

基于平衡损失函数下的信度模型

王天宇 韩 莹 杨永愉*
(北京化工大学 理学院, 北京 100029)

摘 要: 在平方损失的基础上引入惩罚项, 构成一种非对称的平衡损失函数。基于这种平衡损失函数, 运用贝叶斯统计推断理论得到了贝叶斯保费; 结合 Bühlmann 信度模型和 Bühlmann - Straub 信度模型, 给出了最优线性保费的估计结果。

关键词: 信度保费; 平衡损失函数; 贝叶斯决策

中图分类号: F840

引 言

信度理论是研究如何正确、合理地处理不同类型信息的理论。Mowbray^[1]和 Whitney^[2]以公理化的形式提出了信度理论的概念, 指出信度保费是先验均值和后验均值的一种凸组合, 即信度公式。Bailey^[3]证明了信度公式可以由贝叶斯定理导出。目前这一理论已经广泛应用于各种车辆险、人身意外险和火灾险等险种当中。通过信度理论厘定或校正的保险费率更加适合具体保险标的所面临的实际风险状况, 也更加符合商业保险的等价交换原则。

在传统的信度理论中, 通常采用对称的损失函数, 例如平方损失函数、绝对损失函数等。Bühlmann^[4]提出了构造损失随机变量观测值的线性组合, 使平方损失达到最小作为未来损失可信性估计的方法。然而, 自上世纪末以来, 许多学者注意到, 对称损失函数对于过高或者过低估计保险费率所产生的风险, 给出的度量结果是相同的。但是, 过高的保险费率与过低的保险费率在保险实务中所产生的影响及其后果却是不同的。于是, Dey 等^[5]提出了一种非对称损失函数, 即平衡损失函数。Jafari Jozani 等^[6]推广了 Dey 的平衡损失函数, 并研究了平衡损失函数下信度保费的贝叶斯估计及其性质。

本文以一种引入惩罚项的非对称平衡损失函数为基础, 运用贝叶斯统计推断理论得到了贝叶斯保

费, 并结合 Bühlmann 信度模型和 Bühlmann - Straub 信度模型, 对最优线性保费的估计结果进行了研究。

1 平衡损失函数的定义及性质

1.1 基本假定

设非负随机变量 X 表示理赔金额、理赔次数或理赔频率, 对于保险人来说, 随机变量 X 可统称为损失 X 。损失 X 的概率分布由概率空间 $(\Omega, F_\Omega, \{P_\theta, \theta \in \Theta\})$ 所决定, 其中 Ω 表示样本空间; F_Ω 是由 Ω 的所有子集构成的 σ -代数; 分布族 $\{P_\theta, \theta \in \Theta\}$ 中的 θ 称为分布参数。因为保险市场中的风险具有异质性, 所以在保险精算中, 常用分布参数 θ 的不同取值表示异质风险的不同风险水平, 故 θ 也称为风险参数; 给定风险参数 θ 的取值, P_θ 表示保险市场中风险水平为 θ 的某一险种的损失 X 所服从的概率分布。在保险市场中, 若风险是同质的, 则风险参数 θ 取某一固定值; 若风险是异质的, 根据贝叶斯统计理论, 将风险参数 θ 视为随机变量, 它服从参数空间 Θ 上定义的先验分布 $\pi(\theta)$, 在信度理论中, $\pi(\theta)$ 称为结构函数, 它是处理异质风险的重要工具和方法。

同时, 在信度理论中, 定义

$$\mu(\theta) = E(X|\theta)$$

$$\mu = E(X) = E[\mu(\theta)]$$

$$\nu(\theta) = \text{Var}(X|\theta)$$

$$\nu = E[\text{Var}(X|\theta)] = E[\nu(\theta)]$$

$$\alpha = \text{Var}[E(X|\theta)] = \text{Var}[\mu(\theta)]$$

其中 $\mu(\theta)$ 为损失 X 的假设均值, 是参数为 θ 时 X 的条件期望; μ 为损失 X 的总均值, 是不同风险参数下 X 的条件期望的平均; $\nu(\theta)$ 为损失 X 的过程方

收稿日期: 2012-10-26

第一作者: 男, 1991 年生, 硕士生

* 通讯联系人

E-mail: yangyongyu1765@sina.com

差,度量了同一风险水平的内在差异; ν 为损失 X 的同质方差,是 $\nu(\theta)$ 的均值; α 为 X 的异质方差,描述了不同风险水平 θ 下的损失 X 之间的差异^[7]。

1.2 贝叶斯决策与平衡损失函数

给定一个贝叶斯决策问题,并假定决策要素:

(1) 存在一个可观测的损失 X , 它的概率密度函数 $f(x|\theta)$ 依赖于风险参数 θ , 且 $\theta \in \Theta$, Θ 为参数空间;

(2) 在参数空间 Θ 上有一个先验分布 $\pi(\theta)$;

(3) 决策者可能采取的所有行动构成的行动集 $A = \{a\}$;

(4) 在 $\Theta \times A$ 上定义了一个损失函数 $L(\theta, a)$, 它表示风险参数为 θ 时, 决策者采用行动 a 所引起的损失。

对可观察的损失随机变量 X 作 n 次观测, 得样本 $\mathbf{x} = (x_1, x_2, \dots, x_n)$ 。由贝叶斯公式可得, 在样本 \mathbf{x} 给定下, θ 的后验密度函数为

$$\pi(\theta|\mathbf{x}) = \frac{f(\mathbf{x}|\theta)\pi(\theta)}{\int_{\Theta} f(\mathbf{x}|\theta)\pi(\theta)d\theta}$$

损失函数 $L(\theta, a)$ 对后验分布 $\pi(\theta|\mathbf{x})$ 的期望称为后验风险, 记为 $R(a|\mathbf{x})$, 即 $R(a|\mathbf{x}) = E^{\theta|\mathbf{x}}[L(\theta, a)]$ 。

在贝叶斯决策问题中, 从样本空间 $\chi = \{\mathbf{x} = (x_1, x_2, \dots, x_n)\}$ 到行动集 A 上的一个映射 $\delta(\mathbf{x})$ 称为该决策问题的一个决策函数。所有从 χ 到 A 上的决策函数组成的类称为决策函数类, 用 $D = \{\delta(\mathbf{x})\}$ 表示。贝叶斯决策的目标是, 在决策函数类 D 中选取决策函数 $\delta(\mathbf{x})$, 使其后验风险达到最小, 即 $R[\delta(\mathbf{x})|\mathbf{x}] = E^{\theta|\mathbf{x}}[L(\theta, \delta(\mathbf{x}))]$ 取最小^[8]。

本文在平方损失 $[\delta(\mathbf{x}) - \mu(\theta)]^2$ 的基础上引入惩罚项 $[\delta(\mathbf{x}) - \delta_0(\mathbf{x})]^2$, 构成了平衡损失函数

$$L[\mu(\theta), \delta(\mathbf{x})] = \omega [\delta(\mathbf{x}) - \delta_0(\mathbf{x})]^2 + (1 - \omega) [\delta(\mathbf{x}) - \mu(\theta)]^2 \quad \theta \in \Theta, \delta(\mathbf{x}) \in D, \omega \in [0, 1] \quad (1)$$

其中 ω 是给定的权重, $\delta(\mathbf{x})$ 为决策函数, $\delta_0(\mathbf{x})$ 为罚金, 是真实保费 $\mu(\theta)$ 的先验估计。可以证明式(1)定义的平衡损失函数对过高估计与过低估计所造成的损失具有非对称性。

令 $H(\mathbf{x}) = \delta(\mathbf{x}) - \mu(\theta)$, 则

$$\tau(\mathbf{x}) \triangleq \frac{\partial L}{\partial \delta(\mathbf{x})} = 2H(\mathbf{x}) + 2\omega[\mu(\theta) - \delta_0(\mathbf{x})]$$

(1) 当 $\delta_0(\mathbf{x}) \geq \mu(\theta)$, 即先验估计为过高估计

时, 若后验估计 $\delta(\mathbf{x}) < \mu(\theta)$, 有 $\tau(\mathbf{x}) < 0$; 若后验估计 $\mu(\theta) \leq \delta(\mathbf{x}) \leq \delta_0(\mathbf{x})$, 有 $\tau(\mathbf{x})$ 由负变为正; 若后验估计 $\delta(\mathbf{x}) \geq \delta_0(\mathbf{x})$, 有 $\tau(\mathbf{x}) > 0$ 。从而损失 $L[\mu(\theta), \delta(\mathbf{x})]$ 在 $(\mu(\theta), \delta_0(\mathbf{x}))$ 内取得最小值, 故过高估计所造成的损失小于过低估计所造成的损失。

(2) 当 $\delta_0(\mathbf{x}) < \mu(\theta)$, 即先验估计为过低估计时, 同理可得, 过低估计所造成的损失小于过高估计所造成的损失。

2 贝叶斯信度保费

设有损失 X 的 r 组观测数据, 各组依次有 n_1, n_2, \dots, n_r 个观测值 $\mathbf{x}_i = (x_{i,1}, x_{i,2}, \dots, x_{i,n_i})$, 相应的风险参数记为 $\theta_i, i = 1, 2, \dots, r$ 。在此基础上, 当预测风险参数为 θ_i 时, 下一期的信度保费 $\mu(\theta_i) = E(X_{i,n_i+1}|\theta_i)$ 。由式(1)得到相应的平衡损失函数

$$L[\mu(\theta_i), \delta(\mathbf{x}_i)] = \omega (\delta(\mathbf{x}_i) - \delta_0(\mathbf{x}_i))^2 + (1 - \omega) (\delta(\mathbf{x}_i) - \mu(\theta_i))^2 \quad (2)$$

其后验风险

$$R[\delta(\mathbf{x}_i)|\mathbf{x}_i] = E^{\theta_i|\mathbf{x}_i}[L(\mu(\theta_i), \delta(\mathbf{x}_i))] = \omega (\delta(\mathbf{x}_i) - \delta_0(\mathbf{x}_i))^2 + (1 - \omega) E^{\theta_i|\mathbf{x}_i}[(\delta(\mathbf{x}_i) - \mu(\theta_i))^2] \quad (3)$$

考虑后验风险的两种特殊情况, 即当 $\omega = 1$ 时, 信度保费的后验估计就是信度保费的先验估计, 即 $\hat{\delta}(\mathbf{x}_i) = \delta_0(\mathbf{x}_i)$; 当 $\omega = 0$ 时, 信度保费的后验估计 $\hat{\delta}(\mathbf{x}_i) = E^{\theta_i|\mathbf{x}_i}[\mu(\theta_i)]$ 。若 $\omega \in (0, 1)$, 即后验风险 $R[\delta(\mathbf{x}_i)|\mathbf{x}_i]$ 如式(3)所示, 可以证明, 在后验风险最小意义下信度保费的后验估计, 是上述两种特殊情况结果的加权平均, 而加权因子恰好是 ω 。

定理 1 设有损失 X 的 r 组观测数据, 各组依次有 n_1, n_2, \dots, n_r 个观测值 $\mathbf{x}_i = (x_{i,1}, x_{i,2}, \dots, x_{i,n_i})$, 对每一组观测值都有各自的风险参数 $\theta_i, i = 1, 2, \dots, r$ 。则在平衡损失函数 $L[\mu(\theta), \delta(\mathbf{x})]$ 下, 下一期的贝叶斯信度保费估计为 $\hat{\delta}(\mathbf{x}_i) = \omega \delta_0(\mathbf{x}_i) + (1 - \omega) E^{\theta_i|\mathbf{x}_i}[\mu(\theta_i)]$ 。

证明: 考虑平衡损失函数 $L[\mu(\theta), \delta(\mathbf{x})]$, 通过最小化后验风险, 得到平衡损失函数下, $\delta(\mathbf{x}_i)$ 的最优估计。

$$R[\delta(\mathbf{x}_i)|\mathbf{x}_i] = E^{\theta_i|\mathbf{x}_i}[L(\mu(\theta_i), \delta(\mathbf{x}_i))] = \omega (\delta(\mathbf{x}_i) - \delta_0(\mathbf{x}_i))^2 + (1 - \omega) E^{\theta_i|\mathbf{x}_i}[(\delta(\mathbf{x}_i) - \mu(\theta_i))^2]$$

对 $\delta(\mathbf{x}_i)$ 求导可得

$$2\omega(\delta(\mathbf{x}_i) - \delta_0(\mathbf{x}_i)) + 2(1 - \omega)\delta(\mathbf{x}_i) - 2(1 - \omega)E^{\theta|\mathbf{x}_i}[\mu(\theta_i)] = 0$$

故平衡损失函数下的信度保费估计为

$$\hat{\delta}(\mathbf{x}_i) = \omega\delta_0(\mathbf{x}_i) + (1 - \omega)E^{\theta|\mathbf{x}_i}[\mu(\theta_i)]$$

此外,对于同质风险与异质风险中的损失变量之间的协方差结果如定理 2 所示。

定理 2 假设损失 \mathbf{X}_1 和 \mathbf{X}_2 所对应的风险参数相等,均为 θ ,且 $\mu(\theta) = E(\mathbf{X}_1 | \theta) = E(\mathbf{X}_2 | \theta)$,则 $\text{Cov}(\mathbf{X}_1, \mathbf{X}_2) = \text{Var}[\mu(\theta)]$ 。

证明:因为在给定风险参数条件下,损失 \mathbf{X}_1 和 \mathbf{X}_2 是相互独立的,于是有

$$\begin{aligned} \text{Cov}(\mathbf{X}_1, \mathbf{X}_2) &= E(\mathbf{X}_1\mathbf{X}_2) - E(\mathbf{X}_1)E(\mathbf{X}_2) \\ E(\mathbf{X}_1\mathbf{X}_2) &= E[E(\mathbf{X}_1\mathbf{X}_2 | \theta)] = E[E(\mathbf{X}_1 | \theta)E(\mathbf{X}_2 | \theta)] = E[\mu^2(\theta)] \\ E(\mathbf{X}_1)E(\mathbf{X}_2) &= E[E(\mathbf{X}_1 | \theta)]E[E(\mathbf{X}_2 | \theta)] = \{E[\mu(\theta)]\}^2 \end{aligned}$$

所以

$$\text{Cov}(\mathbf{X}_1, \mathbf{X}_2) = E[\mu^2(\theta)] - \{E[\mu(\theta)]\}^2 = \text{Var}[\mu(\theta)]$$

需要指出的是,当 \mathbf{X}_1 和 \mathbf{X}_2 的风险参数不相等时, $\text{Cov}(\mathbf{X}_1, \mathbf{X}_2) = 0$ 。

3 Bühlmann 信度模型

在信度保费估计的实际应用中,由于经验数据的线性形式具有简洁的形式和较高的精度,所以人们常用它预测下一期的理赔状况,以此来确定一个较合理、较公平的信度保费。

假设 B1 ~ B3 条件成立。

B1 给定风险参数 θ_i ,相应的损失 \mathbf{X}_i 的独立同分布样本表示为 $\{X_{ij}, j=1, 2, \dots, n_i\}$;其中损失 \mathbf{X}_i 的假设均值与过程方差分别为 $\mu(\theta_i) = E(X_{ij} | \theta_i)$, $\nu(\theta_i) = \text{Var}(X_{ij} | \theta_i)$, $i=1, 2, \dots, r$ 。

B2 异质风险的风险参数与相应的损失 $(\theta_1, \mathbf{X}_1), (\theta_2, \mathbf{X}_2), \dots, (\theta_r, \mathbf{X}_r)$ 相互独立,且 $\theta_1, \theta_2, \dots, \theta_r$ 独立同分布。

B3 假设损失 \mathbf{X}_i 的罚金为 $\delta_0(\mathbf{X}_i)$,则其期望与协方差分别为 $E[\delta_0(\mathbf{X}_i)] = \mu$, $\text{Cov}[\delta_0(\mathbf{X}_i), X_{ij}] = s_i$, $j=1, 2, \dots, n_i$,其中 s_i 为常数。

令

$$\begin{cases} \mu = E(X_{ij}) = E[E(X_{ij} | \theta_i)] = E[\mu(\theta_i)] \\ \alpha = \text{Var}[E(X_{ij} | \theta_i)] = \text{Var}[\mu(\theta_i)] \\ \nu = E[\text{Var}(X_{ij} | \theta_i)] = E[\nu(\theta_i)] \end{cases} \quad (4)$$

定理 3 假定定理 1 的条件与 B1 ~ B3 假设成立,并且将风险参数为 θ_i 时下一期的信度保费 $\mu(\theta_i)$ 的估计限制在损失样本的线性函数族

$$\left\{ \delta_i = c_{i0} + \sum_{j=1}^{n_i} c_{ij}X_{ij}, c_{i0}, c_{ij} \in R^1 \right\}$$

内,给定平衡损失函数,则下一期的最优线性信度保费和信度因子分别为

$$\hat{\delta}_i^* = (1 - z_i^*)\mu + z_i^*\bar{X}_i$$

$$z_i^* = n_i b_i$$

$$\text{其中 } \bar{X}_i = \sum_{j=1}^{n_i} X_{ij} / n_i, b_i = \frac{\omega s_i + (1 - \omega)\alpha}{\alpha n_i + \nu}$$

证明:由于 $E(X_{ij}) = \mu$, $\text{Var}(X_{ij}) = \alpha$, $i=1, 2, \dots, r, j=1, 2, \dots, n_i$,由定理 2 可得

$$\text{Cov}(X_{ij}, X_{pq}) = 0$$

$$\text{Cov}[\mu(\theta_i), X_{pq}] = 0, i \neq p$$

$$\text{Cov}(X_{ij}, X_{iq}) = \alpha, j \neq q$$

令

$$\begin{aligned} \Psi = E \left\{ \omega \left[\delta_0(\mathbf{X}_i) - c_{i0} - \sum_{j=1}^{n_i} c_{ij}X_{ij} \right]^2 \right\} + E \left\{ (1 - \omega) \left[\mu(\theta_i) - c_{i0} - \sum_{j=1}^{n_i} c_{ij}X_{ij} \right]^2 \right\} \end{aligned} \quad (5)$$

对 c_{i0} 求导并令导数为 0,则有

$$c_{i0} = \left(1 - \sum_{j=1}^{n_i} c_{ij} \right) \mu \quad (6)$$

代入 Ψ ,对 c_{iq} , $q=1, 2, \dots, n_i$ 求导并令导数为 0,可得

$$\omega s_i + (1 - \omega)\alpha = \alpha \sum_{j=1}^{n_i} c_{ij} + \nu c_{iq}$$

$$c_{i1} = c_{i2} = c_{i3} = \dots = c_{in_i}$$

于是, $c_{iq} = \frac{\omega s_i + (1 - \omega)\alpha}{\alpha n_i + \nu} \equiv b_i$ 。则 $\hat{\delta}_i^* = c_{i0} + \sum_{j=1}^{n_i} c_{ij}X_{ij}$

$X_{ij} = (1 - z_i^*)\mu + z_i^*\bar{X}_i$,其中 $z_i^* = n_i b_i$ 。

4 Bühlmann - Straub 信度模型

假设 BS1 ~ BS3 条件成立。

BS1 给定风险参数 θ_i ,相应的损失 \mathbf{X}_i 的独立同分布样本表示为 $\{X_{ij}, j=1, 2, \dots, n_i\}$;其中损失 \mathbf{X}_i 的假设均值与过程方差分别为 $\mu(\theta_i) = E(X_{ij} | \theta_i)$, $\nu(\theta_i) = m_{ij}\text{Var}(X_{ij} | \theta_i)$, $i=1, 2, \dots, r$,其中 m_{ij} 称为观测值 X_{ij} 的权重,表示人们对 X_{ij} 的信任程度,

$$0 < m_{ij} < 1, \sum_{i=1}^r \sum_{j=1}^{n_i} m_{ij} = 1$$

BS2 异质风险的风险参数与相应的损失 $(\theta_1,$

$X_1), (\theta_2, X_2), \dots, (\theta_r, X_r)$ 相互独立,且 $\theta_1, \theta_2, \dots, \theta_r$ 独立同分布。

BS3 假设损失 X_i 的罚金为 $\delta_0(X_i)$, 则其期望与协方差分别为 $E[\delta_0(X_i)] = \mu$, $Cov[\delta_0(X_i), X_{ij}] = t_i, j=1, 2, \dots, n_i$, 其中 t_i 为常数。

在 BS1 ~ BS3 的假设条件下, 与式(4)相比较有

$$\nu' = E[\text{Var}(X_{ij}|\theta_i)] = E[\nu(\theta_i)/m_{ij}] = \nu/m_{ij}$$

定理 4 假定定理 1 的条件与 BS1 ~ BS3 假设成立, 并且将风险参数为 θ_i 时下一期的信度保费 $\mu(\theta_i)$ 的估计限制在损失样本的线性函数族 $\{\delta_i = c_{i0} + \sum_{j=1}^{n_i} c_{ij} X_{ij}, c_{i0}, c_{ij} \in R^1\}$ 内, 给定平衡损失函数, 则下一期的最优线性信度保费和信度因子分别为

$$\hat{\delta}_i^{**} = (1 - z_i^{**})\mu + z_i^{**} \bar{X}_i$$

$$z_i^{**} = b'_i \sum_{j=1}^{n_i} m_{ij}$$

$$\text{其中 } \bar{X}_i = \sum_{j=1}^{n_i} m_{ij} X_{ij} / \sum_{j=1}^{n_i} m_{ij}, b'_i = \frac{\omega t_i + (1 - \omega)\alpha}{\alpha \sum_{j=1}^{n_i} m_{ij} + \nu}。$$

证明过程与定理 3 相似。

5 结束语

本文在平方损失的基础上引入惩罚项, 构成了一种非对称的平衡损失函数。通过惩罚项中含有的调整罚金 $\delta_0(x)$ 使损失达到最小。基于这种平衡损失函数, 得到的信度保费估计具有线性表达式, 形式简单, 在实际使用中具有很大的便利性。由于损失

函数与信度因子是信度保费模型中的两个关键因素, 所以这两者之间是否存在一定的联系也是值得关注的研究课题。

参考文献:

[1] Mowbray A H. How extensive a payroll exposure is necessary to give dependable pure premium? [J]. Proceedings of the Casualty Actuarial Society, 1914, 1: 24-30.

[2] Whitney A W. The theory of experience rating [J]. Proceedings of the Casualty Actuarial Society, 1918, 4: 274-292.

[3] Bailey A L. A generalized theory of credibility [J]. Proceedings of the Casualty Actuarial Society, 1945, 32: 13-20.

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

[5] Dey D K, Ghosh M, Strawderman W E. On estimation with balanced loss functions [J]. Statistics & Probability Letters, 1999, 45: 97-101.

[6] Jafari Jozani M, Marchand E, Parsian A. On estimation with weighted balanced-type loss function [J]. Statistics & Probability Letters, 2006, 76: 773-780.

[7] 韩天雄. 非寿险精算 [M]. 北京: 中国财政经济出版社, 2010.

Han T X. Casualty actuarial [M]. Beijing: China Financial and Economic Press, 2010. (in Chinese)

[8] 茆诗松. 贝叶斯统计 [M]. 北京: 中国统计出版社, 1999.

Mao S S. Bayesian statistics [M]. Beijing: China Statistics Press, 1999. (in Chinese)

On the balanced loss function of credibility models

WANG TianYu HAN Xuan YANG YongYu

(School of Science, Beijing University of Chemical Technology, Beijing 100029, China)

Abstract: On the basis of the square loss function, we introduce a penalty term to constitute a balanced loss function; this is an example of an asymmetric loss function. Based on this balanced loss function, we obtain the Bayesian premium by using the theory of Bayesian statistical inference. Then, by means of the Bühlmann credibility model and the Bühlmann - Straub credibility model, we obtain an estimation of the optimal linear premiums. In real life, these not only provide fair and reasonable credibility premiums, but are also quite convenient.

Key words: credibility premium; balanced loss function; Bayesian decision