

ON BAYESIAN MIXTURE CREDIBILITY

BY

JOHN W. LAU, TAK KUEN SIU AND HAILIANG YANG

ABSTRACT

We introduce a class of Bayesian infinite mixture models first introduced by Lo (1984) to determine the credibility premium for a non-homogeneous insurance portfolio. The Bayesian infinite mixture models provide us with much flexibility in the specification of the claim distribution. We employ the sampling scheme based on a weighted Chinese restaurant process introduced in Lo et al. (1996) to estimate a Bayesian infinite mixture model from the claim data. The Bayesian sampling scheme also provides a systematic way to cluster the claim data. This can provide some insights into the risk characteristics of the policyholders. The estimated credibility premium from the Bayesian infinite mixture model can be written as a linear combination of the prior estimate and the sample mean of the claim data. Estimation results for the Bayesian mixture credibility premiums will be presented.

KEYWORDS

Credibility Theory, Bayesian Mixture Models, Infinite Mixture, Risk Characteristics, Clustering, Weighted Chinese Restaurant Process, Credibility Premium Principle, Dirichlet Process.

1. INTRODUCTION

Credibility theory provides a method to determine the premium of an insurance contract by combining two different sources of information, namely, the collective risk and the individual risk. More specifically, the pure premium of an insurance contract is determined by analyzing how much weight is given to the experience from the individual risk, which can be obtained from the claim history of the individual. It also provides the basic set up for the valuation of an insurance contract and has widely been used by actuaries to determine the premium of an insurance contract. Credibility theory has a long history in actuarial science. The origin of credibility theory goes back to the earlier works by Mowbray (1914) and Whitney (1918), in which the basic ideas of credibility and experience rating have been laid down. The idea of credibility theory has

then been formulated using the statistical ideas in an important paper by Bailey (1950). In particular, the problem of premium rating has been formulated by utilizing the language of parametric Bayesian statistics. The concept of the Bayes premium has been introduced as an estimator for the pure premium. Under some conjugate-prior assumptions on the distributions of the aggregate claim amounts and the risk parameters, a closed-form expression for the Bayes premium can be obtained. Beyond the class of conjugate-prior distributions, closed-form expressions for the Bayes premium rarely exist. In this case, one may need to resort to some numerical procedures for approximating the Bayes premium in order to estimate the pure premium.

The seminal works of Bühlmann (1967, 1970) provide a solid mathematical foundation for credibility and establish a model-based Bayesian approach for credibility by using the least-square approach. The Bühlmann credibility model employs a linear Bayes estimator, which also calls the credibility estimators, to approximate the pure premium. The linear Bayes estimator is optimal in the sense of minimizing the quadratic loss function in the Bayesian decision theory. It provides actuaries with a convenient and flexible method to approximate the pure premium without recourse some complicated numerical procedures. The Bühlmann credibility estimator for the pure premium can be expressed as a linear combination of the collective premium and the sample mean of the individual claim data. This is very easy to interpret and makes the intuition of calculating premiums by experience rating more appealing. The Bühlmann credibility model has become very popular in the actuarial community and is the milestone of the development of credibility theory. For a comprehensive discussion on various developments and methodologies on credibility, see Waters (1993) and Bühlmann and Gisler (2005).

There has been an explosive growth in the development of statistical methods and computation in the past few decades. This facilitates the use of some more computationally intensive statistical models for providing a more flexible way to model the claim data and the risk parameters. Young (1997) develops a semi-parametric credibility model by utilizing a semi-parametric mixture model to represent the insurance losses of a portfolio of risks. She explores the use of the techniques from nonparametric density estimation to estimate the prior means from the loss data and adopts the estimated model to evaluate the predictive mean of the future claims given past claims. This approach provides practitioners with flexibility in specifying the parametric distribution for each risk with unknown mean that varies across the risks. Young (1998) investigates the uncertainty in the estimated prior obtained in Young (1997) due to the randomness in the claim data and calculates the intervals for the corresponding predictive means.

In this article, we introduce a class of Bayesian infinite mixture models developed by Lo (1984) to determine the credibility premium for a non-homogeneous insurance portfolio. The class of Bayesian infinite mixture models can provide a great deal of flexibility in modelling the distributions of the aggregate claim amounts and the risk parameters without imposing stringent

parametric assumption on the distributions. The main idea of the Bayesian infinite mixture models is to represent the distribution of the observations or data as an infinite mixture of kernel density over a random distribution function with prior distribution given by a Dirichlet process. Given that the kernel density is a proper density, for example, a normal density, the mixture representation has the advantage of representing any density function. It also provides a Bayesian non-parametric specification for the distribution of the unknown model parameters. In the context of credibility theory, the class of Bayesian infinite mixture models can provide the flexibility in modelling any shape of the distribution of the aggregate claim amounts. It also provides much flexibility in modelling the prior distribution of the unknown risk parameters. In particular, the prior distribution of the unknown risk parameters is assumed to be random and follows a Dirichlet process. This is in contrast with the parametric Bayesian credibility in which the prior distribution is assumed to be known and specified by a parametric distribution. Due to the generality of the class of the Bayesian infinite mixture models, the computation of the posterior quantities poses a challenging problem. Lo (1984) is the first to provide a compact representation of a posterior quantity as a finite sum over partitions of the data. This greatly simplifies the computation of the posterior quantities and makes the Bayesian infinite mixture models more easy to implement in practical situations.

In the context of credibility theory, one can adopt the Lo approach to compute some posterior quantities by grouping the claim data associated with the same kernel density corresponding to a particular set of risk parameters to form partitions. In practical situations, one needs to sample a large number of partitions when evaluating posterior quantities. We employ a sampling procedure, which calls the weighted Chinese restaurant (WCR) process named by Lo et al. (1996) [see also Ishwaran and James (2001), MacEachern (1994), Neal (2000) and West et al. (1994)], to provide an efficient way to sample partitions. This is a natural procedure to sample partitions sequentially. Since one of the key steps for Bayesian inference is to calculate the posterior quantities, the WCR can also be used to estimate the Bayesian infinite mixture models. Lo et al. (1996) investigate the performance of the sampling procedure in density estimation and find that it performs very well. The clustering of the claim data via the sampling scheme can also provide some insights into the risk characteristics of the policyholders. The estimated credibility premium can be written as a linear combination of the prior estimate and the sample mean of the claim data. The class of Bayesian infinite mixture models provides actuaries with a convenient and flexible way to model the loss (or claim) distribution and estimate the credibility premium as the predictive mean of future claims given past claims. We shall provide both simulation and empirical studies for the Bayesian mixture credibility model and compare its performance with the Bühlmann credibility model.

The rest of this article is structured as follows: The next section presents the Bayesian infinite mixture model and the corresponding credibility premium

formula. Section 3 discusses the WCR for sampling the partitions for the claim data and the associated estimation method. We shall conduct simulation studies on the Bayesian mixture credibility model in Section 4. We shall compare the simulation results for the Bayesian mixture credibility model with those for the Bühlmann credibility model. Estimation results for the Bayesian mixture credibility premiums will be presented and discussed in Section 5. The final section summarizes the paper.

2. BAYESIAN INFINITE MIXTURE MODEL FOR CREDIBILITY PREMIUM

In this section, we present a Bayesian infinite mixture model for modelling the distribution of the aggregate claim amount of a policyholder over different years and the uncertainty of the risk characteristic of the policyholder. Then, we present the credibility premium for a non-homogeneous insurance portfolio.

2.1. Bayesian infinite mixture model for the aggregate claim amounts

First, we fix a complete probability space $(\Omega, \mathcal{F}, \mathcal{P})$, where \mathcal{P} is a real-world probability. Let X_k denote a random variable on $(\Omega, \mathcal{F}, \mathcal{P})$, which represents the aggregate claim amount of a policyholder during the k^{th} policy period (usually one year), for $k = 1, 2, \dots, n + 1$. Let \mathcal{F}_k^X denote the σ -algebra (information set) generated by the process of the aggregate claim amounts X up to and including time k . Following the exposition in Lo (1984), we present the Bayesian infinite mixture model for the probability density function of the aggregate claim amounts in the sequel.

Let U denote a Borel subset of \mathcal{R}^m and $\mathcal{B}(U)$ the σ -algebra generated by the open sets relative to U . Note that U represents the space of the risk parameters. In the context of Bayesian infinite mixture model, the letters u and λ are often used to represent the unknown parameters, which are random variables, and their realizations in order to simplify the notations. Here, we shall follow the convention of the Bayesian infinite mixture model to define the notations of the risk parameters and their realizations as small letters. Write α for a finite measure on $(U, \mathcal{B}(U))$. Let $K(x, u)$ denote a non-negative-valued kernel defined on the product space $(\mathcal{R}^+ \times U, \mathcal{B}(\mathcal{R}^+) \otimes \mathcal{B}(U))$, where $\mathcal{B}(\mathcal{R}^+)$ is the σ -algebra generated by the open sets relative to the non-negative real line \mathcal{R}^+ . We suppose that for each $u \in U$,

$$\int_{\mathcal{R}^+} K(x, u) dx = 1, \quad (2.1)$$

and for each $x \in \mathcal{R}^+$,

$$\int_U K(x, u) \alpha(du) < \infty. \quad (2.2)$$

Some typical examples of the kernel $K(x, u)$ with the support \mathcal{R}^+ for each $u \in U$ are a Poisson probability mass function with mixing intensity parameter $u = \lambda$ and a Gamma density function with a given shape parameter a and a mixing scale parameter $u = b$.

Suppose A denotes the space of distributions on $(U, \mathcal{B}(U))$ and \mathcal{A} is the σ -algebra on A generated by weak convergence. Note that A represents the space of distributions for the risk parameters. For each $G \in A$, we define a random probability density function $f(x|G)$, where $x \in \mathcal{R}^+$, as follows:

$$f(x|G) = \int_U K(x, u) G(du). \tag{2.3}$$

For a given $G \in A$, $f(x|G)$ is a well-defined probability density function by Fubini's theorem. Also, $f(x|G)$ is \mathcal{A} -measurable. This means that $f(x|G)$ is known when the random probability distribution G is given. For more details, see Lo (1984). Since $K(\cdot, u)$ is defined on the non-negative real line, so does $f(x|G)$.

Let $\mathcal{D}_K := \{f(x|G) | G \in A\}$, which represents the space of probability density functions generated by the random probability distributions in A and the kernel $K(\cdot, \cdot)$. Any continuous probability density function can be generated from the space \mathcal{D}_K by suitable choices of the kernel density $K(\cdot, \cdot)$ (see Lo, 1984). Now, we are going to specify the prior probability for the random distribution function G as a Dirichlet process. First, we give a precise definition of a Dirichlet process. Recall that α represents a finite measure on the space of distributions (A, \mathcal{A}) . A probability measure π_α defined on (A, \mathcal{A}) is said to be a Dirichlet process with a shape measure α if for each measurable partition (A_1, \dots, A_M) on A (i.e. $A_i \in \mathcal{A}$, for each $i = 1, 2, \dots, M$),

$$\mathbb{P}(\pi_\alpha(A_1), \dots, \pi_\alpha(A_M)) = \frac{\Gamma(\sum_{i=1}^M \alpha(A_i))}{\prod_{i=1}^M \Gamma(\alpha(A_i))} \prod_{i=1}^M \pi_\alpha(A_i)^{\alpha(A_i)-1}, \tag{2.4}$$

where $\sum_{i=1}^M \pi_\alpha(A_i) = 1$ and $\Gamma(\cdot)$ is the Gamma function. In other words, the vector $(\pi_\alpha(A_1), \dots, \pi_\alpha(A_M))$ is a Dirichlet random vector with parameter vector $(\alpha(A_1), \dots, \alpha(A_M))$.

Without any data, the Bayes estimate of the unknown density $f(x|G)$ is given by the weighted average of $f(x|G)$ over A with weights given by the prior probability π_α . The Bayes estimate is the "best" in the sense of minimizing the quadratic loss function in the Bayesian decision theory. It is shown in Lo (1984) that the Bayes estimate can be written as:

$$E[f(x|G)] = \int_A f(x|G) \pi_\alpha(dG) = \int_U K(x, u) \frac{\alpha(du)}{\alpha(U)}. \tag{2.5}$$

where $\alpha(U)$ is a finite positive real number and $\frac{\alpha(du)}{\alpha(U)}$ is a probability measure.

2.2. Credibility premium

In the context of Bayesian mixture credibility, we suppose that conditional on the distribution G of risk parameters, the aggregate claim amounts $X_1, X_2, \dots, X_n, X_{n+1}$ are independent and identically distributed with a common probability density function $f(x|G)$. Without given knowledge of $G, X_1, X_2, \dots, X_n, X_{n+1}$ are not independent. In the sequel, we shall first present the predictive density function of the future aggregate claim amount X_{n+1} given information \mathcal{F}_n^X generated by past and current aggregate claim amounts. We shall then evaluate the credibility premium using the predictive density function.

Following Lo (1984), we adopt the representation of the posterior distribution of G as a finite sum over partitions of the claim amount data $\{X_1, X_2, \dots, X_n\}$ to evaluate the predictive density function of X_{n+1} given \mathcal{F}_n^X . First, we present some important notations. Let n denote a positive integer and \mathbf{P} denote a partition of $\{1, 2, \dots, n\}$. Write $n(\mathbf{P})$ for the number of cells in the partition \mathbf{P} . Then, $\mathbf{P} := \{C_1, C_2, \dots, C_{n(\mathbf{P})}\}$, where C_i is the i^{th} cell in the partition. Let e_i denote the number of elements in C_i . Then, it is noted in Lo (1984) that both C_i and $e_i, i = 1, 2, \dots, n(\mathbf{P})$, depend on \mathbf{P} . Now, we define the weighted function $W(\mathbf{P})$, which plays an important role for the evaluation of the predictive density function and is defined as:

$$W(\mathbf{P}) := \frac{\phi(\mathbf{P})}{\sum_{\mathbf{P}} \phi(\mathbf{P})},$$

where the function $\phi(\mathbf{P})$ is given as follows:

$$\phi(\mathbf{P}) = \frac{\Gamma(\alpha(U))}{\Gamma(\alpha(U) + n)} \prod_{i=1}^{n(\mathbf{P})} \Gamma(e_i) \int_U \prod_{\ell \in C_i} K(X_\ell, u) \alpha(du).$$

Then, by Lo (1984), the predictive density function of X_{n+1} given \mathcal{F}_n^X is:

$$\begin{aligned} f_{X_{n+1}|\mathcal{F}_n^X}(x) &:= E(f(x|G)|\mathcal{F}_n^X) \\ &= \sum_{\mathbf{P}} \left[\sum_{i=0}^{n(\mathbf{P})} \frac{e_i}{\alpha(U) + n} \int_U K(x, u) \pi(du|C_i) \right] W(\mathbf{P}), \end{aligned}$$

where

$$\pi(du|C_i) = \frac{\prod_{\ell \in C_i} K(X_\ell, u) \alpha(du)}{\int_U \prod_{\ell \in C_i} K(X_\ell, u) \alpha(du)} \text{ for } i = 1, \dots, n(\mathbf{P}), \tag{2.6}$$

$$\pi(du|C_0) = \frac{\alpha(du)}{\alpha(U)}, \tag{2.7}$$

and $e_0 = \alpha(U)$. The credibility premium under the Bayesian infinite mixture model is evaluated as the predictive mean of X_{n+1} given \mathcal{F}_n^X , which is a Bayes

estimator for the pure premium. Let Z_n denote $\frac{\alpha(U)}{\alpha(U) + n}$. Then, the credibility premium is evaluated as:

$$\begin{aligned}
 P_c(X_{n+1} | \mathcal{F}_n^X) &:= E(X_{n+1} | \mathcal{F}_n^X) \\
 &= (1 - Z_n) \left\{ \sum_{\mathbf{P}} \left[\sum_{i=1}^{n(\mathbf{P})} \frac{e_i}{n} \left[\int_{\mathcal{X}^+} \int_U xK(x, u) \pi(du | C_i) dx \right] \right\} \mathcal{W}(\mathbf{P}) \right\} \\
 &+ Z_n \left[\int_{\mathcal{X}^+} \int_U xK(x, u) \pi(du | C_0) dx \right],
 \end{aligned} \tag{2.8}$$

which is a weighted average of the collective premium given by the prior mean of X_k and the sample average of the aggregate claim amounts with weight Z_n given on the collective premium. This is consistent with the Bayesian parametric credibility premium formula for the pure premium.

Example 2.2.1. When the kernel function $K(x, u)$ is a Gamma density function with a given shape parameter $\gamma > 0$ and mixing scale parameter $u = \theta$, let θ be Gamma(a, b). Then, the predictive density function $\int_U K(x, u) \pi(du | C_i)$ becomes an inverse Beta density. That is,

$$\begin{aligned}
 &\int_U K(x, u) \pi(du | C_i) \\
 &= \frac{\Gamma(a + e_i \gamma + \gamma)}{\Gamma(a + e_i \gamma) \Gamma(\gamma)} x^{\gamma-1} \frac{(\sum_{\ell \in C_i} X_\ell + b)^{a+e_i \gamma}}{(\sum_{\ell \in C_i} X_\ell + b + x)^{a+e_i \gamma + \gamma}}.
 \end{aligned} \tag{2.9}$$

The credibility premium is given by:

$$P_c(X_{n+1} | \mathcal{F}_n^X) = Z_n \left(\frac{b\gamma}{a-1} \right) + (1 - Z_n) \sum_{\mathbf{P}} \left[\sum_{i=1}^{n(\mathbf{P})} \frac{e_i}{n} \left(\frac{b + \sum_{\ell \in C_i} X_\ell}{\frac{a-1}{\gamma} + e_i} \right) \right] \mathcal{W}(\mathbf{P}), \tag{2.10}$$

where $a > 1, \gamma > 0$ and $b > 0$.

Example 2.2.2. When the kernel function $K(x, u)$ is a Poisson probability mass function with mixing intensity parameter $u = \lambda$, we let λ be Gamma(a, b), a Gamma distribution with shape parameter $a > 0$ and scale parameter $b > 0$. Then, the predictive density function $\int_U K(x, u) \pi(du | C_i)$ becomes a negative binomial density with parameters $(a + \sum_{\ell \in C_i} X_\ell, \frac{b+e_i}{b+e_i+1})$. That is,

$$\begin{aligned}
 &\int_U K(x, u) \pi(du | C_i) \\
 &= \frac{\Gamma(a + \sum_{\ell \in C_i} X_\ell + x)}{\Gamma(a + \sum_{\ell \in C_i} X_\ell) \Gamma(x + 1)} \left(\frac{e_i + b}{e_i + b + 1} \right)^{a + \sum_{\ell \in C_i} X_\ell} \left(\frac{1}{e_i + b + 1} \right)^x.
 \end{aligned} \tag{2.11}$$

In this case, the credibility premium is given by:

$$P_c(X_{n+1} | \mathcal{F}_n^X) = Z_n \left(\frac{a}{b}\right) + (1 - Z_n) \sum_{\mathbf{P}} \left[\sum_{i=1}^{n(\mathbf{P})} \frac{e_i}{n} \left(\frac{a + \sum_{\ell \in C_i} X_\ell}{b + e_i} \right) \right] W(\mathbf{P}). \quad (2.12)$$

Note that we shall adopt the Bayesian infinite mixture model with a Gamma density function in the simulation experiment and empirical study in Section 4 and Section 5, respectively.

3. ESTIMATION PROCEDURE BY WEIGHTED CHINESE RESTAURANT (WCR) PROCESS

In this section, we shall present the WCR process for sampling partitions and estimate the Bayesian infinite mixture credibility model. We shall employ a numerical scheme for the WCR process developed by Lo et al. (1996) to sample partitions. In particular, a Gibbs version of the WCR process is employed to generate samples from the posterior distribution of partitions $W(\mathbf{P})$. In the sequel, we first describe the Gibbs version of the WCR process.

In the context of the WCR process, we call the k^{th} -element of the set $S := \{1, 2, \dots, n\}$ the k^{th} customer. We shall present the algorithm for the Gibbs version of the WCR process as follows:

Step I: Set an initial partition $\mathbf{P}^{(0)}$ of the set S .

Step II: Determine $\mathbf{P}^{(1)}$ from the following procedures

1. For each $k = 1, \dots, n$, consider a partition $\mathbf{P}_{-k}^{(i)}$ of $\{1, 2, \dots, n\} \setminus \{k\}$ by removing customer k from $\mathbf{P}^{(0)}$
2. For each $k = 1, 2, \dots, n$, re-seat customer k in a new table or an occupied table $C_{j,-k}$ of $\mathbf{P}_{-k}^{(i)}$, for each $j = 1, 2, \dots, n(\mathbf{P}_{-k}^{(i)})$, where $n(\mathbf{P}_{-k}^{(i)})$ represents the number of cells in the partition $\mathbf{P}_{-k}^{(i)}$ according to a predictive seating rule. Customer k is assigned to a new table with a probability proportional to

$$\alpha(U) \times \int_U K(X_k, u) \frac{\alpha(du)}{\alpha(U)},$$

or to the table $C_{j,-k}$ with a probability proportional to

$$e_{j,-k} \times \int_U K(X_k, u) \pi(du | C_{j,-k}) \text{ for } j = 1, \dots, n(\mathbf{P}_{-k}^{(i)})$$

where $e_{j,-k}$ is the number of customers in $C_{j,-k}$ and

$$\pi(du | C_{j,-k}) = \frac{\prod_{\ell \in C_{j,-k}} K(X_\ell, u) \alpha(du)}{\int_U \prod_{\ell \in C_{j,-k}} K(X_\ell, u) \alpha(du)} \text{ for } j = 1, \dots, n(\mathbf{P}_{-k}^{(i)}).$$

3. Obtain $\mathbf{P}^{(1)}$ by repeating (i) and (ii) for each $k = 1, 2, \dots, n$.

Step III: Setting with $\mathbf{P}^{(1)}$, repeat the procedures in Step I and Step II iteratively and obtain a sequence of partitions $\mathbf{P}^{(1)}, \mathbf{P}^{(2)}, \dots$

In Step II of the above algorithm, $\int_U K(x, u) \pi(du | C_{j,-k})$ plays the role of the weight of the predictive density. Customer k is likely to be seated to a table with a large predictive density and a large number of existing customers $e_{j,-k}$. In Step III of the above algorithm, we repeat the above reseating procedure and obtain a sequence of partitions. Then, one can obtain a Bayesian estimate by calculating the average of the posterior quantity (i.e. the posterior expectation) evaluated at the partitions obtained from the Gibbs version of the WCR process.

4. A SIMULATION EXPERIMENT

In this section, we shall conduct a simulation experiment on the credibility premium from the Bayesian infinite mixture model with a Gamma kernel function. We shall compare the estimated credibility premium obtained from the Bayesian infinite mixture model with a Gamma kernel function with that obtained from the Bühlmann credibility model using simulated data. We consider two parametric cases of the Bühlmann credibility model, namely, the Poisson-Gamma case and the Pareto-Uniform case, in the comparison. First, we compare the estimated credibility premiums from the Bayesian infinite mixture model and the two cases of the Bühlmann credibility model using the full data set for the aggregate claim amounts. Then, we provide the comparison on the robustness and the rate of convergence of the estimated credibility premiums as the aggregate claim amounts data emerge.

First, we suppose that the prior distribution of the risk parameter θ is an uniform distribution, $Uni[0, 10]$, on the interval $[0, 10]$, and conditional on the risk parameter θ , the distribution of the aggregate claim amounts data is a Pareto distribution, $Pareto(3, \theta)$, with a given shape parameter $a_0 = 3$ and mode parameter θ . Then, we simulate 10 realizations of the risk parameter θ from $Uni[0, 10]$. Given a realization θ of the risk parameter, we simulate 500 aggregate claim amounts data X_1, X_2, \dots, X_{500} from $Pareto(3, \theta)$. We consider the assumed models for the simulation as if they were the “true” models and the simulated aggregate claim amounts data as if there were the “true” observations in our simulation experiment. In this case, the “true” pure premium is given by:

$$E(X_i | a_0 = 3, \theta) = \theta \times \frac{3}{3-1} = 1.5\theta.$$

Now, we shall calculate the estimated credibility premium P_c from the Bayesian infinite mixture model with a Gamma kernel function using the simulated data with 500 observations. We shall also compare the Bühlmann credibility premium and the Bayesian mixture credibility premium with the “true” pure premium.

We suppose that the parameter of the Dirichlet process $\alpha(dt)$ is the probability distribution function of a Gamma distribution with shape parameter $a = 10$ and scale parameter $b = 2$. This implies that $\alpha(\infty) = 1$ and that $\frac{\alpha(dt)}{\alpha(\infty)}$ is also a Gamma distribution. We further assume that $\gamma = 1$. Then, we run the sampler with 20,000 cycles in the Gibbs version of the WCR process. The first 10,000 samples will be considered the warm-up period while the last 10,000 samples will be taken to evaluate the credibility premium using the formula (2.13).

Then, we shall evaluate the estimated credibility premiums B_{PG} and B_{PU} from the Poisson-Gamma case and the Pareto-Uniform case of the Bühlmann credibility model, respectively, using the simulated data. We employ the formula in Example 16.26 of Klugman et al. (2004) to compute the Bühlmann credibility premium in the Poisson-Gamma case. We assume that the prior parameters of the Gamma distribution are $(a, b) = (10, 2)$. Then, the Bühlmann credibility premium in the Poisson-Gamma case is given by:

$$B_{PG}(X_{n+1} | \mathcal{F}_n^X) = \left(\frac{b}{n+b}\right) \frac{a}{b} + \left(\frac{n}{n+b}\right) \bar{X}_n, \tag{4.1}$$

where $\bar{X}_n := \sum_{i=1}^n X_i$ is the sample average of the aggregate claim amounts data.

Now, we consider the Bühlmann credibility model in the Pareto-Uniform case. In this case, we suppose that the prior parameters of the uniform distribution are $(L, U) = (1, 10)$ and that the shape parameter of a Pareto distribution is given as $a_0 = 3$. Then, the Bühlmann credibility premium in the Pareto-Uniform case is given by:

$$B_{PU}(X_{n+1} | \mathcal{F}_n^X) = \left(\frac{k}{n+k}\right) \frac{L+U}{2} + \left(\frac{n}{n+k}\right) \bar{X}_n, \tag{4.2}$$

where the constant factor k is:

$$k = \frac{4}{(a-1)(a-2)} \left(1 + 3 \frac{UL}{(U-L)^2}\right). \tag{4.3}$$

Table 4.1 presents the “true” pure premium, the Bühlmann credibility premiums in both the Poisson-Gamma and Pareto-Uniform cases, and the Bayesian infinite mixture credibility premium evaluated using the full simulated data, with 500 observations for each simulated path and each risk parameter.

Figure 4.2 presents the plots of the “true” pure premium, the Bühlmann credibility premiums in both the Poisson-Gamma and Pareto-Uniform cases, and the Bayesian mixture credibility premium using the simulated data in successive periods for each of the 10 risk parameters.

Based on the simulation results, the Bayesian mixture credibility premium seems to be more close to the “true” pure premium than the Bühlmann one

TABLE 4.1
CREDIBILITY PREMIUMS V.S. THE “TRUE” PURE PREMIUM

Case	1	2	3	4	5	6	7	8	9	10
θ	2.8842	7.4075	1.9461	6.5859	5.3743	3.028	3.1345	9.7325	8.6013	4.0759
True	4.3262	11.1113	2.9191	9.8789	8.0614	4.5419	4.7017	14.5988	12.9019	6.1139
Premium Principles	Estimated Premiums									
$B_{PG}(X_{n+1} \mathcal{F}_n^X)$	4.4317	11.4788	2.8907	9.804	8.0933	4.6033	4.7797	14.774	13.2637	6.1848
$B_{PU}(X_{n+1} \mathcal{F}_n^X)$	4.4353	11.4719	2.8966	9.7997	8.0914	4.6066	4.7828	14.7623	13.2543	6.1858
$P_c(X_{n+1} \mathcal{F}_n^X)$	4.3468	11.2831	2.83	9.6347	7.9508	4.5157	4.6893	14.5265	13.04	6.0724

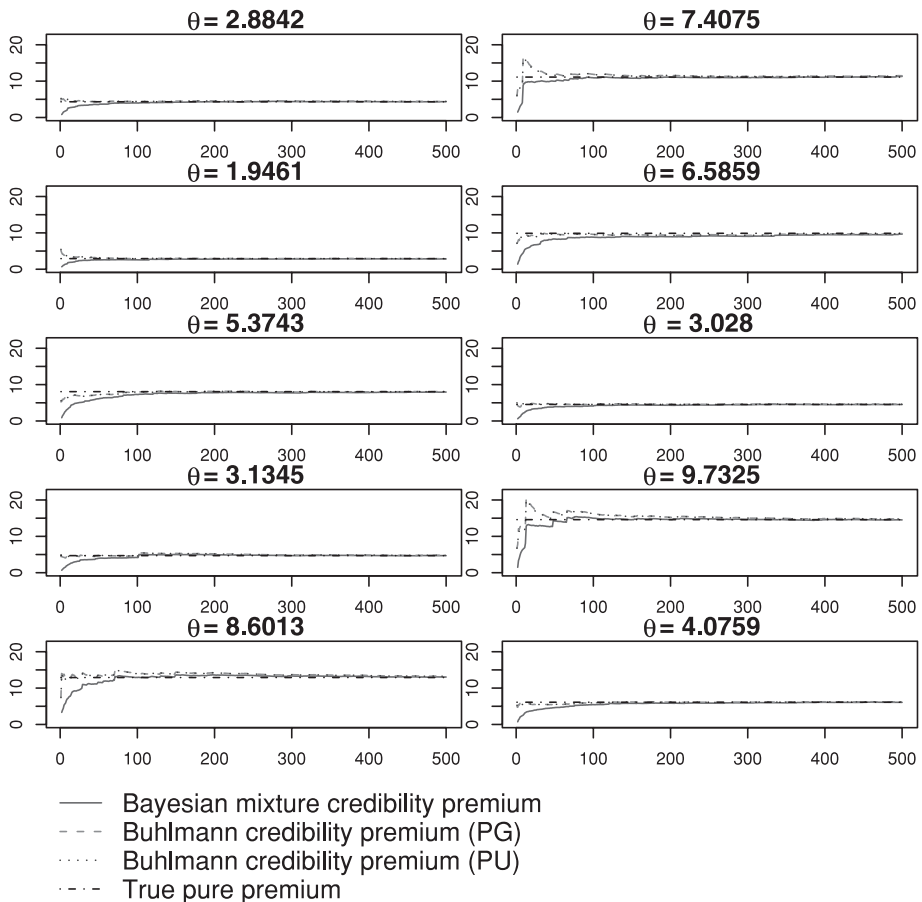


FIGURE 4.2.

in most of the cases. In addition, the Bayesian mixture credibility premium is more robust than the Bühlmann credibility premium. Even though we choose a set of prior parameters, which gives a very different estimate to the “true” pure premium, the predictive estimate will converge quickly to the corresponding “true” pure premium.

5. A REAL-DATA EXAMPLE AND COMPARISON

In this section, we shall provide a real-data example for the Bayesian mixture credibility premium and compare it with the Bühlmann credibility premium using the Danish fire insurance loss data from years 1988 to 1990. The data consist of 663 fire insurance losses in Danish Krone (DKK) and were downloaded from

<http://www.math.ethz.ch/mcneillftp/DanishData.txt>

First, we shall provide a comparison between the the credibility premium from the Bayesian infinite mixture model with a Gamma kernel function and those obtained from the Poisson-Gamma and Pareto-Uniform cases of the Bühlmann credibility model using the full set of the aggregate claim amounts data. Then, we shall make a comparison between the Bayesian mixture credibility premiums with a Gamma kernel function and the Bühlmann credibility premiums in the Poisson-Gamma and Pareto-Uniform cases in successive periods. In the parametric Bayesian case, Waters (1993) provides a study on the credibility factors in successive periods. Here, following a similar analysis with Waters (1993), we investigate the credibility factors of the Bayesian mixture credibility model and those of the Bühlmann credibility model in successive periods. We shall compute the credibility premium from the Bayesian infinite mixture model using the Gibbs version of the WCR process in Section 3. All the computations were done by C.

For evaluating the Bayesian mixture credibility premium with a Gamma kernel and the Bühlmann credibility premiums in both the Poisson-Gamma and Pareto-Uniform cases, we assume the same parameter values as those in Section 4 and adopt the corresponding methods there. Table 5.1 presents the estimated credibility premiums from the Bayesian infinite mixture model and the Bühlmann credibility models using full set of the aggregate claim amounts data with 663 observations.

TABLE 5.1
CREDIBILITY PREMIUMS USING FULL DATA

Premium Principles	Estimated Premiums (in Millions)
$B_{PG}(X_{n+1} \mathcal{F}_n^X)$	3.7091
$B_{PU}(X_{n+1} \mathcal{F}_n^X)$	3.7126
$P_c(X_{n+1} \mathcal{F}_n^X)$	3.3404

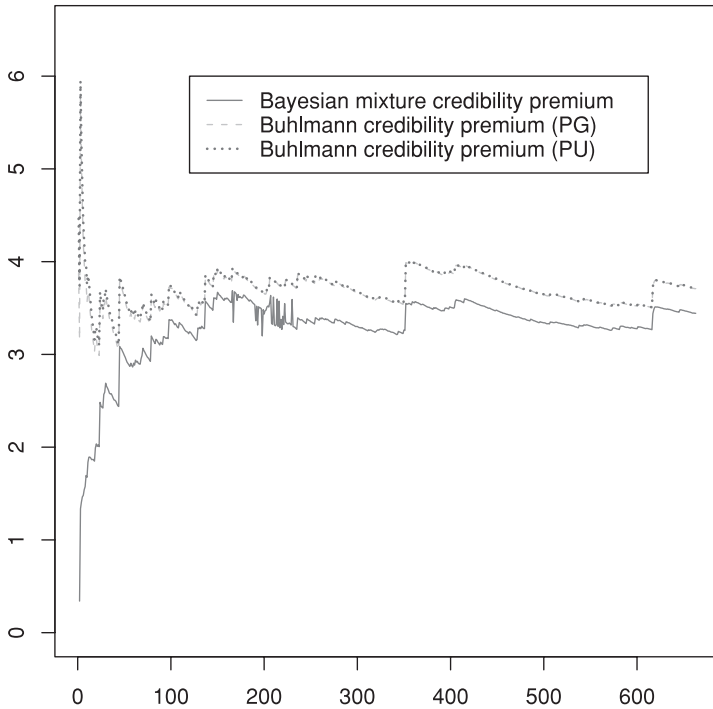


FIGURE 5.2.

From Table 5.1, we see that the Bayesian mixture credibility premiums are lower than the Bühlmann credibility premiums while the Bühlmann credibility premiums in both the Poisson-Gamma and Pareto-Uniform cases are very similar.

Figure 5.2 presents the plots of the Bayesian mixture credibility premiums and the Bühlmann credibility premiums in successive periods.

In Figure 5.2, the same as in Table 5.1, we observe that the Bayesian mixture credibility premiums are systematically lower than the Bühlmann credibility premiums in successive periods while the Bühlmann credibility premiums in both the Poisson-Gamma and Pareto-Uniform cases are very close to each other in successive periods.

Figure 5.3 displays the plots of the credibility factors from the Bayesian infinite mixture model and those from the Bühlmann credibility models in successive periods.

From Figure 5.3, we observe that the Bayesian mixture credibility factor converges more quickly to the level one compared with the Bühlmann credibility factors in both the Poisson-Gamma and Pareto-Uniform cases as data emerge. The convergence of the Bühlmann credibility factors to the level one in the two cases are very similar.

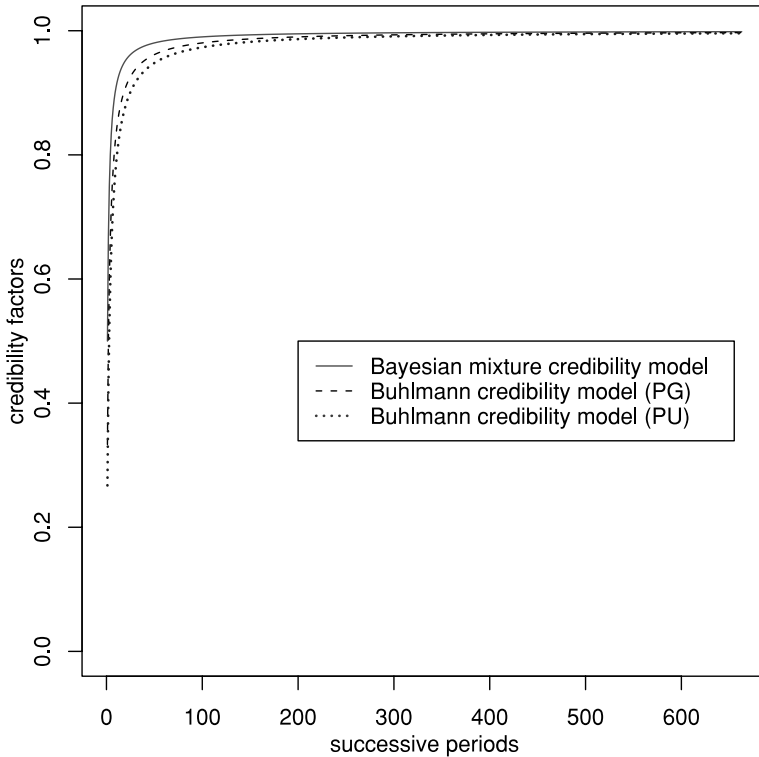


FIGURE 5.3.

6. FUTURE RESEARCH

Comparing with the parametric Bayesian credibility theory, the Bayesian mixture credibility theory is a more flexible and general approach for determining the credibility premium of a non-homogeneous insurance portfolio. The robustness analysis with respect to the prior processes and the method for choosing the shape parameter of the Dirichlet prior process when the sample size is small also represent interesting research problems. It seems that the classical density estimation can be applied to estimate the shape parameter of the Dirichlet prior process. Some techniques in neural network may provide some insights in developing efficient estimation methods for the shape parameter of the Dirichlet prior process. Formulas for the credibility premiums involving the median, quantiles and higher moments of the predictive distribution can be obtained in the Bayesian infinite mixture model. This provides us with a convenient and flexible way to investigate and develop other credibility premium principles, such as the premium principle involving the first four cumulations by Ramsay (1994), the risk-adjusted credibility premiums with distorted

probabilities by Wang and Young (1998), the scenario-based premiums in Siu and Yang (1999), and others.

ACKNOWLEDGMENTS

The authors would like to thank the referee for careful reading of the paper and many helpful comments and suggestions, Professor Albert Lo for providing us with his lecture note for Bayesian nonparametric statistical methods and related topics, Professor Howard Waters for his discussion on credibility theory and providing us with his note on credibility theory, and Professor Angus Macdonald for his discussion on some data set on insurance claims. This work was supported by Research Grants Council of HKSAR (Project No: HKU 7239/04H).

REFERENCES

- BAILEY, A. (1950) Credibility procedures. *Proceedings of the Casualty Actuarial Society*, XXXVII, 7-23 and 94-115.
- BÜHLMANN, H. (1967) Experience rating and credibility. *ASTIN Bulletin*, **4**, 199-207.
- BÜHLMANN, H. (1970) *Mathematical Methods in Risk Theory*. New York: Springer-Verlag.
- BÜHLMANN, H. and GISLER, A. (2005) *A Course in Credibility Theory and Its Applications*. New York: Springer Verlag.
- ISHWARAN, H. and JAMES, L.F. (2001) Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, **96**, 161-173.
- KLUGMAN, S.A., PANJER, H.H. and WILLMOT, G.E. (2004) *Loss Models: From Data to Decisions*, 2nd Edition. New Jersey: John Wiley & Sons.
- LO, A.Y. (1984) On a class of Bayesian nonparametric estimates I: Density estimates. *Ann. Statist.*, **12**, 351-357.
- LO, A.Y., BRUNNER, L.J. and CHAN, A.T. (1996) Weighted Chinese restaurant processes and Bayesian mixture models. Research Report, Hong Kong University of Science and Technology. Available at <http://www.erin.utoronto.ca/~jbrunner/papers/wcr96.pdf>
- MACEachern, S.N. (1994) Estimating normal means with a conjugate style Dirichlet Process Prior. *Communications in statistics – Simulation and computation*, **23**, 727-741.
- MOWBRAY, A.H. (1914) How extensive a payroll exposure is necessary to give a dependable pure premium? *Proceedings of the Casualty Actuarial Society*, **I**, 24-30.
- NEAL, R.M. (2000) Markov Chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, **9**, 249-265.
- RAMSAY, C.M. (1994) Loading gross premiums of risk without using utility theory. *Transactions of Society of Actuaries*, **XLV**, 305-349.
- SIU, T.K. and Yang, H. (1999) Subjective risk measures: Bayesian predictive scenarios analysis. *Insurance: Mathematics & Economics*, **25(2)**, 157-169.
- WANG, S.S. and YOUNG, V.R. (1998) Risk-adjusted credibility premiums using distorted probabilities. *Scand. Actuarial Journal*, **2**, 143-165.
- WATERS, H. (1993) Credibility Theory. Department of Actuarial Mathematics and Statistics, School of Mathematical and Computer Sciences, Heriot-Watt University.
- WEST, M., MÜLLER, P. and ESCOBAR, M.D. (1994) Hierarchical Priors and Mixture Models, With Applications in Regression and Density Estimation, in *A tribute to D.V. Lindley, eds. A.F.M. Smith and P.R. Freeman*, New York: Wiley.
- WHITNEY, A.W. (1918) The theory of experience rating. *Proceedings of the Casualty Actuarial Society*, **IV**, 274-292.

YOUNG, V.R. (1997) Credibility using semiparametric models. *ASTIN Bulletin*, **27**, 273-285.

YOUNG, V.R. (1998) Robust Bayesian credibility using semiparametric models. *ASTIN Bulletin*, **28**, 187-203.

JOHN W. LAU

*Department of Mathematics
University of Bristol
Bristol, United Kingdom
Email: John.Lau@bristol.ac.uk*

TAK KUEN SIU

*Department of Actuarial Mathematics and Statistics
School of Mathematical and Computer Sciences and
the Maxwell Institute for Mathematical Sciences
Heriot-Watt University
Edinburgh, United Kingdom
E-mail: T.K.Siu@ma.hw.ac.uk*

HAILIANG YANG

*Department of Statistics and Actuarial Science
The University of Hong Kong
Pokfulam Road, Hong Kong
Email: hlyang@hkusua.hku.hk*