险参数 θ_i 的贝叶斯估计为

$$\delta_{Stein} = (n_{i\cdot}+\alpha-1)\left(\sum_{k=0}^{T_0} C_{T_0}^k \omega^{-k}(1-\omega)^k (T_i - T_0 + k + \beta)^{-(n_{i\cdot}+\alpha)}\right) \Big/ \left(\sum_{k=0}^{T_0} C_{T_0}^k \omega^{-k}(1-\omega)^k (T_i - T_0 + k + \beta)^{-(n_{i\cdot}+\alpha-1)}\right).$$

证　为方便，记 $\Phi = E(\delta/\theta_i - \ln(\delta/\theta_i) - 1)$，$\delta$ 为 θ_i 的估计，由双重期望公式知，

$$\Phi = E(E(\delta/\theta_i - \ln(\delta/\theta_i) - 1 \mid N = n)),$$

要使 Φ 最小化，只需

$$Q = E(\delta/\theta_i - \ln(\delta/\theta_i) - 1 \mid N = n)$$

达到最小即可. 对 Q 关于 δ 求偏导，令导数等于 0，即

$$\partial Q/\partial\delta = E((1/\theta_i - 1/\delta) \mid N = n) = 0,$$

则有

$$\delta = 1/(E(\theta_i^{-1} \mid N = n)). \tag{5}$$

把(4) 式代入 $E(\theta_i^{-1} \mid N = n) = \int_0^{+\infty} \theta_i^{-1}\pi_*(\theta_i \mid N = n)\,d\theta_i$，整理(5) 式即为所求.

定理 5　若风险满足假设 1 和假设 2，则在 Linex 损失函数下，使期望损失达到最小，得到风险参数 θ_i 的贝叶斯估计为

$$\delta_{Linex} = \frac{1}{a}\ln\left(\left(\sum_{k=0}^{T_0} C_{T_0}^k \omega^{-k}(1-\omega)^k (T_i - T_0 + k + \beta)^{-(n_{i\cdot}+\alpha)}\right) \Big/ \left(\sum_{k=0}^{T_0} C_{T_0}^k \omega^{-k}(1-\omega)^k (T_i - T_0 + k + \beta + a)^{-(n_{i\cdot}+\alpha)}\right)\right),$$

其中 $L(\theta,\delta) = \exp(a(\delta-\theta)) - a(\delta-\theta) - 1$，这里 $a > 0$ 为已知的参数.

证　为方便，记 $\Phi = E(\exp(a(\delta-\theta_i)) - a(\delta-\theta_i) - 1)$，$\delta$ 为 θ_i 的估计，由双重期望公式知 $\Phi = E(E(\exp(a(\delta-\theta_i)) - a(\delta-\theta_i) - 1 \mid N = n))$，要使 Φ 最小化，只需

$$Q = E(\exp(a(\delta-\theta_i)) - a(\delta-\theta_i) - 1 \mid N = n)$$

达到最小即可，对 Q 关于 δ 求偏导，令导数等于 0，即

$$\partial Q/\partial\delta = E((a\exp(a(\delta-\theta_i)) - a) \mid N = n) = 0,$$

则有

$$\delta = -\frac{1}{a}\ln(E(\exp(-a\theta_i) \mid N = n)), \tag{6}$$

其中 $E(\exp(-a\theta_i) \mid N = n) = \int_0^{+\infty} \exp(-a\theta_i)\pi_*(\theta_i \mid N = n)\,d\theta_i$，把(4) 式代入上式，整理(6) 式即为所求.

注 1　当 $a \to 0^+$ 时，$|\delta-\theta|$ 有界，Linex 损失与 $a^2(\delta-\theta)^2$ 为同阶无穷小量. 事实上，

$$\lim_{a\to 0^+}(\exp(a(\delta-\theta)) - a(\delta-\theta) - 1)/(a^2(\delta-\theta)^2) = 1/2,$$

因此该 Linex 损失函数也趋于对称，与平方损失函数十分接近. 于是对(6) 式取极限，记 $X = (\theta_i \mid N = n)$ 为条件变量，由 L. Hospital 法则，有

$$\lim_{a\to 0^+}\delta_{Linex} = \lim_{a\to 0^+}(E(X\exp(-aX)))/(E(\exp(-aX))) = E(X).$$

因此在平方损失函数下的贝叶斯估计是在 Linex 损失函数下的一种特殊情况.

在 3 个损失函数下，得到了风险参数 θ_i 的 3 个贝叶斯估计 δ_{square}、δ_{Stein} 和 δ_{Linex}，因此有必要比较 3 个估计的大小关系. 为了后续证明，给出如下引理.

引理 2[16]　设 X 是任意随机变量，若 $f(x)$ 是凸(convex) 函数，则

$$E(f(X)) \geqslant f(E(X)).$$

定理 6　在一定正则条件下，若 3 个估计都存在，则有 $\delta_{square} \geqslant \max\{\delta_{Linex},\delta_{Stein}\}$.

证　先证明 $\delta_{square} \geqslant \delta_{Linex}$，等价于证明

$$E(\theta_i \mid N = n) \geqslant -\frac{1}{a}\ln(E(\exp(-a\theta_i) \mid N = n)),$$

或者

$$E(\exp(-a\theta_i)\mid N=n)\geqslant \exp(-aE(\theta_i\mid N=n)),\tag{7}$$

令 $X=(\theta_i\mid N=n)$ 为条件变量,记 $f(x)=\mathrm{e}^{-ax}$,则 $f(x)$ 是凸函数,由引理 2 知,

$$E(f(X))\geqslant f(E(X)),$$

因此(7)式成立.

接下来证明 $\delta_{square}\geqslant\delta_{Stein}$,等价于

$$E(\theta_i\mid N=n)\geqslant 1/(E(\theta_i^{-1}\mid N=n)),\tag{8}$$

令 $Y=(\theta_i^{-1}\mid N=n)$,$g(x)=1/x$,容易证明 $g(x)$ 在$(0,+\infty)$内是凸函数,由引理 2 知,

$$E(g(Y))\geqslant g(E(Y)),$$

因此(8)式成立. 综上所述,定理 6 得证.

3 风险参数的信度估计

从定理3~定理5的结果可知,为求贝叶斯估计必须知道先验分布的具体形式,然而在实际应用中,先验分布的具体分布形式是很难获得的. 即使先验分布已知,但在大多数情况下贝叶斯估计没有显性解. 针对这一缺点,H. Bühlmann(1967)[14] 将风险保费的估计限定在样本的线性函数中,得到的估计恰能表达为样本均值和聚合均值的加权平均,这一估计形式被称为信度估计,该方法被称为信度理论,在非寿险的保费厘定中有重要的应用[17].

在零膨胀泊松模型中,若先验分布不再具有假设 2 的 Gamma 分布形式,则无法得到风险参数的贝叶斯估计. 这时将采用信度理论的方法,让风险参数的估计直接限定在样本均值和聚合均值的加权和形式中,求解风险参数的最优估计. 为方便,引入如下记号:

$$X_{ij}=N_{ij}/(1-\omega),\ \bar{X}_i=\frac{1}{T_i}\sum_{j=1}^{T_i}X_{ij},\ E(\theta_i)=\mu,$$

$$\mathrm{Var}(\theta_i)=\tau^2,\ E(\mathrm{Var}(X_{ij}\mid\theta_i))=\sigma^2.$$

定理 7 在满足假设 1 的情况下,若将风险参数的估计限定在样本均值和聚合均值的加权和中,则求解最小化问题

$$\min_C E(\theta_i-(C\bar{X}_i+(1-C)\mu))^2,$$

得到 θ_i 的信度估计:

$$\delta_{cre}=Z\bar{X}_i+(1-Z)\mu,$$

其中信度因子 $Z=T_i\tau^2/(\sigma^2+T_i\tau^2)$.

证 为方便,记 $\Phi=E(\theta_i-(C\bar{X}_i+(1-C)\mu))^2$,对 Φ 关于 C 求偏导,令导数等于 0,即

$$E((\theta_i-(C\bar{X}_i+(1-C)\mu))(\bar{X}_i-\mu))=0,$$ 解得

$$C=E((\theta_i-\mu)(\bar{X}_i-\mu))/E(\bar{X}_i-\mu)^2.$$

由 $E(N_{ij}\mid\theta_i)=(1-\omega)\theta_i$,则 $E(X_{ij}\mid\theta_i)=E(\bar{X}_i\mid\theta_i)=\theta_i$,由双重期望公式知,$E(X_{ij})=\mu$,且

$$E((\theta_i-\mu)(\bar{X}_i-\mu))=\mathrm{Cov}(\theta_i,\bar{X}_i)=\mathrm{Var}(\theta_i)=\tau^2,$$

$$E(\bar{X}_i-\mu)^2=\mathrm{Var}(\bar{X}_i)=E(\mathrm{Var}(\bar{X}_i\mid\theta_i))+\mathrm{Var}(\theta_i)=\sigma^2/T_i+\tau^2.$$

综上可得

$$C=\mathrm{Var}(\theta_i)/(E(\mathrm{Var}(\bar{X}_i\mid\theta_i))+\mathrm{Var}(\theta_i))=T_i\tau^2/(\sigma^2+T_i\tau^2).$$

推论 1 若 θ_i 具有假设 2 的 Gamma 分布形式,则风险参数 θ_i 的信度估计为

$$\delta_{cre}=Z_1\bar{X}_i+(1-Z_1)\mu,$$

其中信度因子 $Z_1=T_i(1-\omega)/(T_i+\beta+\omega(1+\alpha-T_i))$.

注 2 考虑信度因子 Z 需要满足 $0\leqslant Z\leqslant 1$,且 Z 关于样本容量单调递增,即当 $T_i\to\infty$ 时,$Z\to 1$,从而信度估计越接近 $\bar{X}_i$. 这符合传统信度的解释,样本容量越大,样本均值越可靠. 由强大数定律知,$\bar{X}_i\to\theta_i$,a. s. ,由 Slutsky 定理得 $\delta_{cre}\to\theta_i$,a. s. . 因此信度估计是风险参数的强相合估计. 当 $\omega=0$ 时,风险参数的信度估计退化为泊松模型下的结果.

在统计学意义上风险参数的信度估计是一个较好的估计,但在保险实际中,一般样本容量并不足够大,这时可采用均方误差作为衡量的标准. 注意到,信度估计的均方误差为

$$E(\delta_{cre}-\theta_i)^2=E(Z(\bar{X}_i-\theta_i)+(1-Z)(\mu-\theta_i))^2=Z^2E(\bar{X}_i-\theta_i)^2+(1-Z)^2E(\mu-\theta_i)^2=Z^2E(\mathrm{Var}(X_{ij}\mid\theta_i))/T_i+(1-Z)^2\mathrm{Var}(\theta_i)=Z^2\sigma^2/T_i+(1-Z)^2\tau^2,$$

根据平方损失函数下风险参数的贝叶斯估计的定义可知 $E(\delta_{square}-\theta_i)^2\leqslant E(\delta_{cre}-\theta_i)^2\to 0$,所以 $E(\delta_{square}-\theta_i)^2\to 0$,即这 2 个估计是平方收敛的,且在均方误差意义下,δ_{square} 优于 δ_{cre}.

4 模拟研究

为了进一步说明这 4 个估计的优劣以及收敛速率的快慢,采用数值模拟的方法进行比较. 在模拟中,取零膨胀参数 $\omega=0.3$,先验分布参数 $\alpha=2$,$\beta=4$,尺度参数 $a=0.5$. 取 7 个不同的风险参数值

θ =0.35、0.40、0.45、0.50、0.55、0.60、0.65,并且取3个不同的样本容量 n = 20、50、150. 对每一个参数 θ 与样本容量 n 的组合,使用 Matlab 软件生成零膨胀泊松分布随机数,重复抽样5 000次. 在频率意义下分别利用4种估计方法得到参数 θ 的估计均值以及均方标准差,将相应的结果列于表1 ~ 表3中.

表1 当 n = 20时4个估计的模拟结果

θ	Stein 损失函数		Linex 损失函数		平方损失函数		平方损失函数	
	$\bar{\delta}_{Stein}$	均方标准差	$\bar{\delta}_{Linex}$	均方标准差	$\bar{\delta}_{square}$	均方标准差	$\bar{\delta}_{cre}$	均方标准差
0.35	0.327 6	0.129 0	0.382 5	0.129 8	0.388 5	0.133 4	0.387 9	0.129 6
0.40	0.364 0	0.142 3	0.418 6	0.137 4	0.425 1	0.140 6	0.423 4	0.137 0
0.45	0.402 9	0.150 0	0.457 1	0.141 0	0.464 3	0.143 8	0.461 7	0.140 9
0.50	0.442 9	0.162 9	0.496 6	0.150 8	0.504 4	0.153 2	0.501 1	0.151 1
0.55	0.481 5	0.175 1	0.534 7	0.160 0	0.543 1	0.162 0	0.539 0	0.160 2
0.60	0.517 4	0.185 3	0.570 2	0.166 5	0.579 2	0.167 8	0.574 9	0.167 9
0.65	0.555 1	0.196 6	0.607 3	0.175 2	0.616 9	0.175 8	0.612 7	0.176 7

表2 当 n = 50时4个估计的模拟结果

θ	Stein 损失函数		Linex 损失函数		平方损失函数		平方损失函数	
	$\bar{\delta}_{Stein}$	均方标准差	$\bar{\delta}_{Linex}$	均方标准差	$\bar{\delta}_{square}$	均方标准差	$\bar{\delta}_{cre}$	均方标准差
0.35	0.342 1	0.092 5	0.367 7	0.093 6	0.370 3	0.094 9	0.369 8	0.093 1
0.40	0.386 6	0.102 2	0.412 1	0.101 8	0.415 1	0.102 9	0.413 8	0.101 5
0.45	0.431 4	0.107 2	0.456 8	0.105 5	0.460 1	0.106 6	0.458 1	0.105 2
0.50	0.471 9	0.117 3	0.497 1	0.113 6	0.500 8	0.114 4	0.498 4	0.113 5
0.55	0.519 0	0.123 0	0.544 1	0.118 7	0.548 1	0.119 6	0.545 2	0.119 4
0.60	0.561 6	0.130 2	0.586 6	0.124 7	0.590 9	0.125 3	0.587 5	0.125 4
0.65	0.605 2	0.138 6	0.630 0	0.132 1	0.634 7	0.132 5	0.631 0	0.133 5

表3 当 n = 150时4个估计的模拟结果

θ	Stein 损失函数		Linex 损失函数		平方损失函数		平方损失函数	
	$\bar{\delta}_{Stein}$	均方标准差	$\bar{\delta}_{Linex}$	均方标准差	$\bar{\delta}_{square}$	均方标准差	$\bar{\delta}_{cre}$	均方标准差
0.35	0.345 4	0.058 4	0.354 6	0.058 4	0.355 5	0.058 7	0.355 2	0.058 2
0.40	0.394 6	0.063 5	0.403 8	0.063 3	0.404 9	0.063 6	0.404 3	0.063 3
0.45	0.443 7	0.068 8	0.452 9	0.068 5	0.454 1	0.068 8	0.453 2	0.068 5
0.50	0.489 5	0.071 6	0.498 6	0.070 7	0.499 9	0.070 9	0.498 6	0.070 7
0.55	0.539 2	0.075 0	0.548 3	0.074 2	0.549 8	0.074 4	0.548 6	0.074 4
0.60	0.587 2	0.079 8	0.596 2	0.078 8	0.597 8	0.079 0	0.596 6	0.079 3
0.65	0.636 0	0.083 6	0.645 0	0.082 5	0.646 7	0.082 7	0.645 0	0.082 9

从表1 ~ 表3结果可知,随着样本容量增加,4种估计方法的估计值与真实值越来越接近,且各估计的均方标准差也越来越小,能满足实际使用的需要. 在样本容量给定的情况下,4种估计方法的均方标准差都关于风险参数单调递增. 当样本容量较小时,在均方标准差意义下,Linex 损失下的贝叶斯估计优势更明显,Stein 损失下的贝叶斯估计对参数估计偏小,且参数真实值越大过低估计越明显. 平方损失与 Linex 损失下的贝叶斯估计对在真实参数较小时的参数估计偏大,并且前者的贝叶斯估计的偏大程度比后者更大. 平方损失与 Linex 损失下的贝叶斯估计对在真实参数较大时的参数估计偏小,并且前者的贝叶斯估计的偏小程度比后者更小. 总体来说,在均方标准差意义下,贝叶斯估计的优良性依次为 δ_{Linex}、δ_{square}、δ_{Stein}. 而信度估计 δ_{cre} 的均方误差比 δ_{square} 稍大一些(当 θ_i 值较小时 δ_{cre} 的均方误差甚至出现小于 δ_{square} 的均方误差的情形). 在实际使用中,决策者应结合估计的优良性和损失函数的(非)对称性等具体情况作出选择.

5 参考文献

[1] Lambert D. Zero-inflated Poisson regression with an application to defects in manufacturing [J]. Technometric, 1992,34(1):1-14.

[2] Falkner K, Mitter H, Moltchanova E, et al. A zero-inflated Poisson mixture model to analyse spread and abundance of the Western Corn Rootworm in Austria [J]. Agricultural Systems, 2019, 174(11):105-116.

[3] Wang Fu Kwun, Hailemariam S S. Sampling plans for the zero-inflated Poisson distribution in the food industry [J]. Food Control, 2018, 85(43):359-368.

[4] Chen Tian, Pan Wu, Wan Tang, et al. Variable selection

for distribution-free models for longitudinal zero-inflated count responses [J]. Statistics in Medicine,2016,35(16):2770-2785.

[5] Chen Kun,Huang Rui,Chan Ngai Hang,et al. Subgroup analysis of zero-inflated Poisson regression model with applications to insurance data [J]. Insurance:Mathematics and Economics,2019,86(2):8-18.

[6] Yau K K W,Wang Kui,Lee A H. Zero-inflated negative binomial mixed regression modeling of over-dispersed count data with extra zeros [J]. Biometrical Journal,2003,45(4):437-452.

[7] Denuit M,Maréchal X,Pitrebois S,et al. Actuarial modelling of claim counts: risk classification,credibility and Bonus-Malus systems [M]. England:John Wily and Sons,2007.

[8] Winkelmann R. Econometric analysis of count data [M]. 5th ed. Berlin:Springer,2008.

[9] 温利民,张美,程子红,等. 帕累托索赔分布中风险参数的经验贝叶斯估计 [J]. 应用概率统计,2015,31(3):225-237.

[10] 盛建为,钱夕元. 零膨胀对数级数分布的参数估计 [J]. 华东理工大学学报:自然科学版,2019,45(3):507-510.

[11] Wen Limin,Zhang Xiankun,Zheng Dan,et al. The credibility models under Linex loss functions [J]. Chin Quart J of Math,2012,27(3):397-402.

[12] Zellner A. Biased predictors,rationality and the evaluation of forecasts [J]. Ecnomics Letters,1986,21(1):45-48.

[13] James W,Stein C. Estimation with quadratic loss [C]//Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability. Berkeley:California University Press,1961,1:361-379.

[14] Bühlmann H. Experience rating and credibility [J]. Astin Bulletin,1967,4(3):199-207.

[15] 韦来生,张伟平. 贝叶斯分析 [M]. 合肥:中国科学技术大学出版社,2013:141-142.

[16] Casella G,Berger R L. Statistical inference [M]. 2nd ed. Pacific Grove:Duxbury,2002:189-190.

[17] Bühlmann H,Gisler A. A course in credibility theory and its applications [M]. Netherlands:Springer,2005.

The Bayesian Estimation of Risk Parameters in Zero-Inflated Poisson Model

ZHANG Liangchao,ZHOU Jinliang,WEN Limin*

(College of Mathematics and Information Science,Jiangxi Normal University,Nanchang Jiangxi 330022,China)

Abstract: The Bayesian model is established and the parameter estimation of the zero-inflated Poisson distribution is discussed in the paper. Bayesian estimates of risk parameters are obtained under the squared loss function,the Linex loss function,and the Stein loss function. Furthermore,the credibility theory is introduced,and the credibility estimates of the risk parameters are obtained under the mean square loss. In addition,the consistency of the estimates are proved. Finally,the convergence of the estimates is compared by numerical simulation.

Key words: zero-inflated Poisson model; Bayes estimators; parameter estimation

(责任编辑:曾剑锋)