Available online at www.sciencedirect.com
SCIENCE DIRECT®
Insurance: Mathematics and Economics 33 (2003) 585–593
INSURANCE MATHEMATICS & ECONOMICS
www.elsevier.com/locate/econbase

Semiparametric credibility ratemaking using a piecewise linear prior[☆]

Xiaowei Huang[a], Lixin Song[a], Yanchun Liang[b,*]

[a] College of Mathematics, Jilin University, Changchun 130012, PR China
[b] College of Computer Science and Technology, Jilin University, Changchun 130012, PR China

Received June 2002; received in revised form 3 August 2003; accepted 25 August 2003

Abstract

A practical estimate on the credibility formula is presented, where a piecewise linear function is taken as the approximation of the prior distribution and applied to the credibility theory. The convergence of the approximation is analyzed. Simulation results for the lognormal–lognormal mixture show the effectiveness of the proposed estimate on the credibility.

Keywords: Piecewise linear function; Credibility theory; Bayesian estimation

1. Introduction

One of the important issues for actuaries to handle is how to analyze various losses in performing the insurance. It is essential to determine the appropriate premiums for policyholders. The premiums consist of net and gross parts. The net premiums are that need to be determined mainly. An appropriate premium is often needed to be estimated. The appropriate premium means that it should be large enough for the insurance company to remain the solvent and should be lower enough to enable the insurance product to be competitive.

If the insurer knows the exact loss distribution of a risk, then the equitable net premium to be charged is the expectation of the loss distribution. It is an important issue for the actuaries to estimate the loss distribution. However, in most cases, the exact loss distributions are unknown because risks are heterogeneous. The insurance losses with different risks follow different loss distributions. The only information on the individual policyholder or the classes of the homogeneous risks is the historical claim data or the theoretical considerations. Therefore, the net premium is the conditional expectation of the future claims under a given prior claim data for the risk. The credibility ratemaking in the credibility theory is to deal with this kind of problem. The main purposes of the credibility ratemaking are to seek the estimation on the conditional mean under the given risk and take the obtained estimation as a credibility formula for the policyholders. For the detailed description on the above discussion, please refer to Dannenburg et al. (1996) and Willmot (1994).

☆ Supported by the Natural Science Foundation of PR China (grant no. 10271049) and the Key Project of Ministry of Education of PR China (grant no. 02090).
* Corresponding author. Tel.: +86-431-849-9879; fax: +86-431-516-6063.
E-mail addresses: ycliang@email.jlu.edu.cn, ycliang@public.cc.jl.cn (Y. Liang).

0167-6687/$ – see front matter
doi:10.1016/j.insmatheco.2003.08.007

According to Bühlmann (1967), the most accurate estimator (in squared-error loss) is the predictive mean. If taking Bayesian analysis into consideration, the predictive mean could be calculated using the conditional distribution and the prior distribution. The disadvantage of this method is that the prior distribution should be given and the resulting model may not represent the information brought by the historical data. One method for solving this problem is to employ empirical Bayesian analysis to estimate parameters in the model by using the historical data. Young (1997) proposed a semiparametric model for the credibility ratemaking where the kernel-density estimation was used to estimate the prior distribution of the conditional mean. The resulting predictive mean occasionally diverged upward from the true predictive mean for large claims beyond the 95th percentile of the marginal distribution. To reduce this divergence, Young (2000) suggested using a loss function with a linear combination of a squared-error term and a constancy penalty term. The modified model encourages the estimator to be close to a constant for large claims, however, the divergence still exists in some extent.

In this paper, we make use of the semiparametric model proposed in Young (1997, 2000). Note that although the kernel density estimation could be used successfully to estimate the probability function, whereas there exist some limitations in the kernel estimator, such as boundary effect, bias, lack of local adaptivity, and the tendency to flatten peaks and valleys of the density function. These are the potential difficulties in using the kernel-density-based estimation method. Considering that the piecewise linear function has better characteristics in simplicity and intuition than the kernel, we use the piecewise linear function as the estimate of the prior distribution and to obtain the estimates for the credibility formula. Simulations show that the proposed credibility estimates have better improvement in precision than those from existing methods. The credibility model and the piecewise linear function are introduced in Section 2. The method for estimating the prior density is proposed and the asymptotic property of the approach is analyzed in Section 3. The credibility formula of the estimator is discussed in Section 4. Numerical simulations are performed using a lognormal–lognormal mixture in Section 5.

2. Credibility model and piecewise linear function

Assume that the underlying claim per unit of exposure is a random variable denoted by Y, whose distribution function is related to a unknown parameter θ. In the spirit of Bayes, the unknown parameter θ could be viewed as a random variable. The corresponding probability density function is denoted as $\pi(\theta)$, which is also called the structure function (Bühlmann, 1970). Thus, the structure function $\pi(\theta)$ is the prior distribution of θ. Denote that the values of the realization of θ are $\theta_1, \theta_2, \dots, \theta_I$. Given $\theta = \theta_i$, the conditional probability density function of Y is denoted as $f(y|\theta_i)$.

The observed data are the average claims per unit of exposure $x_i = (x_{i1}, x_{i2}, \dots, x_{iT_i})$ with an associated exposure vector $w_i = (w_{i1}, w_{i2}, \dots, w_{iT_i})$, where x_{it} represents the arithmetic average of w_{it} claims. If a risk is a class of homogeneous policies, x_{it} is the average claim per policy and w_{it} the number of the policies in the tth policy period of the ith risk class, respectively.

Assume that θ is the conditional mean of Y, that is, $E[Y|\theta] = \theta$ holds. Assume that the parameters except for the conditional mean θ are fixed across the risks. The loss distribution of a given risk is, therefore, characterized by its conditional mean. The structure function $\pi(\theta)$ describes the function relationship that the conditional mean θ varies from risk to risk.

For the purpose of convenience in the following discussion, the two assumptions are proposed.

Assumption 1. The marginal density of Y is nonzero on $[a, b]$, except for the possibility at $y = a$ or $y = b$.

Assumption 2. The sample mean $\bar{x}$ weighted by the exposure vector w is sufficient statistic for parameter θ, and the function form of $f(y|\theta)$ is closed under averaging.

In fact, Assumption 1 means that the random variable Y and the conditional mean θ are defined on the bounded interval. Assumption 2 imposes some restrictions on the distribution of Y. Families of densities that satisfy these conditions are used commonly in actuarial science are (1) the normal density with mean θ and fixed variance σ^2, (2) the gamma density with mean $\theta = \alpha/\beta$ and fixed shape parameter α, and (3) the inverse Gaussian density with mean θ and fixed $\lambda = \theta^3/\mathrm{Var}[x|\theta]$.

According to Assumption 1 and considering that θ varies in a bounded interval $[a, b]$, we make an equal-partition $a = \theta_0^0 < \theta_1^0 < \cdots < \theta_m^0 = b$. As usual, m is determined based on the sample size I, that is, $m = \mathrm{O}(I^k)$ such that the partition satisfies $\max_{1\le j\le m}(\theta_j^0 - \theta_{j-1}^0) \le d \cdot I^k$, where d is a constant and $0 < k < 1$. For the choice of parameter m in detail, please refer to Prakasa Rao (1983).

For $1 \le j < m$, let

$$I_j(\theta) = I_{(\theta_{j-1}^0 \le \theta < \theta_j^0)} = \begin{cases} 1 & \text{if } \theta_{j-1}^0 \le \theta < \theta_j^0, \\ 0 & \text{otherwise,} \end{cases} \qquad I_m(\theta) = I_{(\theta_{m-1}^0 \le \theta \le \theta_m^0)}, \qquad c = (c_0, c_1, \dots, c_m)^{\mathrm{T}} \tag{2.1}$$

and

$$\hat{\pi}_m(\theta; c) = \sum_{j=1}^{m} \left(\frac{c_j - c_{j-1}}{\theta_j^0 - \theta_{j-1}^0}\theta - \frac{c_j\theta_{j-1}^0 - c_{j-1}\theta_j^0}{\theta_j^0 - \theta_{j-1}^0} \right) I_j(\theta). \tag{2.2}$$

Therefore, $\hat{\pi}_m(\theta; c)$ is a piecewise linear function that connects the points $(\theta_0^0, c_0), (\theta_1^0, c_1), \dots, (\theta_m^0, c_m)$ and c a parameter vector of function $\hat{\pi}_m$.

Moreover, in order to ensure that $\hat{\pi}_m(\theta; c)$ is a probability density function, the following constraints should be satisfied

(A1) $\int_a^b \hat{\pi}_m(\theta; c)\,\mathrm{d}\theta = 1$, that is, $c_0 + 2\sum_{j=1}^{m-1} c_j + c_m = 2\delta$, $\delta = \dfrac{1}{\theta_j^0 - \theta_{j-1}^0}$.

(A2) $c_j > 0$, $j = 0, 1, \dots, m$.

3. Prior density estimate using piecewise linear function

The commonly used criteria to estimate the parameter vector c is to minimize the squared-error loss function or maximize the log-likelihood function. Because the estimation of the log-likelihood function leads to seek the roots of a polynomial of degree n and is often troublesome, we use the least square estimate in this paper. The observed data are x_i and w_i, and the sample mean is $\bar{x}_i = \sum_{t=1}^{T_i} w_{it}x_{it}/w_i$. Employing the central limit theorem and Assumption 2 we can use $w_i = \sum_{t=1}^{T_i} w_{it}$ to estimate parameters θ_i $(i = 1, \dots, I)$ successively (Serfling, 1980). Thus we obtain the estimated observations of θ denoted by $(\tilde{\theta}_1, \tilde{\theta}_2, \dots, \tilde{\theta}_I)$. In view of the criteria and the constraints mentioned above, the estimate on the prior density is transformed into an optimization problem under two types of constraints, that is, equality and inequality ones. Lagrangian multiplier method and Kuhn–Tucker condition (Rockfellar, 1970) are used here to handle this constrained optimization problem.

The optimization problem is to minimize

$$\Phi(c) = \sum_{i=0}^{m-1} \left[\frac{c_i + c_{i+1}}{2\delta} - \frac{n_i}{I} \right]^2$$

under the constrained conditions (A1) and (A2).

The Lagrangian function is

$$L(c,\lambda,h)=\sum_{i=0}^{m-1}\left[\frac{c_i+c_{i+1}}{2\delta}-\frac{n_i}{I}\right]^2-\lambda\left(c_0+2\sum_{i=1}^{m-1}c_j+c_m-2\delta\right)-\sum_{i=0}^{m}h_ic_i, \tag{3.1}$$

where $n_i=\sum_{j=1}^{I} I_{\{\theta_i^0\le\tilde{\theta}_j<\theta_{i+1}^0\}}$, λ and h are the Lagrangian multipliers, where $h=(h_0,h_1,\dots,h_m)\in\mathbf{R}^{m+1}$.

Take the partial derivatives of the function $L(c,\lambda,h)$ with respect to parameters and applying the Kuhn–Tucker condition, we have

$$\begin{aligned}
&\frac{1}{2\delta}c_0+\frac{1}{2\delta}c_1-\delta\lambda-h_0\delta=\frac{n_0}{I},\\
&\frac{1}{2\delta}c_{j-1}+\frac{1}{\delta}c_j+\frac{1}{2\delta}c_{j+1}-2\delta\lambda-h_j\delta=\frac{n_j+n_{j-1}}{I},\quad j=1,\dots,m-1,\\
&\frac{1}{2\delta}c_{m-1}+\frac{1}{2\delta}c_m-\delta\lambda-h_m\delta=\frac{n_{m-1}}{I},\\
&c_0+2\sum_{j=1}^{m-1}c_j+c_m=2\delta,\qquad h_ic_i=0,\quad i=0,1,\dots,m,\qquad h_i\ge 0,\quad i=0,1,\dots,m.
\end{aligned} \tag{3.2}$$

The linear algebraic equations can be rewritten as the following matrix form:

$$\begin{pmatrix}
\frac{1}{2\delta} & \frac{1}{2\delta} & 0 & \cdots & \cdots & -\delta\\
\frac{1}{2\delta} & \frac{1}{\delta} & \frac{1}{2\delta} & 0 & \cdots & -2\delta\\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\
0 & \cdots & 0 & \frac{1}{2\delta} & \frac{1}{2\delta} & -\delta\\
1 & 2 & \cdots & 2 & 1 & 0
\end{pmatrix}
\begin{pmatrix} c_0\\ c_1\\ \vdots\\ c_m\\ \lambda\end{pmatrix}
=\begin{pmatrix}
\frac{n_0}{I}+h_0\delta\\
\frac{n_0+n_1}{I}+h_1\delta\\
\vdots\\
\frac{n_{m-1}}{I}+h_m\delta\\
2\delta
\end{pmatrix}. \tag{3.3}$$

If the initial value of h are given the solution of Eq. (3.3) could be obtained easily. The problem here is how to choose the value of h.

For any initial value of h, solve Eq. (3.3) to obtain c. If $c_i\le 0$ or $h'c\ne 0$, update h until $c_i>0$ and $h'c=0$ such that the Kuhn–Tucker constraints in Eq. (3.3) are satisfied, where "$'$" represents the transpose of a vector and $i=0,1,\dots,m$. The executed steps of the algorithm are as follows.

Define $e=(0,\dots,0,1,0,\dots,0)'$, where only the jth element is 1, $j=[(m+1)/2]$. Designate a suitable small positive ε. Assign $i=1$ and $k=0$.

Step 1. Set $h^i=(0,\dots,0,1,\dots,1)$, where i is the number of "1" in the vector.
Step 2. Solve Eq. (3.3) to find c.
Step 3. Calculate $s=h'c$.
Step 4. If $|s|<\varepsilon$ and $c_j>0$ calculate $\Phi(c)$ and go to Step 5, where j is the index to make $h_j^i=1$. Otherwise if $k<m$ set $h^i=B(h^i)$ and $k=k+1$ and go to Step 2, else if $i>[(m+1)/3]$ end, otherwise set $i=i+1$ and go to Step 1, where B is a backshift operator, that is, $B(a_1,a_2,\dots,a_n)=(a_2,\dots,a_n,a_1)$.
Step 5. If $\Phi(c)<\Phi(e)$ set $e=c$. If $i>[(m+1)/3]$ end, otherwise set $i=i+1$ and go to Step 1.

The final e is the found solution. If e is still the original the search is failure. But this case does not appear provided that the true prior distribution exists.

Define

$$\rho(g(\theta),g^m(\theta))=[E(g(\theta),g^m(\theta))^2]^{1/2},\quad g\in B_r, \tag{3.4}$$

where

$$B_r = \{g \in C^r[a,b] : -\infty < m_0 \le g(\theta) \le M < +\infty\}, \quad r = 1 \text{ or } 2.$$

Theorem 1. *For an arbitrary prior density $\pi(\theta) \in B_r$, $r = 1$ or 2, denoted by $g(\theta)$, there exist a sequence of piecewise linear functions $g^m(\theta)$, s.t.*:

$$\rho(g(\theta), g^m(\theta)) \to 0, \quad m \to \infty. \tag{3.5}$$

Proof. Take the equal-partition of $[a, b]$, we have

$$g^m(\theta) = \int_0^\theta \sum_{j=1}^m \frac{g(\theta_j) - g(\theta_{j-1})}{\theta_j - \theta_{j-1}} I_j(s)\,\mathrm{d}s + g(0).$$

Note that when $\theta_l \le \theta < \theta_{l+1}$, $0 \le l < m-1$, $g^m(\theta)$ also has the following form:

$$\begin{aligned} g^m(\theta) &= \sum_{j=1}^l \int_{\theta_{j-1}}^{\theta_j} \frac{g(\theta_j) - g(\theta_{j-1})}{\theta_j - \theta_{j-1}}\,\mathrm{d}s + g(0) + \int_{\theta_l}^{\theta} \frac{g(\theta_{l+1}) - g(\theta_l)}{\theta_{l+1} - \theta_l}\,\mathrm{d}s \\ &= g(\theta_l) + \frac{g(\theta_{l+1}) - g(\theta_l)}{\theta_{l+1} - \theta_l}(\theta - \theta_l) = \frac{g(\theta_{l+1}) - g(\theta_l)}{\theta_{l+1} - \theta_l}\theta - \frac{g(\theta_{l+1})\theta_l - g(\theta_l)\theta_{l+1}}{\theta_{l+1} - \theta_l}. \end{aligned}$$

Therefore, when $\theta \in [\theta_l, \theta_{l+1})$ we have

$$\begin{aligned} |g(\theta) - g^m(\theta)| &= \left| \int_0^\theta g^{(1)}(s)\,\mathrm{d}s + g(0) - \int_0^\theta \sum_{j=1}^m \frac{g(\theta_j) - g(\theta_{j-1})}{\theta_j - \theta_{j-1}} I_j(s)\,\mathrm{d}s - g(0) \right| \\ &= \left| \int_0^\theta \left(g^{(1)}(s) - \sum_{j=1}^m \frac{g(\theta_j) - g(\theta_{j-1})}{\theta_j - \theta_{j-1}} I_j(s) \right) \mathrm{d}s \right| \\ &= \left| \sum_{j=1}^l \int_{\theta_{j-1}}^{\theta_j} \left(g^{(1)}(s) - \frac{g(\theta_j) - g(\theta_{j-1})}{\theta_j - \theta_{j-1}} \right) \mathrm{d}s + \int_{\theta_l}^{\theta} \left(g^{(1)}(s) - \frac{g(\theta_{l+1}) - g(\theta_l)}{\theta_{l+1} - \theta_l} \right) \mathrm{d}s \right| \\ &= \left| \int_{\theta_l}^{\theta} (g^{(1)}(s) - g^{(1)}(\zeta))\,\mathrm{d}s \right|, \quad \zeta \in [\theta_l, \theta_{l+1}]. \end{aligned}$$

If $r = 1$:

$$|g(\theta) - g^m(\theta)| \le 2\|g^{(1)}\|_\infty(\theta - \theta_l) \le 2\|g^{(1)}\|_\infty(\theta_{l+1} - \theta_l) = \frac{2(b-a)}{m}\|g^{(1)}\|_\infty.$$

If $r = 2$:

$$\begin{aligned} &|g(\theta) - g^m(\theta)| \\ &\quad = \left| \int_{\theta_l}^{\theta} g^{(2)}(w)(s - \zeta)\,\mathrm{d}s \right| \quad \le \frac{1}{2}\|g^{(2)}\|_\infty (s - \zeta)^2 \Big|_{\theta_l}^{\theta} \le \frac{1}{2}\|g^{(2)}\|_\infty(\theta_{l+1} - \theta_l)^2 = \frac{(b-a)^2}{2m^2}\|g^{(2)}\|_\infty, \end{aligned}$$

which leads to the following estimate:

$$\rho(g(\theta), g^m(\theta)) \to 0, \quad m \to \infty.$$

This completes the proof. □

Theorem 1 provides the theoretical basis for us to apply the proposed piecewise linear estimate methodology to approximate the prior density. It shows that any prior density function can be approximated with any accuracy by using a piecewise linear function once the requirement on accuracy and the metric distance are designated. The more historical claim data the actuaries gain, the more accurate the piecewise prior estimate would be.

4. Estimation of the credibility formula

In the Bayesian spirit, for a given loss function $L = L(y, d(\bar{x}))$ with the future claim y and the claim predictor $d(\cdot)$, the credibility estimator $\hat{d}$ is the function that minimizes the expected loss $E[L(y, d(\bar{x}))]$, where the expected loss is found with respect to the joint density of the sample mean and the future claim. The squared-error loss criteria is used here to formulate the credibility formula, which has the following form:

$$L(y, d(\bar{x})) = (y - d(\bar{x}))^2. \tag{4.1}$$

According to the existing result in Bühlmann (1967), we know that the minimum of the expected loss is the predictive mean, which is the posterior mean of θ here. When the sample mean $\bar{x}$ is given, the predictive mean can be estimated using the formula:

$$\hat{\mu}(\bar{x}) \stackrel{\triangle}{=} \hat{E}[\theta|\bar{x}] = \int \theta \hat{\pi}(\theta|\bar{x})\,\mathrm{d}\theta. \tag{4.2}$$

The piecewise linear function $\hat{\pi}(\theta)$ is employed here to estimate the prior distribution $\pi(\theta)$, which can be written as

$$\hat{\pi}(\theta) = \sum_{j=1}^{m} \left(\frac{c_j - c_{j-1}}{\theta_j - \theta_{j-1}}\theta - \frac{c_j\theta_{j-1} - c_{j-1}\theta_j}{\theta_j - \theta_{j-1}} \right) I_j(\theta) = \sum_{j=1}^{m} (c_j\theta - d_j) I_j(\theta), \tag{4.3}$$

where

$$c_j = \frac{c_j - c_{j-1}}{\theta_j - \theta_{j-1}}, \qquad d_j = \frac{c_j\theta_{j-1} - c_{j-1}\theta_j}{\theta_j - \theta_{j-1}}.$$

Then the credibility estimator has the form:

$$\hat{E}(\theta|\bar{x}) = \frac{\int \theta f(\bar{x}|\theta)\hat{\pi}(\theta)\,\mathrm{d}\theta}{\int f(\bar{x}|\theta)\hat{\pi}(\theta)\,\mathrm{d}\theta} = \frac{\sum_{j=1}^{m} \int_{\theta_{j-1}}^{\theta_j} \theta f(\bar{x}|\theta)(c_j\theta - d_j)\,\mathrm{d}\theta}{\sum_{j=1}^{m} \int_{\theta_{j-1}}^{\theta_j} f(\bar{x}|\theta)(c_j\theta - d_j)\,\mathrm{d}\theta}. \tag{4.4}$$

In fact, the estimator approaches the true value of the expected claim with probability 1. It shows that if an actuary gets more information, that is, if the amount of the exposure $w = \sum_{i=1}^{I} w_i \to \infty$, then $\hat{\mu}(\bar{x})$ approaches the true value θ_0 for a given risk. Indeed, under a certain regularity condition, De Groot (1970) and Walke (1969) proved the following results:

$$\lim_{w\to\infty} \hat{\mu}(\bar{x}) = \lim_{w\to\infty} \frac{\int \theta f(\bar{x}|\theta)\hat{\pi}(\theta)\,\mathrm{d}\theta}{\int f(\bar{x}|\theta)\hat{\pi}(\theta)\,\mathrm{d}\theta} = \frac{\theta_0\hat{\pi}(\theta_0)}{\hat{\pi}(\theta_0)} = \theta_0$$

holds with probability 1, if

$$E(\bar{x}|\theta) = \theta, \qquad \mathrm{Var}(\bar{x}|\theta) = \frac{\mathrm{Var}(Y|\theta)}{w}$$

and the mass of the density function $f(\bar{x}|\theta)$ concentrates at the point $\bar{x} = \theta_0$. Thus, if the actuary gets more claim information for a given policyholder, the estimated expected claim approaches the true value with probability 1.

5. Simulation in a lognormal–lognormal mixture

The lognormal distribution is often used by actuaries to model the distributions of the claim intensity and the total claim. In this section, we simulate the data from the following density:

$$f(x|\theta) = \frac{1}{\sigma x\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}\left[\ln\left(\frac{x}{\theta}\right)\right]^2\right\}, \quad x > 0, \tag{5.1}$$

where $\sigma > 0$ is a known parameter and

$$\pi(\theta) = \frac{1}{\tau\theta\sqrt{2\pi}} \exp\left\{-\frac{1}{2\tau^2}\left[\ln\left(\frac{\theta}{\mu}\right)\right]^2\right\}, \quad \theta > 0, \tag{5.2}$$

where $\mu > 0$ and $\tau > 0$ are known parameters, that is, $\ln X|\theta \sim N(\ln\theta, \sigma^2)$ and $\ln\theta \sim N(\ln\mu, \tau^2)$. The marginal distribution of X is also lognormal, $\ln X \sim N(\ln\mu, \sigma^2 + \tau^2)$.

Assume that we are given individual claim data, that is, $w_{it} = 1$, for each risk i and policy period t, and $x = y$. We also assume that the claim data are given for a specific policy holder $x = (x_1, x_2, \dots, x_n) \in [0, +\infty]^n$. The posterior distribution of θ is also lognormal $\ln\theta|x \sim N(\ln\mu^*, \tau^{*2})$, where

$$\mu^* = \exp\left(\frac{\sigma^2\ln\mu + \tau^2\nu}{\sigma^2 + n\tau^2}\right), \quad \nu = \sum_{i=1}^{n}\ln(x_i), \qquad \tau^{*2} = \frac{\sigma^2\tau^2}{\sigma^2 + n\tau^2}. \tag{5.3}$$

Thus the predictive distribution of X_{n+1} is lognormal for a given x, $(\ln X_{n+1})|x \sim N(\ln\mu^*, \sigma^2 + \tau^{*2})$. It follows that the true predictive mean is a function of ν:

$$\mu(x) = E(X_{n+1}|x) = \exp\left(\frac{\sigma^2\ln\mu + \tau^2\nu}{\sigma^2 + n\tau^2} + \frac{\sigma^2(\sigma^2 + (n+1)\tau^2)}{2(\sigma^2 + n\tau^2)}\right). \tag{5.4}$$

In this paper, we simulate the data from the above-mentioned distribution with $\sigma^2 = 0.25$, $\tau^2 = 0.50$ and $\mu = 2000\,\mathrm{e}^{-0.25}$. For each simulation, we calculate 1000 values of θ. For each realization value of θ, we calculate $w_i = 6$ claims.

During the calculation the time-consuming part is to estimate the prior distribution of θ. The range of θ is taken from zero to its 99.9th percentile, that is, $\theta \in [0, 15958]$. $m = 100$ is taken corresponding to $I = 1000$. Parameters $c_0, c_1, \dots, c_m$ in the piecewise linear function are estimated using the square error loss criteria. Calculated results are shown in Fig. 1. In the figure the solid line represents the true prior distribution of θ and the dotted line represents the estimated prior distribution using the piecewise linear function.

For $n = 1$, we compare our estimated predictive mean $\hat{\mu}(x)$ with the true mean $\mu(x)$ and Young's estimated result obtained in 2000. The range of x is taken from zero to its 99.9th percentile, that is, $x \in [0, 22632]$. In Fig. 2, the solid line represents the true mean, the dotted line represents our estimated results using the piecewise linear prior distribution, and the black solid line represents Young's estimated results using the kernel estimation of prior distribution and a loss function with a penalty term. From Fig. 2 it can be seen that the estimated predictive mean obtained using the proposed approach in this paper is very close to the true mean when $x < 20{,}000$, but Young's estimated results approximate well only when $x < 6500$. Fig. 2 also shows that when previous claims are larger than 20,000, our estimator tends to diverge downward slightly. In this case, if the insured is unfortunate to have large claims in the previous policy periods, the amount of the future claims that he will be charged is less than that he should be charged to some extent, whereas the insured who has unlucky claim experience tends to be overly penalized using Young's estimator. From this point of view, we provide actuaries a credibility formula in favor of unlucky insureds, which has solved the problem proposed by Young and De Vylder (2000) to some extent.

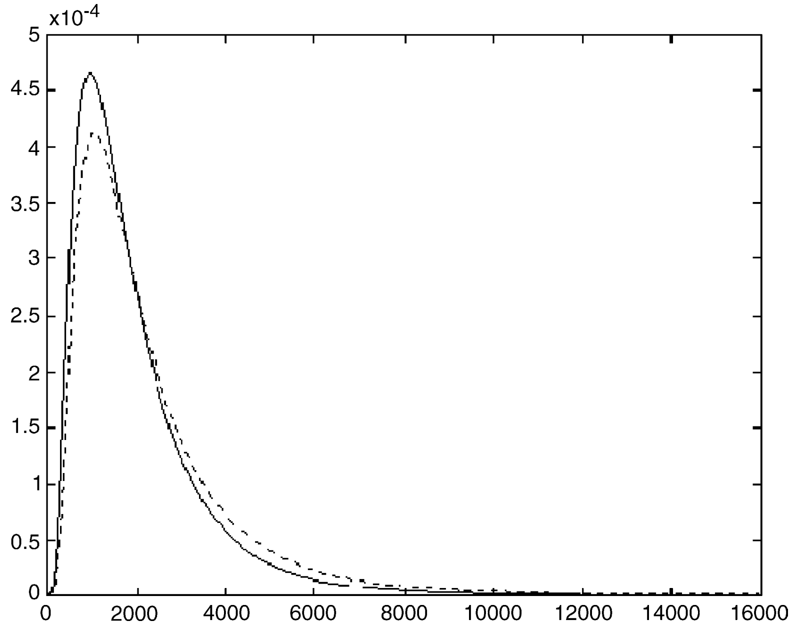

Fig. 1. Prior distribution's estimation.

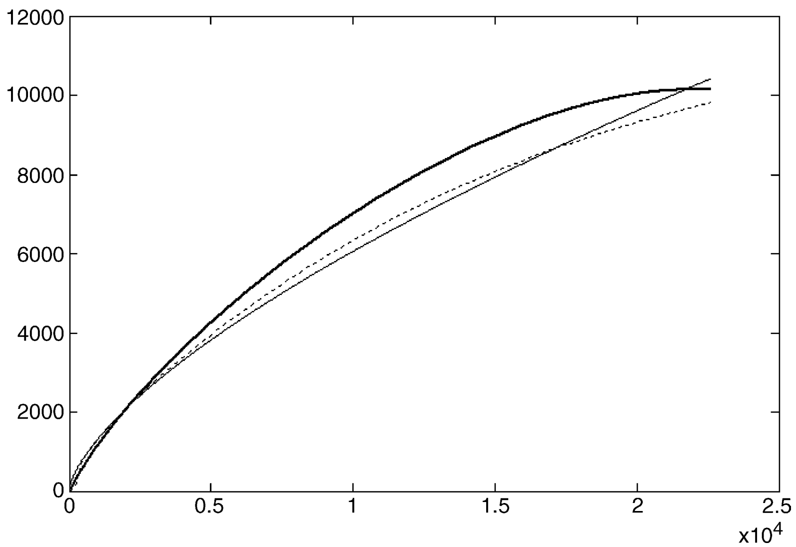

Fig. 2. Credibility formula estimators.

6. Conclusions

We propose an estimate on the credibility formula based on the piecewise linear function. Theoretical analysis and numerical simulations in the lognormal–lognormal case demonstrate that the proposed credibility estimate is fairly accurate. Numerical simulations also show that this credibility estimate is favorable to unfortunate insureds.

Acknowledgements

The authors would like to thank the anonymous referees and Prof. Rob Kaas for the very useful comments and suggestions on the earlier draft of this paper, which greatly improved our presentation.

References

Bühlmann, H., 1967. Experience rating and credibility. ASTIN Bulletin 4, 199–207.
Bühlmann, H., 1970. Mathematical Models in Risk Theory. Springer, New York.
Dannenburg, D.R., Kaas, R., Goovaerts, M.J., 1996. Practical Actuarial Credibility Models. Institute of Actuarial Science and Econometrics, University of Amsterdam, Amsterdam, The Netherlands.
De Groot, M.H., 1970. Optimal Statistical Decisions. McGraw-Hill, New York.
Prakasa Rao, B.L.S., 1983. Nonparametric Functional Estimation. Academic Press, USA.
Rockfellar, R.T., 1970. Convex Analysis. Princeton University Press, Princeton, NJ.
Serfling, R.J., 1980. Approximation Theorems of Mathematical Statistics. Wiley, New York.
Walke, A.M., 1969. On the asymptotic behaviour of posterior distributions. Journal of the Royal Statistical Society, Series B 31, 80–88.
Willmot, G.E., 1994. Introduction Credibility Theory Institute of Insurance and Person Research. University of Waterloo, Waterloo, Ont.
Young, V.R., 1997. Credibility using semiparametrics models. ASTIN Bulletin 27, 273–285.
Young, V.R., 2000. Credibility using semiparametric models and a loss function with a constancy penalty. Insurance: Mathematics and Economics 26, 151–156.
Young, V.R., De Vylder, FI.E., 2000. Credibility in favor of unlucky insureds. North American Actuarial Journal 4 (1), 107–113.