应用概率统计 第 32 卷
第 5 期 2016 年 10 月

Chinese Journal of Applied Probability and Statistics
Oct., 2016, Vol. 32, No. 5, pp. 463-475
doi: 10.3969/j.issn.1001-4268.2016.05.003

# The Maximum Entropy Method to the Credibility Estimation *

HU Yingying WU Lijun⋆ SUN Yi

(*College of Mathematics and System Science, Xinjiang University, Urumqi, 830046, China*)

**Abstract:** Besides the claims data in the past, certain assumptions about the distribution of claims $X_i\,(i=1,2,\ldots,n)$ are required to derive the credibility premium in the classical theory. In the paper, the credibility premium can be calculated via the maximum entropy method if we know nothing about the distribution of claims $X_i\,(i=1,2,\ldots,n)$. Furthermore, two corollaries are obtained under certain assumptions, that is, new claims have more weight than the old ones and the classical credibility formula is a special case of the credibility premium derived in the present paper. Finally, the simulation study is presented to illustrate that the credibility premium in the present paper is better than other models if the mean square error is taken as the evaluation criterion.

**Keywords:** credibility estimation; entropy; the maximum entropy method; Lagrange multiplier

**2010 Mathematics Subject Classification:** 62P05; 91B30

## §1. Introduction

In the actuarial field, credibility theory is an empirical model used to calculate the premium. Given risk and data of claims in the past, actuaries can compute the feasible premium by applying credibility theory.

Bühlmann [1] proposed the argument that credibility premium should be a weighted average of individual premium and collective premium, i.e., $P_c = Z\overline{X} + (1-Z)\mu$, where $\overline{X} = \big(\sum\limits_{i=1}^{n} X_i\big)/n$, $\mu$ is the collective premium and $Z = n\tau^2/(n\tau^2+\sigma^2)$ with $\sigma^2 = \mathsf{E}[\mathsf{Var}\,(X_i\,|\,\theta)]$, $\tau^2 = \mathsf{Var}\,[\mathsf{E}(X_i\,|\,\theta)]$, $\mu(\theta) = \mathsf{E}(X_i\,|\,\theta)$. The dominant advantage of credibility premium formula outweighing Bayes premium is that we can calculate premium by credibility formula even if we do not know the distribution of $X_i$. Based on the classical Bühlmann model, Bühlmann and Straub [2] introduced natural weight $\omega$, which extends the Bühlmann model to a large extent.

*The project was supported by National Natural Science Foundation of China (11361058).

⋆Corresponding author, E-mail: xjmath@xju.edu.cn.

Received August 8, 2014. Revised February 4, 2015.

It has been proved in bankruptcy theory the insurance companies must go bankrupt if they only charge net premium, see [3] and [4]. A method to tackle this problem is to replace the squared loss function in the classical credibility premium with other loss functions. An extension to the credibility formula from the loss viewpoint was given by Gómez-Déniz [5,6]. He proposed the balanced square loss function,

$$L(\delta,\theta)=\omega\cdot h(\theta)\cdot(\delta-\delta_0)^2+(1-\omega)\cdot h(\theta)\cdot(\delta-\theta)^2,$$

where $\delta_0$ is chosen as a prior "target" estimator of $\theta$, obtained for instance from the criterion of maximum likelihood estimator, least squares, or unbiased among others. Then he established the credibility theory for this loss function. In domestic, some scholars made lots of contribution to this area. For instance, Wen and Wu [7] derived credibility formula under the exponential premium principle, and Wen et al. [8] obtained the corresponding results under balanced loss function and the similar results under a new type of generalized weighted loss function by Wen and Mei [9].

Frees et al. [10] developed links between credibility theory and longitudinal (or panel) data models. They demonstrated how longitudinal data models can be applied to the credibility rate-making problem. The usual credibility models induces time dependence among annual claim characteristics (such as number of accidents, associated costs, etc.) via the sharing of common random effects, see [11,12] for more information. Wen et al. [13] proved the Bühlmann credibility model with dependent structure between risks. Zheng et al. [14] considered credibility model with time effect within risks. Teng and Wu [15] derived double dependent credibility premium under the exponential premium principle.

The maximum-entropy principle provides a means to obtain least-biased statistical inference whenever partial and insufficient information is available. The maximum entropy method (MEM) was first proposed by Jaynes [16]. Since then, he did numerous research in this field and established theoretical basis which is used to solve quantitative problem by exploring the maximum entropy method. Later on, Kullback [17] proposed the Kullback minimum crossing entropy method which is similar to the maximum entropy method (MEM).

MEM has become popular in many areas since it appeared, such as thermodynamics, information theory, biotechnology, industry, agriculture and so on. Nevertheless, in the actuarial field, only several scholars combined MEM and insurance together successfully. For instance, Jessop [18] used the minimum crossing entropy method to determine the weight and estimate the target of utility function. Besides, on the basis of the utility function, Abbas [19] adjusted the utility function to the cumulative probability function. And

Landsman and Makov [20] estimated credibility factor and the distribution of the scale parameters in dispersed exponential family with MEM method. What's more, Darooneh [21] deduced the utility function of non-life insurance market. And Payandeh Najafabadi et al. [22] calculated the approximate Bayes estimators by MEM.

Consider that many scholars have done a lot of research by exploring maximum entropy method, but it appears that the applications of it in the field of studying insurance premium is still rare. This paper will make some tries in this area.

The structure of this paper is organized as follows. We introduce the maximum entropy method in the second part; and in the third part, by exploring the maximum entropy method, a main theorem with its applications is given. As some kind of verification, we also provide numerical simulation with a table in the forth part to show that the credibility premium obtained in this paper is better than other credibility models.

## §2. The Maximum Entropy Method

Entropy is the measure of the degree of uncertainty of an event in information theory, the greater its value, the less information we know about the event.

In 1948, Shannon published a famous paper, which established the quantitative metrics about uncertainty by entropy.

**Definition 1** Let $\xi$ be a random variable and it has $n$ possible outcomes as $A_1, A_2, \ldots, A_n$, with the probability $\mathsf{P}(\xi = A_i) = p_i$ for $i = 1, 2, \ldots, n$. Then the uncertainty of $\xi$, in other words, the information entropy is defined as follows,

$$S(\xi) = -\sum_{i=1}^{n} p_i \log p_i.$$

It is worthy to note that, for a given random event, the bigger its entropy, the more uncertainty it has. If the entropy is 0, then the event is determined.

Shannon's definition of entropy plays an important role in quantification of the degree of uncertainty. However, because this measure is related to the probability, the probability distribution is required to calculate the specific values of entropy, then the extent of the uncertainty of information can be calculated. Jaynes made a significant contribution to widely used of entropy, entropy became a reasoning tool for a simple measure through his work. He was the first people who recognized that in many probabilistic experiment, the probability is unknown in advance so that the entropy can not be calculated. But he believed that we can make use of the observation data to calculate the probability distribution.

The maximum entropy principle proposed by Jaynes provides us with such a selection criterion, "In making inference on the basis of partial information we must use that probability distribution which has maximum entropy subject to whatever is known. This is the only unbiased assignment we can make; to use any other would amount to arbitrary assumption of information which by hypothesis we do not have".

From the viewpoint of the insurance, the maximum entropy method indicates that we know nothing about the probability distribution of claims $X_i$ $(i = 1, 2, \cdots, n)$ except observation data. In addition, the probability satisfies the axiomatic conditions:

$$\sum_{i=1}^{n} p_i = 1, \qquad \text{for } p_i \geqslant 0 \ (i = 1, 2, \ldots, n),$$

and the observation data can be represented by its statistical moments (such as the mean, variance, etc.).

Then under given constraints, the problem of solving optimal probability distribution by the maximum entropy method is amount to solving the problem of conditional extreme values. Lagrange multiplier method is often used to tackle this kind of problems and determine the probability distribution.

**Definition 2** Suppose a discrete random variable $X$ take distinct values $X_1, X_2, \ldots$, $X_n$ with corresponding probabilities $\mathsf{P}(X = X_i) = p_i$ for $i = 1, 2, \ldots, n$ and constraints $g_1(\cdot), g_2(\cdot), \ldots, g_m(\cdot)$ which satisfy $\sum\limits_{i=1}^{n} g_r(x_i) = a_r$ for $r = 1, 2, \ldots, m$. Then the Lagrange multiplier which is subject to the above constraints is that

$$L = -\sum_{i=1}^{n} p_i \ln(p_i) - \sum_{r=1}^{m} \lambda_r \Big[\sum_{i=1}^{n} g_r(x_i) - a_r\Big],$$

for some $\lambda_1, \lambda_2, \ldots, \lambda_m$.

In fact, parameters $p_1, p_2, \ldots, p_n$ can be estimated by maximizing the Lagrange multiplier.

As can be seen, the more constraints of statistical moments, the more information we get, and the smaller degree of uncertainty in the solution of the problem. Thus the probability distribution obtained according to the maximum entropy method will be the one making the entropy maximum under constraints of statistical moments.

## §3. Main Results

### 3.1 The Maximum Entropy Method to the Credibility Estimation

Based on the net premium principle, we can obtain the following credibility estimation by applying the maximum entropy method.

**Theorem 3** Suppose the claims of a contract in the past $n$ years are $X_1, X_2, \ldots, X_n$, then an optimal linear nonhomogeneous estimator of the future claim $X_{n+1}$ is given by

$$\widehat{X}^*_{n+1} = \alpha^*_0 + \sum_{k=1}^{n} \alpha^*_k X_k, \tag{1}$$

which satisfies the following two conditions:

(i) minimize the expected loss function of $\mathsf{E}\big[X_{n+1} - \alpha_0 - \sum\limits_{i=1}^{n} \alpha_i X_i\big]^2$;

(ii) maximize the entropy,

with

$$\alpha^*_0 = \mathsf{E}(X_{n+1}) \Big/ \Big\{ 1 + \sum_{i=1}^{n} \exp \Big[ \lambda \sum_{j=1}^{n} \mathsf{Cov}\,(X_i, X_j) \frac{\mathsf{E}(X_{n+1})}{\mathsf{E}(X_i)} \Big] \Big\},$$

$$\alpha^*_k = \alpha^*_0 \exp \Big[ \lambda \sum_{j=1}^{n} \mathsf{Cov}\,(X_k, X_j) \frac{\mathsf{E}(X_{n+1})}{\mathsf{E}(X_k)} \Big] \Big/ \mathsf{E}(X_k),$$

where $\lambda$ is the solution of the following equation,

$$\begin{aligned} &\sum_{i=1}^{n} \exp \Big[ \lambda \sum_{j=1}^{n} \mathsf{Cov}\,(X_i, X_j) \frac{\mathsf{E}(X_{n+1})}{\mathsf{E}(X_i)} \Big] \sum_{j=1}^{n} \mathsf{Cov}\,(X_i, X_j) \frac{\mathsf{E}(X_{n+1})}{\mathsf{E}(X_i)} \\ = &\Big[ 1 + \sum_{i=1}^{n} \exp \Big[ \lambda \sum_{j=1}^{n} \mathsf{Cov}\,(X_i, X_j) \frac{\mathsf{E}(X_{n+1})}{\mathsf{E}(X_i)} \Big] \Big] \sum_{j=1}^{n} \mathsf{Cov}\,(X_{n+1}, X_j). \end{aligned}$$

**Proof** We shall first solve the minimization problem, i.e.,

$$\min \mathsf{E} \Big[ X_{n+1} - \alpha_0 - \sum_{i=1}^{n} \alpha_i X_i \Big]^2.$$

For simplicity, let

$$H = \mathsf{E} \Big[ X_{n+1} - \alpha_0 - \sum_{i=1}^{n} \alpha_i X_i \Big]^2. \tag{2}$$

Differentiate equation (2) on $\alpha_0$ and let the resulting formula equal to 0, it follows that

$$\alpha_0 = \mathsf{E}(X_{n+1}) - \sum_{i=1}^{n} \alpha_i \mathsf{E}(X_i),$$

which can be written as follows,

$$f_1(\underline{\alpha}) \triangleq \frac{\alpha_0}{\mathsf{E}(X_{n+1})} + \sum_{i=1}^{n} \alpha_i \frac{\mathsf{E}(X_i)}{\mathsf{E}(X_{n+1})} - 1 = 0,$$

with $\underline{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_n)$.

Take $\alpha_0$ into equation (2), then differentiate it on $\alpha_j$ and let the resulting formula equal to 0, we have that

$$\mathsf{Cov}\,(X_{n+1}, X_j) - \sum_{i=1}^{n} \alpha_i \mathsf{Cov}\,(X_i, X_j) = 0,$$

which is equivalent to

$$f_2(\underline{\alpha}) \triangleq \sum_{j=1}^{n} \mathsf{Cov}\,(X_{n+1}, X_j) - \sum_{j=1}^{n}\sum_{i=1}^{n} \alpha_i \mathsf{Cov}\,(X_i, X_j) = 0.$$

The corresponding Lagrange multiplier formula can be expressed in the following way,

$$\begin{aligned} L = & -\sum_{i=1}^{n} p_i \ln(p_i) - \lambda_0 f_1(\underline{\alpha}) - \lambda f_2(\underline{\alpha}) \\ = & -\frac{\alpha_0}{\mathsf{E}(X_{n+1})} \ln\Big(\frac{\alpha_0}{\mathsf{E}(X_{n+1})}\Big) - \sum_{i=1}^{n} \alpha_i \frac{\mathsf{E}(X_i)}{\mathsf{E}(X_{n+1})} \ln\Big(\alpha_i \frac{\mathsf{E}(X_i)}{\mathsf{E}(X_{n+1})}\Big) \\ & - \lambda_0\Big(\frac{\alpha_0}{\mathsf{E}(X_{n+1})} + \sum_{i=1}^{n} \alpha_i \frac{\mathsf{E}(X_i)}{\mathsf{E}(X_{n+1})} - 1\Big) \\ & - \lambda\Big[\sum_{j=1}^{n} \mathsf{Cov}\,(X_{n+1}, X_j) - \sum_{j=1}^{n}\sum_{i=1}^{n} \alpha_i \mathsf{Cov}\,(X_i, X_j)\Big]. \end{aligned} \tag{3}$$

Differentiate (3) on $\alpha_0$, it follows that

$$\alpha_0 = \mathsf{E}(X_{n+1}) \cdot \exp(-1 - \lambda_0).$$

Take the $\alpha_0$ into (3), we have

$$\begin{aligned} L = & (1+\lambda_0)\exp(-1-\lambda_0) - \sum_{i=1}^{n} \alpha_i \frac{\mathsf{E}(X_i)}{\mathsf{E}(X_{n+1})} \ln\Big(\alpha_i \frac{\mathsf{E}(X_i)}{\mathsf{E}(X_{n+1})}\Big) - \lambda_0 \exp(-1-\lambda_0) \\ & - \lambda_0 \cdot \sum_{i=1}^{n} \alpha_i \frac{\mathsf{E}(X_i)}{\mathsf{E}(X_{n+1})} + \lambda_0 - \lambda\Big[\sum_{j=1}^{n} \mathsf{Cov}\,(X_{n+1}, X_j) - \sum_{j=1}^{n}\sum_{i=1}^{n} \alpha_i \mathsf{Cov}\,(X_i, X_j)\Big]. \end{aligned} \tag{4}$$

Similarly as above, differentiate (4) on $\alpha_k$, and let the resulting formula equal to 0, we obtain

$$\alpha_k = \frac{\mathsf{E}(X_{n+1})}{\mathsf{E}(X_k)} \exp\Big[\lambda \mathsf{Cov}\,(X_k, n\overline{X}) \frac{\mathsf{E}(X_{n+1})}{\mathsf{E}(X_k)} - 1 - \lambda_0\Big].$$

Take $\alpha_k$ into (4), then

$$L = \exp[-1-\lambda_0] + \lambda_0 - \lambda \mathsf{Cov}\,(X_{n+1}, n\overline{X}) + \sum_{i=1}^{n} \exp\Big[\lambda \mathsf{Cov}\,(X_i, n\overline{X}) \frac{\mathsf{E}(X_{n+1})}{\mathsf{E}(X_i)} - 1 - \lambda_0\Big]. \tag{5}$$

Tackle the equation (5) with respect to $\lambda_0$ with the same operations as above, it follows that,

$$\lambda_0 = \ln\Big[1 + \sum_{i=1}^{n} \exp\Big(\lambda \mathsf{Cov}\,(X_i, n\overline{X}) \frac{\mathsf{E}(X_{n+1})}{\mathsf{E}(X_i)}\Big)\Big] - 1,$$

and then

$$L = \ln\Big[1 + \sum_{i=1}^{n} \exp\Big(\lambda \mathsf{Cov}\,(X_i, n\overline{X}) \frac{\mathsf{E}(X_{n+1})}{\mathsf{E}(X_i)}\Big)\Big] - \lambda \mathsf{Cov}\,(X_{n+1}, n\overline{X}).$$

Differentiate the above equation on $\lambda$, and let the resulting formula equal to 0, and after a series of simple calculation, an equation about $\lambda$ can be presented as follows,

$$\begin{aligned}&\sum_{i=1}^{n}\exp\left(\lambda\mathsf{Cov}\,(X_i,n\overline{X})\frac{\mathsf{E}(X_{n+1})}{\mathsf{E}(X_i)}\right)\cdot\mathsf{Cov}\,(X_i,n\overline{X})\frac{\mathsf{E}(X_{n+1})}{\mathsf{E}(X_i)}\\ =&\left[1+\sum_{i=1}^{n}\exp\left(\lambda\mathsf{Cov}\,(X_i,n\overline{X})\frac{\mathsf{E}(X_{n+1})}{\mathsf{E}(X_i)}\right)\right]\cdot\mathsf{Cov}\,(X_{n+1},n\overline{X}).\end{aligned}\tag{6}$$

We consider $\lambda$ as the solution of equation (6), and solve $\alpha_0$ and $\alpha_k$, respectively.

$$\begin{aligned}\alpha_0^* &= \mathsf{E}(X_{n+1})\Big/\left\{1+\sum_{i=1}^{n}\exp\left[\lambda\mathsf{Cov}\,(X_i,n\overline{X})\frac{\mathsf{E}(X_{n+1})}{\mathsf{E}(X_i)}\right]\right\};\\ \alpha_k^* &= \alpha_0^*\cdot\exp\left[\lambda\mathsf{Cov}\,(X_k,n\overline{X})\frac{\mathsf{E}(X_{n+1})}{\mathsf{E}(X_k)}\right]\Big/\mathsf{E}(X_k).\end{aligned}$$

Consequently,

$$\widehat{X}_{n+1}^* = \alpha_0^* + \sum_{k=1}^{n}\alpha_k^* X_k, \tag{7}$$

as desired. $\square$

**Remark 4** Given risk parameter $\theta$, suppose $X_i$ $(i=1,2,\dots,n)$ are independent and identically distributed according to an exponential family of distribution. Moreover, suppose that $\pi$ is a conjugate prior distribution. Then, under the square-error loss function, the Bayes premium can be written in the form of credibility premium formula. Payandeh Najafabadi et al.[22] generalized this situation in the following way: when claims $X_i$ $(i=1,2,\dots,n)$ have been distributed according to the log-concave distributions, the Bayes premium can also be expressed in the form of the linear credibility formula by the MEM. The present paper is different from [22], we make no distribution assumption about claims $X_i$ in this paper, under this condition, the estimation of the premium $\widehat{X}_{n+1}$ is limited to some linear function class which is similar to the classical credibility theory and combined with the MEM to calculate the optimal premium $\widehat{X}_{n+1}^*$.

### 3.2 Applications

Claims for different years are given the same weight in classical Bühlmann credibility model. However, this assumption is clearly unreasonable in practice. Therefore, the time effect is ought to be considered. We refer to [14] and [23] in this field. According to Theorem 3 and given a proper dependent structure, we can prove that the sequence $\{\alpha_k^*\}_{1\leqslant k\leqslant n}$ is nondecreasing.

We first give some assumptions:

**Assumption 5** Suppose random variables $X_1, X_2, \ldots, X_n$ have their own risk parameters $\Theta_1, \Theta_2, \ldots, \Theta_n$, respectively. Given time effect $\Theta$, suppose that $X_1, X_2, \ldots, X_n$ are mutually independent and have the same distribution. Let

$$\mathsf{E}(X_i \,|\, \theta_i) = \mu(\theta_i); \qquad \mathsf{Var}\,(X_i \,|\, \theta_i) = \sigma^2(\theta_i).$$

**Assumption 6** The distribution of risk parameters $\Theta_i$ is $\pi_i(\theta)$, define structure parameters as follows,

$$\mathsf{E}[\mu(\theta_i)] = \mu; \qquad \mathsf{E}[\sigma^2(\theta_i)] = \sigma^2; \qquad \mathsf{Cov}\,[\mu(\theta_i), \mu(\theta_j)] = \gamma(\theta, |i-j|),$$

where $\gamma(\theta, t)$ is a non-increasing function with respect to $t$. For example, we can take $\gamma(\theta, t) = \rho^t \cdot \tau^2$, $0 < \rho < 1$.

**Corollary 7** Under the Assumptions 5 and 6, the sequence $\{\alpha_k^*; k = 1, 2, \ldots, n\}$ satisfies the following inequalities:

(i) If $\lambda > 0$, $k \leqslant [n/2]$, then $\alpha_1^* \leqslant \alpha_2^* \leqslant \cdots \leqslant \alpha_{[n/2]-1}^* \leqslant \alpha_{[n/2]}^* = \alpha_{[n/2]+1}^* = \cdots = \alpha_n^*$.

(ii) If $\lambda < 0$, $k \geqslant [n/2]$, then $\alpha_1^* = \alpha_2^* = \cdots = \alpha_{[n/2]-1}^* \leqslant \alpha_{[n/2]}^* \leqslant \alpha_{[n/2]+1}^* \leqslant \cdots \leqslant \alpha_n^*$.

**Proof** We only give the proof of the first part, and that of the second part is omitted as its proof is similar to the first part.

Define a sign function Sign as follows,

$$\mathrm{Sign}(x) = \begin{cases} 1 & \text{if } x > 0; \\ 0 & \text{if } x = 0; \\ -1 & \text{if } x < 0. \end{cases}$$

If $\lambda > 0$, $k \leqslant [n/2]$,

$$\begin{aligned} \mathrm{Sign}\Big(\frac{\alpha_{k+1}^*}{\alpha_k^*} - 1\Big) &= \mathrm{Sign}\Big(\frac{\exp[\lambda \mathsf{Cov}\,(X_{k+1}, n\overline{X})(\mathsf{E}(X_{n+1})/\mathsf{E}(X_{k+1}))]}{\exp[\lambda \mathsf{Cov}\,(X_k, n\overline{X})(\mathsf{E}(X_{n+1})/\mathsf{E}(X_k))]} - 1\Big) \\ &= \mathrm{Sign}(\exp[\lambda \mathsf{Cov}\,(X_{k+1} - X_k, n\overline{X})] - 1) \\ &= \mathrm{Sign}[\mathsf{Cov}\,(X_{k+1} - X_k, n\overline{X})] \\ &= \mathrm{Sign}\Big[\sum_{i=1}^n \mathsf{Cov}\,(X_{k+1}, X_i) - \sum_{i=1}^n \mathsf{Cov}\,(X_k, X_i)\Big]. \end{aligned} \tag{8}$$

Consider that

$$\sum_{i=1}^n \mathsf{Cov}\,(X_i, X_j) = \sum_{i=1}^n \mathsf{Cov}\,(\mu(\theta_i), \mu(\theta_j)) + \sigma^2 = \sum_{i=1}^n \gamma(\theta, |i-j|) + \sigma^2.$$

It follows that

$$\begin{aligned}\operatorname{Sign}\Big(\frac{\alpha^*_{k+1}}{\alpha^*_k}-1\Big) &= \operatorname{Sign}\Big[\sum_{i=1}^{n}\gamma(\theta,|i-k-1|)-\sum_{i=1}^{n}\gamma(\theta,|i-k|)\Big] \\ &= \operatorname{Sign}[\gamma(\theta,k)-\gamma(\theta,n-k)]. \end{aligned} \tag{9}$$

Notice that $\gamma(\theta,k)$ is a non-increasing function. Thus $\operatorname{Sign}(\alpha^*_{k+1}/\alpha^*_k-1)\geqslant 0$ if $k\leqslant[n/2]$. □

**Remark 8** If we consider a special time effect in the model of Theorem 3, the conclusion of Corollary 7 illustrates that the new claims have more weight than the old claims, which is consistent with reality.

As is known to all that the credibility estimator with geometric weights has the similar property mentioned as above. Estimator in this form have previously been studied by Gerber and Jones [24,25] and Sundt [26], under the assumption of Bühlmann model, the estimator can be calculated recursively by

$$\widehat{X}_{n+1}=(1-c)\cdot X_n+c\cdot\widehat{X}_n;\qquad \widehat{X}_1=\mu. \tag{10}$$

According to this recursion, we can get

$$\widehat{X}_{n+1}=(1-c)\cdot\sum_{k=1}^{n}c^{n-k}X_k+c^n\mu, \tag{11}$$

with $\widehat{X}_1=\mu$, $c=\tau^2/(\tau^2+\sigma^2)$.

Actually, the credibility estimator with geometric weights is derived via the recursive model, that is, the credibility premium $\widehat{X}_{n+1}$ is the weighted average of the claims $X_n$ and the credibility premium $\widehat{X}_n$. However, this assumption is not required in the present paper.

When a credibility model is established, we often consider the mutual independence of random variables. The following content is a special case of Theorem 3 in which the assumption of independence is considered.

**Assumption 9** Given the risk parameter $\theta$, suppose random variables $X_1,X_2,\ldots,X_n$ are mutually independent and identically distributed. And

$$\mathsf{E}(X_i\,|\,\theta)=\mu(\theta);\qquad \mathsf{Var}\,(X_i\,|\,\theta)=\sigma^2(\theta).$$

**Assumption 10** The distribution of the risk parameter $\Theta$ is $\pi(\theta)$, and define the structure parameters as follows,

$$\mathsf{E}[\mu(\theta)]=\mu;\qquad \mathsf{E}[\sigma^2(\theta)]=\sigma^2;\qquad \mathsf{Var}\,[\mu(\theta)]=\tau^2.$$

**Corollary 11** Under the Assumptions 9 and 10, the optimal linear nonhomogeneous estimates $\widehat{X}^*_{n+1}$ can be written as:

$$\widehat{X}^*_{n+1} = (1 - Z^*)\mu + Z^*\overline{X},$$

with $Z^* = n\tau^2/(n\tau^2 + \sigma^2)$, $\overline{X} = n^{-1}\sum\limits_{i=1}^{n} X_i$.

**Proof** Firstly, we should find the solution of $\lambda$.

Since

$$\mathsf{Cov}\,(X_i, n\overline{X}) = n\tau^2 + \sigma^2; \qquad \mathsf{Cov}\,(X_{n+1}, n\overline{X}) = n\tau^2.$$

Take the above two formulas into (6), it follows that

$$\lambda = \frac{1}{n\tau^2 + \sigma^2}\ln\Big(\frac{\tau^2}{\sigma^2}\Big).$$

Then we can get

$$\alpha_0^* = \frac{\sigma^2}{n\tau^2 + \sigma^2}\mu; \qquad \alpha_k^* = \frac{\tau^2}{n\tau^2 + \sigma^2}.$$

$\widehat{X}^*_{n+1} = \alpha_0^* + \sum\limits_{k=1}^{n} \alpha_k^* X_k = (1 - Z^*)\mu + Z^*\overline{X}$, with $Z^* = n\tau^2/(n\tau^2 + \sigma^2)$, $\overline{X} = n^{-1}\sum\limits_{k=1}^{n} X_k$. So this completes the proof. $\square$

**Remark 12** Obviously, Corollary 11 is a special case of Theorem 3, the result coincides with the classical Bühlmann credibility estimator, Theorem 3 can thus be seen as an extension of classical credibility estimator. Note that Corollary 11 is the same as classical Bühlmann credibility estimators, so we omit the problem of parameter estimation here.

**Remark 13** The classical credibility theory is derived by minimizing the expected loss function under the assumption that the claims $X_i$ $(i = 1, 2, \ldots, n)$ are independent and identically distributed given $\theta$. What's more, the coefficient $\alpha_i$ $(i = 1, 2, \ldots, n)$ calculated is equal to each other. However, we cannot obtain the coefficient $\alpha_i$ by just minimizing the expected loss function because no assumption about the distribution of $X_i$ is made in this paper, so the MEM is used to optimize it for the second time under the condition of minimizing the expected loss function and then the coefficient $\alpha_i$ is derived. In addition, the classical credibility theory is proved as a special case of Theorem 3 in Corollary 11, which illustrates that using the MEM does not affect minimizing the expected loss function.

## §4. Numerical Simulation

We give a simulation study to compare the credibility premium in our paper with classical credibility premium, the credibility estimator with geometric weights and the credibility premium in [22], respectively.

We assume that the claims in the $i$th year are $X_i \sim \mathrm{N}(\Theta_i, \sigma^2)$, for $i = 1, 2, \ldots, n$, $\Theta = (\Theta_1, \Theta_2, \ldots, \Theta_n)' \sim \mathrm{N}(\mu \mathbf{1}_n, \tau^2 \cdot \Sigma)$, where $\Sigma$ is defined in the following,

$$\Sigma = \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-2} & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-3} & \rho^{n-2} \\ \vdots & & & \ddots & & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & \rho & 1 \end{pmatrix}.$$

Obviously, the correlation between two claims decreases as the time distance between them is large.

First we assume that $\rho = 0.3$, $\tau^2 = 2.5$, $\sigma^2 = 8.1$, $\mu = 0.7$. The random values $X_i$ $(i = 1, 2, \ldots, n+1)$ is generated according to the distribution assumptions above, where $X_i$ $(i = 1, 2, \ldots, n)$ are taken as the sample values, and $X_{n+1}$ is seen as the real value to be estimated. In addition, the mean square error between $X_{n+1}$ and $\widehat{X}_{n+1}$ is taken as the evaluation criterion in this paper. For convenient expression, we use $\mathrm{MSE}^1$, $\mathrm{MSE}^2$, $\mathrm{MSE}^3$ and $\mathrm{MSE}^4$ to represent for the mean square error in Theorem 3, the classical credibility premium, the credibility estimator with geometric weights and the credibility premium in [22], respectively.

We consider $n = 5, 10, 20$, respectively, finish the simulation 10 000 times and list the corresponding results as follows.

**Table 1 Results of numerical simulation**

| MSE | $\mathrm{MSE}^1$ | $\mathrm{MSE}^2$ | $\mathrm{MSE}^3$ | $\mathrm{MSE}^4$ |
|---|---|---|---|---|
| $n = 5$ | 68.3697 | 73.3302 | 109.3460 | 79.7394 |
| $n = 10$ | 67.9487 | 71.4389 | 108.1630 | 79.0551 |
| $n = 20$ | 66.5341 | 69.0905 | 106.3300 | 78.1634 |

As we can see, the MSE in these four models become smaller when the year $n$ is bigger. Furthermore, the $\mathrm{MSE}^1$ in our model is smaller than the other three models. The $\mathrm{MSE}^2$ of the classical credibility theory is bigger than the $\mathrm{MSE}^1$ because all claims $X_i$ $(i = 1, 2, \ldots, n)$ have the same weight in the classical credibility theory. Though the credibility estimator with geometric weights has the property that new claims have more weight than the old ones, the $\mathrm{MSE}^3$ of the credibility estimator with geometric weights is the biggest because it is derived under the certain assumption that the credibility premium $\widehat{X}_{n+1}$ is the weighted average of the claims $X_n$ and the credibility premium $\widehat{X}_n$. Finally, the credibility premium in [22] is derived via the MEM, and it also has the characteristic that new claims have more weight than the old ones, but it is obtained under the assumption that the claims $X_i$ $(i = 1, 2, \ldots, n)$ have been distributed according to the logconcave distributions. Therefore, the MSE$^4$ of the credibility premium in [22] is bigger than that of the model in Theorem 3. In conclusion, the credibility premium in Theorem 3 is better if we take MSE as the evaluation criterion.

## References


[1] Bühlmann H. Experience rating and credibility [J]. *ASTIN Bull.*, 1967, **4(3)**: 199–207.

[2] Bühlmann H, Straub E. Glaubwürdigkeit für Schadensätze [J]. *Mitt. Ver. Schweiz. Ver.*, 1970, **70**: 111–133.

[3] Asmussen S. *Ruin Probabilities* [M]. Singaproe: World Scientific, 2000.

[4] Gerber H U. *An Introduction to Mathematical Risk Theory* (*S.S. Huebner Foundation Monograph Series, No. 8*) [M]. Philadelphia: Irwin, 1979.

[5] Gómez-Déniz E. On the use of the weighted balanced loss function to obtain credibility premiums [C] // *International Conference on Mathematical and Statistical Modeling in Honor of Enrique Castillo. June 28-30, 2006.* 2006: 1–12.

[6] Gómez-Déniz, E. A generalization of the credibility theory obtained by using the weighted balanced loss function [J]. *Insurance Math. Econom.*, 2008, **42(2)**: 850–854.

[7] Wen L M, Wu X Y. Experience rating under the exponential premium principle [J]. *Sci. Sin. Math.*, 2011, **41(10)**: 861–876. (in Chinese)

[8] Wen L M, Lin X, Wang J L. The credibility model under balance loss function [J]. *Chinese J. Appl. Probab. Statist.*, 2009, **25(5)**: 553–560. (in Chinese)

[9] Wen L M, Mei G P. The credibility estimators of risk premium under a new type of generalized weighted premium principle [J]. *Acta Math. Appl. Sin.*, 2013, **36(2)**: 257–268. (in Chinese)

[10] Frees E W, Carriére J F, Valdez E. Annuity valuation with dependent mortality [J]. *J. Risk Insur.*, 1996, **63(2)**: 229–261.

[11] Denuit M, Dhaene J, Goovaerts M, et al. *Actuarial Theory for Dependent Risks: Measures, Orders and Models* [M]. Chichester, UK: Wiley, 2005.

[12] Denuit M, Maréchal X, Pitrebois S, et al, *Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems* [M]. Chichester, UK: Wiley, 2007.

[13] Wen L M, Gong H L, Wang J L. The Bühlmann credibility models with dependent risk structure [J]. *Acta Math. Appl. Sin.*, 2010, **33(4)**: 732–740. (in Chinese)

[14] Zheng D, Zhang Y, Wen L M. The credibility models with time changeable effects [J]. *J. Jiangxi Norm. Univ.* (*Nat. Sci.*), 2012, **36(3)**: 249–252. (in Chinese)

[15] Teng Y, Wu L J. Credibility premiums with double common effects obtained under the exponential premium principle [J]. *Appl. Math. J. Chinese Univ. Ser. A.*, 2013, **28(4)**: 417–423. (in Chinese)

[16] Jaynes E T. Information theory and statistical mechanics [J]. *Phys. Rev.*, 1957, **106(4)**: 620–630.

[17] Kullback S. *Information Theory and Statistics* [M]. New York: Wiley, 1959.

[18] Jessop A. Entropy in multiattribute problems [J]. *J. Multi-Crit. Decis. Anal.*, 1999, **8(2)**: 61–70.

[19] Abbas A E. Maximum entropy utility [J]. *Oper. Res.*, 2006, **54(2)**: 277–290.

[20] Landsman Z, Makov U E. Credibility evaluation for the exponential dispersion family [J]. *Insurance Math. Econom.*, 1999, **24(1-2)**: 23–29.

[21] Darooneh A H. Utility function from maximum entropy principle [J]. *Entropy*, 2006, **8(1)**: 18–24.

[22] Payandeh Najafabadi A T, Hatami H, Omidi Najafabadi M. A maximum-entropy approach to the linear credibility formula [J]. *Insurance Math. Econom.*, 2012, **51(1)**: 216–221.

[23] Atanasiu V. Application of the credibility theory based on important mathematical properties [J]. *Appl. Sci.* (*APPS*), 2013, **15**: 13–29.

[24] Gerber H U, Jones D A. Credibility formulae with geometric weights [C] // *Proceedings of the Business and Economic Section, American Statistical Association*. Washington: American Statistical Association, 1973: 229–230.

[25] Gerber H U, Jones D. Credibility formulas of the updating type [J]. *Trans. Society Actuaries*, 1975, **27**: 31–52.

[26] Sundt B. Credibility and old estimates [J]. *Insurance Math. Econom.*, 1987, **6(3)**: 189–194.

# 最大熵方法下的信度估计

胡莹莹　吴黎军　孙　毅

(新疆大学数学与系统科学学院, 乌鲁木齐, 830046)

**摘　要:** 在经典的信度理论中, 计算信度保费时除了要知道历史索赔数据, 还要对$X_i\,(i=1,2,\ldots,n)$的分布做出一定假设. 本文在索赔数据$X_i\,(i=1,2,\ldots,n)$的分布信息完全未知时, 运用最大熵的方法给出了相应的信度保费. 并且在一定的假设条件下得到两个性质, 即: 对年份较近的索赔比年份较远的索赔赋予更大的权重; 经典的信度保费是本文信度保费的一种特殊情况. 最后进行数值模拟, 结果表明: 在均方误差标准下, 本文的结果优于其它三个模型.

**关键词:** 信度估计; 熵; 最大熵方法; 拉格朗日乘子

**中图分类号:** O212.9