

## 熵损失函数下的信度模型

王娜娜

(宜春职业技术学院公共基础部, 江西 宜春 336000)

**摘 要:** 本文研究了信度模型问题. 利用熵损失函数, 获得了风险保费的信度估计和经验 Bayes 信度估计. 所获结果是对现有风险保费信度估计和经验 Bayes 信度估计的一个补充.

**关键词:** 熵损失函数; 信度估计; 保费估计

MR(2010) 主题分类号: 62C12; 62F10

中图分类号: O212.1

文献标识码: A

文章编号: 0255-7797(2015)06-1372-07

### 1 引言

在风险理论<sup>[1]</sup>中, 如何对风险定义合适的保费是非常重要的. 在对风险定价时, 必须考虑到投保人的情况. 由于投保人年龄、性别等的不同, 他们的保费也有所不同. 这种保费的不一致性, 称此风险具有非齐次性. 制定这种风险的保费通常采用信度理论<sup>[2,3]</sup>.

在信度理论中, 假设风险  $X$  可以由风险参数  $\Theta$  识别. 由于风险的非齐次性,  $\Theta$  是一个随机变量, 它的取值服从分布  $\pi(\theta)$ , 称为结构分布或经验分布. 风险  $X$  具有一系列观测值, 根据这些观测值估计或预测未来的风险  $X$ . 信度原理就是利用先验分布  $\pi(\theta)$  与样本信息估计风险保费的一种方法.

但是大部分信度理论的结果都是在纯保费原理下获得的, 而纯保费不能满足保费的正的安全负荷. 解决这个问题的办法之一是通过修改损失函数. 这种想法最早是由 Gerber<sup>[4]</sup>提出, 他将平方损失函数修改为指数加权平方损失函数, 得到了 Esscher 保费原理下的风险保费的信度估计. Furman, Zitikis 在广义加权损失函数<sup>[5]</sup>下, 温利民分别在新型广义加权损失函数<sup>[6]</sup>、Linex 损失函数<sup>[7]</sup>、指数损失函数<sup>[8]</sup>等损失函数下, 张佳佳在相对损失函数<sup>[9]</sup>下研究了信度模型问题.

利用熵损失函数已经有如下的研究: 文献 [10] 研究了熵损失函数下两参数广义指数分布形状参数的 Bayes 估计, 文献 [11] 研究了在熵损失函数下定数截尾情形的参数估计, 文献 [12] 研究了熵损失函数下广义线性混合模型协方差矩阵谱的分解估计等等. 目前作者还未见有报道利用熵损失函数研究信度模型.

本文将在熵损失函数下, 考虑风险保费的信度估计与保费估计. 全文分为四节, 第二节给出信度模型的基本假设, 第三节给出熵损失函数下的信度估计, 最后给出多合同信度模型.

### 2 信度模型的基本假设

我们定义熵损失函数:

$$L(X, P) = \frac{P}{X} - \ln \frac{P}{X} - 1. \quad (2.1)$$

\*收稿日期: 2014-04-17

接收日期: 2014-06-04

作者简介: 王娜娜 (1982-), 女, 江西宜春, 讲师, 主要研究方向: 概率论与数理统计.

假设 1 非负随机变量  $X$  可以由风险参数  $\Theta$  来识别, 风险参数的先验分布为  $\pi(\theta)$ .

假设 2 给定  $\Theta$ , 随机变量序列  $X_1, X_2, \dots, X_n$  独立同分布于  $X$ , 具有相同的分布函数  $F_X(x, \theta)$ , 记  $\underline{X}_n = (X_1, X_2, \dots, X_n)$  表示到时刻为止的索赔经历.

当风险参数  $\Theta$  已知时, 可以用  $\Theta$  的某个函数来预测第  $n+1$  年的损失  $X_{n+1}$ .

**定理 2.1** 在熵损失函数 (2.1) 下, 若风险参数  $\Theta$  已知, 则第  $n+1$  年的损失  $X_{n+1}$  的最优预测为

$$H(X|\Theta) = \frac{1}{E(X^{-1}|\Theta)}. \quad (2.2)$$

由于  $\text{Cov}(X^{-1}, X|\Theta) \geq 0$ , 则  $H(X|\Theta) \geq E(X|\Theta)$ , 其中  $E(X|\Theta) \geq 0$  为纯保费. 因此风险保费  $H(X|\Theta)$  具有正的安全负荷.

在实际中, 风险参数  $\Theta$  是未知的, 因此风险保费  $H(X|\Theta)$  需要由样本来估计.

### 3 风险保费的估计

若没有索赔样本, 即  $n=0$  时, 可以用一个常数来估计风险保费  $H(X|\Theta)$ .

首先用某个常数  $C$  来预测第  $n+1$  年的损失  $X_{n+1}$ , 在熵损失函数 (2.1) 下求解

$$\min_C E[L(X_{n+1}, C)] = \min_C E\left[\frac{C}{X_{n+1}} - \ln \frac{C}{X_{n+1}} - 1\right],$$

得到

$$C = \frac{1}{E[X^{-1}|\Theta]} := H_C(X),$$

称  $H_C(X)$  为聚合保费.

其次, 用某个常数  $R$  来估计风险保费  $H(X|\Theta)$ , 即求解

$$\min_R E[L(H(X|\Theta), R)] = \min_R E\left[\frac{R}{H(X|\Theta)} - \ln \frac{R}{H(X|\Theta)} - 1\right],$$

得到

$$R = \frac{1}{E[H(X|\Theta)]} := H_R(X),$$

称  $H_R(X)$  为风险保费的聚合估计.

根据第二节的假设, 我们有样本  $X_1, X_2, \dots, X_n$ , 在此基础上来估计风险保费  $H(X|\Theta)$ , 假设  $\mathfrak{L}$  为  $\underline{X}_n$  的可测函数类.

**定理 3.1** 在熵损失函数 (2.1) 下, 得到  $X_{n+1}$  的最优预测为

$$H_C(\underline{X}_n) = \frac{1}{E[X_{n+1}^{-1}|\underline{X}_n]},$$

称  $H_C(\underline{X}_n)$  为熵损失函数 (2.1) 下的 Bayes 保费.

**证** 由 Bayes 定理, 求解下列最小化问题

$$\min_{h \in \mathfrak{L}} E[L(X_{n+1}, h(\underline{X}_n))] = \min_{h \in \mathfrak{L}} E\left[\frac{h(\underline{X}_n)}{X_{n+1}} - \ln \frac{h(\underline{X}_n)}{X_{n+1}} - 1 | \underline{X}_n\right].$$

设

$$\phi = E\left[\frac{h(\underline{X}_n)}{\underline{X}_{n+1}} - \ln \frac{h(\underline{X}_n)}{\underline{X}_{n+1}} - 1 | \underline{X}_n\right].$$

令  $\frac{\partial \phi}{\partial h} = 0$ , 可以得到以下方程

$$E\left[\frac{1}{\underline{X}_{n+1}} - \frac{1}{h(\underline{X}_n)} | \underline{X}_n\right] = 0,$$

于是

$$h(\underline{X}_n) = \frac{1}{E[\underline{X}_{n+1}^{-1} | \underline{X}_n]} := H_C(\underline{X}_n).$$

定理得证.

**定理 3.2** 在熵损失函数 (2.1) 下, 风险保费的最优估计为

$$H_R(\underline{X}_n) = \frac{1}{E[H(X|\Theta)^{-1} | \underline{X}_n]},$$

称为熵损失函数 (2.1) 下的风险保费的 Bayes 估计.

**证** 求解下列问题

$$\min_{h \in \mathcal{L}} E[L(H(X|\Theta), h(\underline{X}_n))] = \min_{h \in \mathcal{L}} E\left[\frac{h(\underline{X}_n)}{H(X|\Theta)} - \ln \frac{h(\underline{X}_n)}{H(X|\Theta)} - 1 | \underline{X}_n\right],$$

可以得到定理的结论.

**例** 设  $X_1, X_2, \dots, X_n, X_{n+1}$  在风险参数  $\Theta = \theta$  给定时为独立同分布的随机变量, 且  $X_i \sim U(a, \theta), \theta \sim U(c, d)$ , 其中  $a, c, d$  已知,  $a > 0, c > 0$ , 则

$$f(X|\theta) = \frac{1}{\theta - a}, \pi(\theta) = \frac{1}{d - c},$$

那么聚合保费为

$$H_C(X) = \frac{1}{E[X^{-1} | \Theta]} = \frac{\theta - a}{\ln \theta - \ln a}.$$

聚合估计为

$$H_R(X) = \frac{1}{E[H(X|\Theta) | \Theta]} = \frac{\theta - a}{\ln \theta - \ln a}.$$

而

$$\pi(\theta | \underline{X}_n) = \frac{(1-n)(\theta-a)^{-n}}{(d-a)^{-n+1} - (c-a)^{-n+1}},$$

因此 Bayes 保费为

$$H_C(\underline{X}_n) = \frac{1}{E[\underline{X}_{n+1}^{-1} | \underline{X}_n]} = \frac{(d-a)^{-n+1} - (c-a)^{-n+1}}{(1-n) \int_c^d (\theta-a)^{-n+1} \ln\left(\frac{\theta}{a}\right) d\theta}.$$

Bayes 估计为

$$H_R(\underline{X}_n) = \frac{1}{E[H(X|\Theta)^{-1} | \underline{X}_n]} = \frac{(d-a)^{-n+1} - (c-a)^{-n+1}}{(1-n) \int_c^d (\theta-a)^{-n+1} \ln\left(\frac{\theta}{a}\right) d\theta}.$$

## 4 多合同信度模型

### 4.1 基本假设

假设有  $k$  个被保险人的保单组合, 每份保单都有若干年索赔记录, 则数据样本为  $\{X_{ij}, i = 1, 2, \dots, k, j = 1, 2, \dots, n_i\}$ . 由于风险的非齐次性, 对每份保单有一种风险属性  $\Theta_i$ . 我们的目标是估计  $E(X_{i,n+1}|\Theta_i) = \mu(\theta_i)$ .

设  $X_i = (X_{i1}, X_{i2}, \dots, X_{in_i})'$  表示第  $i$  份保单的索赔经历,  $n_i$  表示第  $i$  份保单索赔样本的容量, 记

$$\underline{X}_{n_k} = (X'_1, X'_2, \dots, X'_k)'$$

为所有的索赔样本.

结合经典 Bühlmann-Straub 模型, 提出 3 下面几个假设:

假设 4.1 对第  $i$  份合同,  $i = 1, 2, \dots, k$ , 给定风险参数  $\Theta_i = \theta_i$ , 随机变量序列  $\{X_{ij}, j = 1, 2, \dots, n_i\}$  独立同分布, 且条件期望与方差为

$$E(X_{ij}^{-1}|\Theta_i) = \mu(\theta_i), \text{Var}(X_{ij}^{-1}|\Theta_i) = s^2(\theta_i), i = 1, 2, \dots, k; j = 1, 2, \dots, n_i.$$

假设 4.2 随机变量  $\Theta_1, \Theta_2, \dots, \Theta_k$  相互独立, 并且具有相同的先验分布  $\pi(\theta)$ .

假设 4.3 合同  $\{(X_i, \Theta_i), i = 1, 2, \dots, k\}$  之间相互独立.

为了方便, 作以下符号的假定:

$$E[\mu(\theta_i)] = \mu, E(s^2(\theta_i)) = s^2, \text{Var}(\mu(\theta_i)) = \sigma^2.$$

可以验证

$$E[X_{ij}^{-1}] = \mu, \text{Var}(X_{ij}^{-1}) = \sigma^2, i = 1, 2, \dots, k; j = 1, 2, \dots, n_i.$$

### 4.2 信度估计

在多合同模型下, 熵损失函数 (2.1) 下第  $i$  个合同的风险保费为

$$P(\Theta_i) = \frac{1}{E[X_{ij}^{-1}|\Theta_i]}, \quad i = 1, 2, \dots, k.$$

Bayes 保费为

$$H_{C_i}(\underline{X}_{n_k}) = \frac{1}{E[\mu(\theta_i)|\underline{X}_{n_k}]}.$$

要求得 Bayes 保费, 必须知道  $\Theta_i$  与所有样本的联合分布. 而在实际运用中, 一般很难知道样本的具体分布, 特别是先验分布的具体形式. 因此用 Bayes 保费来定价保险产品受到很大的限制. 但是, 若将  $\mu(\theta_i)$  限定在样本的线性函数类

$$\{\widehat{\mu(\theta_i)} = h_0 + \sum_{m=1}^k \sum_{l=1}^{n_i} h_{ml} X_{ml}^{-1}, \quad h_0, h_{ml} \in R\}$$

中, 则可得到样本的最优线性估计, 在信度理论中称之为信度保费.

为求第  $i$  个合同的信度保费, 我们先求解下面的最小化问题

$$\min_{h_0, h_{ml} \in R} E[(\mu(\theta_i) - h_0 - \sum_{m=1}^k \sum_{l=1}^{n_i} h_{ml} X_{ml}^{-1})^2]. \quad (4.1)$$

求解上式得到下面的定理.

**定理 4.1** 在多合同模型 4.1–4.3 的假设下, 求解 (4.1) 式得到  $\mu(\theta_i)$  的最优线性估计

$$\widehat{\mu(\theta_i)} = Z_i \bar{\mu}_i + (1 - Z_i) \mu,$$

其中信度因子为  $Z_i = \frac{n_i \sigma^2}{n_i \sigma^2 + s^2}$ , 且  $\bar{\mu}_i = \frac{1}{n_i} \sum_{l=1}^{n_i} X_{ml}^{-1}$ ,  $h_{m\cdot} = \sum_{l=1}^{n_i} h_{ml}$ .

**证** 为下面记号方便, 令

$$\varphi = E[(\mu(\theta_i) - h_0 - \sum_{m=1}^k \sum_{l=1}^{n_i} h_{ml} Y_{ml})^2],$$

其中  $Y_{ml} = X_{ml}^{-1}$ .

首先关于  $h_0$  求导并令导数为 0, 即  $\frac{\partial \phi}{\partial h_0} = 0$ , 得到

$$h_0 = E[(\mu(\theta_i) - \sum_{m=1}^k \sum_{l=1}^{n_i} h_{ml} Y_{ml})] = \mu - \mu \sum_{m=1}^k \sum_{l=1}^{n_i} h_{ml},$$

将上式代入  $\varphi$  中, 于是

$$\varphi = E[(\mu(\theta_i) - \mu + \mu \sum_{m=1}^k \sum_{l=1}^{n_i} h_{ml} - \sum_{m=1}^k \sum_{l=1}^{n_i} h_{ml} Y_{ml})^2],$$

对上式两边关于  $h_{m'l'}$  求导, 其中  $m' = 1, 2, \dots, k; l' = 1, 2, \dots, n_i$ , 并且令  $\frac{\partial \phi}{\partial h_{m'l'}} = 0$ , 于是

$$E[(\mu(\theta_i) - \mu + \mu \sum_{m=1}^k \sum_{l=1}^{n_i} h_{ml} - \sum_{m=1}^k \sum_{l=1}^{n_i} h_{ml} Y_{ml})(\mu - Y_{m'l'})] = 0,$$

化简后得到

$$\text{Cov}(\mu(\theta_i), Y_{m'l'}) = \sum_{m=1}^k \sum_{l=1}^{n_i} h_{ml} \text{Cov}(\mu(\theta_i), Y_{m'l'}) = \sum_{l=1}^{n_i} h_{m'l} \text{Cov}(Y_{m'l}, Y_{m'l'}), \quad (4.2)$$

而

$$\text{Cov}(Y_{m'l}, Y_{m'l'}) = E[\text{Cov}(Y_{m'l}, Y_{m'l'}) | \Theta] + \text{Cov}(E(Y_{m'l} | \Theta), E(Y_{m'l'} | \Theta)),$$

那么当  $l' \neq l$  时,  $\text{Cov}(Y_{m'l}, Y_{m'l'}) = \sigma^2$ , 当  $l' = l$  时,  $\text{Cov}(Y_{m'l}, Y_{m'l'}) = s^2 + \sigma^2$ , 且

$$\text{Cov}(\mu(\theta_i), Y_{m'l'}) = \begin{cases} 0, & m' \neq i, \\ \sigma^2, & m' = i, \end{cases}$$

于是 (4.2) 式可化为

$$\begin{cases} \sum_{l=1}^{n_i} h_{m'l} \sigma^2 + h_{m'l'} s^2 = 0, & m' \neq i, \\ \sum_{l=1}^{n_i} h_{m'l} \sigma^2 + h_{m'l'} s^2 = \sigma^2, & m' = i, \end{cases} \quad (4.3)$$

由 (4.3) 式可得  $m' \neq i$  时,  $h_{m'l} = -\frac{\sigma^2}{s^2} h_{m'l'}$ ,  $m' = i$  时,  $h_{i'l} = \frac{\sigma^2}{s^2} (1 - h_{i'l'})$ . 即

$$m' \neq i, \quad h_{m'l} = 0; \quad m' = i, \quad h_{i'l} = \frac{\sigma^2}{n_i \sigma^2 + s^2}.$$

因此  $\mu(\theta_i)$  的最优线性信度估计为

$$\begin{aligned} \widehat{\mu(\theta_i)} &= \mu - \mu \sum_{m=1}^k \sum_{l=1}^{n_i} h_{ml} + \sum_{m=1}^k \sum_{l=1}^{n_i} h_{ml} Y_{ml} \\ &= \mu - \mu \sum_{l=1}^{n_i} h_{il} + \sum_{l=1}^{n_i} h_{il} Y_{il} \\ &= \mu - \mu \sum_{l=1}^{n_i} \frac{\sigma^2}{n_i \sigma^2 + s^2} + \sum_{l=1}^{n_i} \frac{\sigma^2}{n_i \sigma^2 + s^2} Y_{il} \\ &= \frac{n_i \sigma^2}{n_i \sigma^2 + s^2} \bar{\mu}_i + \frac{s^2}{n_i \sigma^2 + s^2} \mu \\ &= Z_i \bar{\mu}_i + (1 - Z_i) \mu. \end{aligned}$$

定理得证.

**注 1** 若  $n_1 = n_2 = \cdots = n_i = n$ , 此时称为平衡模型, 于是  $Z_i = \frac{n\sigma^2}{n\sigma^2 + s^2}$ ,  $\bar{\mu}_i = \frac{1}{n} \sum_{l=1}^n Y_{ml}$ .

**注 2** 若  $s^2 \rightarrow \infty$ , 则  $Z_i \rightarrow 0$ , 此时有  $\widehat{\mu(\theta_i)} \rightarrow \mu$ .

**注 3** 若  $\sigma^2 \rightarrow \infty$ , 则  $Z_i \rightarrow 1$ , 此时有  $\widehat{\mu(\theta_i)} \rightarrow \bar{\mu}_i$ .

**命题**  $E[(\mu(\theta_i) - \widehat{\mu(\theta_i)})^2] = Z_i \frac{s^2}{n_i}$ .

**证**

$$\begin{aligned} E[(\mu(\theta_i) - \widehat{\mu(\theta_i)})^2] &= E[(\mu(\theta_i) - Z_i \bar{\mu}_i + (1 - Z_i) \mu)^2] \\ &= E[(\mu(\theta_i) - \bar{\mu}_i + (1 - Z_i)(\bar{\mu}_i - \mu))^2] \\ &= E[(\mu(\theta_i) - \bar{\mu}_i)^2] + 2(1 - Z_i) E[(\mu(\theta_i) - \bar{\mu}_i) \cdot (\bar{\mu}_i - \mu)] \\ &\quad + (1 - Z_i)^2 E[(\bar{\mu}_i - \mu)^2] \\ &= \frac{s^2}{n_i} + 2(1 - Z_i) \left(-\frac{s^2}{n_i}\right) + (1 - Z_i)^2 \left(\frac{s^2}{n_i} + \sigma^2\right) \\ &= (1 - Z_i) \sigma^2 = Z_i \cdot \frac{s^2}{n_i}. \end{aligned}$$

## 参 考 文 献

- [1] 卡尔斯等. 现代精算风险理论 [M]. 北京: 科学出版社, 2005.
- [2] 严颖, 成世学, 程侃. 保险精算方法 (三) 信度理论 [J]. 数理统计与管理, 1996, 15(06): 59-64.
- [3] 成世学. 关于可信性模型的若干评注 [J]. 应用概率统计, 2002, 18(4): 438-448.

- [4] Gerber H U. Credibility for esscher premium[J]. Mitteilungen der Vereinigung Schweizerischer Versicher Ungsmathematiker, 1980, 3: 307–312.
- [5] Edward Furman, Ricardas Zitikis. Weighted premium calculation principles[J]. Insurance: Math. Econ., 2008, 42(1): 459–465.
- [6] 温利民, 梅国平. 新型广义加权保费原理下风险保费的信度估计 [J]. 应用数学学报, 2013, 36(2): 257–267.
- [7] 温利民等. LINEX 损失函数下的信度模型 (英文)[J]. 数学季刊, 2012, 27(3): 397–402.
- [8] Wen LiMin, et al. The Credibility Premiums for Exponential Principle[J]. Acta Math. Sinica, 2011, 27(11): 2217–2228.
- [9] 张佳佳. 相对损失函数下的信度保费 (英文)[J]. 应用概率统计, 2007, 23(2): 157–164 .
- [10] 王国富. 熵损失函数下两参数广义指数分布形状参数的 Bayes 估计 [J]. 统计与决策, 2010,(1): 154–155.
- [11] 王德辉, 宋立新. 在熵损失函数下定数截尾情形的参数估计 – 指数分布情形 [J]. 应用概率统计, 1999, 15(2): 176–186.
- [12] 西日古力, 吴黎军. 熵损失函数下广义线性混合模型协方差矩阵谱的分解估计 [J]. 统计与决策, 2012, (22): 7–9.
- [13] 史建红, 关丽娜. 非对称损失函数下 Burr XII 型分布可靠性指标的 Bayes 估计 [J]. 数学杂志, 2013(1): 124–131.
- [14] 胡桂开, 彭萍, 罗汉. 平衡损失下回归系数最优线性无偏估计的相对效率 [J]. 数学杂志, 2010, 30, 1059–1064.

## CREDIBILITY MODELS UNDER THE ENTROPY LOSS FUNCTION

WANG Na-na

*(Department of Basic, Yichun Vocational Technical College, Yichun 336000, China)*

**Abstract:** In this paper, we study the problem of the reliability model. Using entropy loss function, we obtain the credibility estimator and empirical Bayes estimator of risk premium. The result is a complement to the existing risk premium reliability estimation and empirical Bayes reliability estimation.

**Keywords:** entropy loss function; credibility estimator; premium estimator

**2010 MR Subject Classification:** 62C12; 62F10