Insurance: Mathematics and Economics 82 (2018) 87–94

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

journal homepage: www.elsevier.com/locate/ime

Bayesian ratemaking with common effects modeled by mixture of Polya tree processes[☆]

Jianjun Zhang, Chunjuan Qiu [*], Xianyi Wu

School of Statistics, East China Normal University, China

ARTICLE INFO

Article history:
Received April 2017
Received in revised form April 2018
Accepted 15 June 2018
Available online 3 July 2018

Keywords:
Bayesian nonparametrics
Common effects
Credibility models
Experience ratemaking
MCMC and Polya tree process

ABSTRACT

In classical models for Bayesian ratemaking, claims are usually assumed to be independent over risks. However, this assumption may be violated because there are situations that could derive possible dependence among the insured individuals. This paper aims to investigate the typical problem of experience ratemaking to account for a special type of dependence that is known as common effects in the literature. Polya tree processes are employed to model the common effects and, by means of an MCMC scheme, the corresponding Bayesian premiums are numerically computed. This provides a useful alternative to the well known results on Bayesian ratemaking with common effects.

1. Introduction

One of the most important tasks for actuaries is to determine adequate premiums for risks to be insured under certain premium principles derived from economical and operational considerations, to reflect the distributional features of the risks. These considerations may be choice of error functions, utilities, additivity for comonotonic or independent risks, and so on, see, for example, Gerber (1979), Kass et al. (2001), Wu and Zhou (2006), among a vast body of literature, for the details. In real operations, it is rarely the case that the loss distributions can be perfectly specified and, hence, actuaries generally need to fix insurance premiums with information extracted from claims data, giving rise to the so called experience ratemaking. Experience ratemaking has long been playing a central role in both the theoretical aspects and real operations of general insurance in which Bayesian methodologies are of fundamental importance (Makov et al., 1996). Because of mainly the ease in computation and robustness over the full Bayesian inference, credibility methods, of which the modern version was introduced to general insurance by Bühlmann (1967), are the most maturely developed Bayesian ratemaking that works under square errors and assigns premiums based on the simple linear combination of the historical claims data (see Norberg, 2004, Bühlmann and Gisler, 2005, Cai et al., 2015 and Pan et al., 2008, for examples).

We are concerned with full Bayesian ratemaking of which one of the extensively adopted settings is of the form (see e.g., (Bühlmann and Gisler, 2005):

Assumption 1.1.

1. The primitive is a portfolio of n individuals $i = 1, 2, \ldots, n$, of which each has contributed a vector $\boldsymbol{X}_i = (X_{i1}, X_{i2}, \ldots, X_{i,m_i})$ of m_i chronologically recorded claims, so as to predict future claims X_{i,m_i+1} for every individual i with the information provided by the whole observations.
2. Heterogeneity over the risks is characterized by n independent latent variables $\Theta_i, i = 1, 2, \ldots, n$, such that, given Θ_i, $\boldsymbol{X}_i$ is a sample of m_i i.i.d. observations from a perfectly specified density $f(x|\Theta_i)$.

This setting simply states that the random vectors $(\boldsymbol{X}_i, \Theta_i), i = 1, 2, \ldots, n$ are independent across individuals (*independence over risks*) and for each i, the claims $X_{i1}, X_{i2}, \ldots X_{i,m_i}$ are conditionally independent given Θ_i (*conditional independence over time*). However, in this complex world, there exist many important insurance scenarios where these classical assumptions are certainly violated. On the one hand, for a single individual i, there may exist certain conditional dependence over time so that considerable attention has been drawn to experience ratemaking with time dependence structures, see, e.g., the work by Bolancé et al. (2008) and Frees and Wang (2005) under the framework of credibility theory. On the other hand, as argued by Yeo and Valdez (2006) , Wen et al. (2009)

☆ This work was partially supported by the National Social Science Foundation Key Program under Grant No. 17ZDA091, Shanghai Philosophy and Social Science Foundation under Grant No. 2015BGL001, NSFC under Grant No. 71771089 and the 111 Project under Grant No. B14019.

* Corresponding author.
E-mail address: cjqiu@stat.ecnu.edu.cn (C. Qiu).

https://doi.org/10.1016/j.insmatheco.2018.06.007
0167-6687/

and a few other followers, there also exist many important insurance practices where the dependence over risks are common. A classic example is house insurance for which geographic proximity of the insureds may result in exposures to common catastrophes such as typhoons, floods, fire, earthquake and so on. For another example, consider a group insurance coverage of an organization that provides health insurance to its employees subjected to the same working conditions and environment. For example in the case of workers' compensation in a manufacturing firm, a severe mechanical breakdown or explosion can have damaging impacts on the health of many of the employees. Also, it may involve several insureds for motor vehicle insurance at once in a collision.

The impacts of dependence over risks in the community of insurance have long been recognized and discussed. Albrecher et al. (2002), Albrecher et al. (2011) and Valdez and Mo (2002) discussed the ruin probabilities with dependent risks. The effects for stop-loss premiums were analyzed by Heilmann (1986), Hurlimann (1993) and Müller (1997). The credibility formulas of claim occurrence frequencies under dependent risks can be found in Purcaru and Denuit (2003, 2003). Quite recently, Valdez (2014) provided empirical evidence for the dependence between the claim occurrences frequencies.

As argued by Valdez (2014), it proves quite powerful in a variety of areas to model dependence between random variables by mixing distributions of individuals, resulting in the so-called mixture models or hierarchical Bayesian models. Another equivalent name of a mixture model in insurance ratemaking is the so-called risk model with common effects by Yeo and Valdez (2006) who proposed a rather special full Bayesian model with claim dependence caused by certain common factors and investigated the corresponding Bayesian premiums under normally distributed claim amounts combined with their conjugate prior distributions, also normal distributions. A remarkable extension of Yeo and Valdez (2006) to the credibility theory is Wen et al. (2009) who considered Bühlmann and Bühlmann–Straub's models with common effects to account for dependence over risks induced by common stochastic effects but without any particular requirements for the form of claims distributions.

As a member of the family of hierarchical Bayesian models in mathematical perspective, Bayesian/credibility models with common effects may essentially allow for various variants and extensions yet to be examined so as to add useful members to the existing toolbox for ratemaking.

Classically, the latent variables Θ_i in Assumption 1.1 are assumed to be independent and identical in certain probability distribution $\pi(\theta_i)$ (referred to as structure function or prior distribution). The mostly adopted priors are certain conjugate ones for mathematical ease to tackle. With the development in the past decades stemmed from the cornerstone by Ferguson (1973) on Dirichlet processes and rather recent prosperity of Bayesian methodologies for nonparametric statistics, the community of ratemaking may be benefited from the modeling by means of Bayesian nonparametrics that allow to assign more flexible and robust priors to Θ_i. There have indeed been a small body of applications of Bayesian nonparametrics to insurance ratemaking under Dirichlet process priors. Examples include Zehnwirth (1979) who investigated the Bayesian ratemaking under a Dirichlet process prior, Lau et al. (2006) who discussed the credibility premiums under mixture of Dirichlet processes developed by Lo (1984), and Fellingham et al. (2015) who employed Dirichlet Process mixture to analyze healthcare group claims, among a few others.

As important alternatives to Dirichlet processes in Bayesian nonparametrics, Polya tree processes (see Section 2.1 for details) not only provide highly flexible non-parametric solutions to traditional parametric models but also possess many attractive properties over Dirichlet process models. The most appealing is that a Polya tree process can be made to be priors for continuous, even absolutely continuous distributions, by suitably choosing their super parameters, while Dirichlet processes can only give probability 1 to the set of discrete probability measures. Though Polya tree models have been introduced into probability community for several decades by Freedman (1963), Fabius (1964) and Ferguson (1974), in contrast to the latterly invented ones based on Dirichlet processes (Ferguson, 1973), as argued by Robert and Casella (2011), computational difficulty blocked possible advances of statistical modeling using Polya trees in especially real world applications until the contemporary intensive use of Markov chain Monte Carlo (MCMC) after the genuine starting point (Gelfand and Smith, 1990) in the mainstream statistical community. Recent development of Polya trees in statistics can be found in, for example, Berger and Guglielmi (2001) who used a mixture of Polya trees (MPT) in testing whether data arise from a family of parametric distributions versus a mixture of Polya trees alternative, Paddock (1999) who examined the properties of so-called "randomized Polya trees", Polya trees with slightly perturbed, random partitions, with applications to multivariate data and Hanson and Johnson (2002) who considered the linear regression model with error terms arising from a random distribution assigned an MPT prior.

Clearly, it is appealing to take the compromise to characterize common effects phenomenon by ideas from Bayesian nonparametrics, especially by the use of Polya tree processes. This paper intends to make efforts in this aspect by examining the setting described in Assumption 1.1 but with the priors modeled by mixtures of Polya tree processes, so as to make further contribution to this clearly overlooked discipline. To be specific, instead of the full Bayesian models proposed by Yeo and Valdez (2006) for normally distributed claims and the credibility model by Wen et al. (2009), we consider the Bayesian premiums of future risks derived by minimizing the expectation of some general error functions in the framework of decision-making, under Assumption 1.1, but with the characteristic parameters Θ_i following mixture of Polya tree processes (see Section 2.1 for details), as compromises between the solutions by full parametric Bayesian and fully distribution-free credibility methods.

As is generally the case in Bayesian statistics, it is also impossible to acquire analytical solutions to experience premiums because of mostly the inherent intricacy of the posterior distributions and frequently the irregularity of error functions. It turns out that the final solutions of ratemaking under this type of models sacrifice the tractability in mathematics and have to depend largely on the computation of the posterior expectations regarding the parameters. Because it is usually impossible to directly compute posterior expectations involving the parameters under Polya tree process priors, for the ratemaking to be numerically computed, we are devoted to the MCMC schemes of the corresponding Bayesian premiums.

The remainder of this paper is organized as follows. The general definition of mixture of Polya trees is reviewed in Section 2 for later reference. Also discussed there are the detailed description of Bayesian premiums in general decision-making framework and the computational difficulty they face. The MCMC scheme for the conditional expectations and thus the corresponding Bayesian premiums are presented in Section 3. Section 4 reports a simulation experiment on log-normally distributed claim losses. Some conclusions are discussed in Section 5.

2. Problem formulation

This section consists of three parts. The first recalls the definition of Polya tree processes on probability measures, which was first introduced to Bayesian statistics by Ferguson (1974) and further discussed in detail by Lavine (1994, 1994) and Mauldin et al. (1992) as a compromise between Dirichlet Processes and the more general tail free processes, so as to host a larger class of tractable priors than Dirichlet processes and provide a flexible framework for Bayesian analysis of nonparametric problems. Then we discuss in general Bayesian premiums under Polya tree process models so as to identify the quantity to be computed for acquiring the Bayesian premiums desired. Also discussed is the possible difficulty preventing direct computation of those Bayesian premiums.

2.1. Mixture of Polya tree processes

Conventionally, let $\{(\Omega, \mathscr{B}, G) : G \in \mathscr{P}\}$ be a family of probability spaces derived by a common measurable space $(\Omega, \mathscr{B})$ and indexed by $\mathscr{P}$, where Ω is a separable space, $\mathscr{B}$ the associated Borel σ-algebra and $\mathscr{P}$ the class of all probability measures on $\mathscr{B}$, equipped with the usual σ-algebra generated by weak convergence of probability measures and a probability measure. Defining a Polya tree prior needs the following notation:

(1) For any $j \in \{1, 2, \dots\}$, use E^j to stand for the set of all sequences of 0 s and 1 s of length j and write $E = \bigcup_{j=1}^{\infty} E^j$ for the set of all finite sequences of 0s and 1s with representative element ε. Let $B_\emptyset = \Omega$ and, for any $\varepsilon \in E$, $(B_{\varepsilon 0}, B_{\varepsilon 1})$ be a binary partition of B_ε, so that, $\pi_j = \{B_\varepsilon : \varepsilon \in E^j\}$ forms a partition of Ω (referred to as the level j partition) for every j and every $\Pi = \{\pi_j, j = 1, 2, \dots\}$ is indeed a binary tree of partitions of Ω with π_{j+1} refining π_j. Note that degenerate splits are not prohibited, for example, $B_\varepsilon = B_{\varepsilon 0} \bigcup \varnothing$. The union $\bigcup_{j=1}^{\infty} \pi_j = \{B_\varepsilon : \varepsilon \in E\}$ is assumed so rich a class of subsets of Ω that it generates the Borel σ-field $\mathscr{B}$, i.e., $\mathscr{B} = \sigma(\bigcup_{j=1}^{\infty} \pi_j)$.

(2) For every fixed $G \in \mathscr{P}$, associate to any $\varepsilon \in E$ a notation

$$Y_\varepsilon = G(B_{\varepsilon 0} | B_\varepsilon) \text{ (hence } Y_\emptyset = G(B_0)), \quad (1)$$

so that, for every $m \geq 1$ and every $\varepsilon = \epsilon_1 \cdots \epsilon_m \in E^m$, $G(B_\varepsilon) = \prod_{j=1}^{m} Y_{\epsilon_1 \cdots \epsilon_{j-1}}^{1-\epsilon_j} (1 - Y_{\epsilon_1 \cdots \epsilon_{j-1}})^{\epsilon_j}$, where the term for $j = 1$ is understood as $Y_\emptyset^{1-\epsilon_1}(1 - Y_\emptyset)^{\epsilon_1}$. This establishes an equivalence between the sets $\{G(B) : B \in \Pi\}$ and $\{Y_\varepsilon : \varepsilon \in E\}$.

(3) Let $\alpha_\varepsilon = \alpha(B_\varepsilon)$ be a positive set function on $\bigcup_{j=1}^{\infty} \pi_j$ with $\alpha(\emptyset) = 0$ if $\emptyset \in \bigcup_{j=1}^{\infty} \pi_j$ and write $\mathscr{A} = \{\alpha_\varepsilon : \varepsilon \in E\}$.

Definition 2.1. Let M be a positive integer or infinite and G_0 (referred to as centering distribution) a distribution on $(\Omega, \mathscr{B})$. A probability measure on $\mathscr{P}$ (or, equivalently, the random probability measure G on $(\Omega, \mathscr{B})$) is called a Polya tree process with parameters $(\Pi, \mathscr{A}, M, G_0)$, writing $G \sim \mathrm{PT}(\Pi^M, \mathscr{A}^M; G_0)$, if

(a) the elements of $\{Y_\varepsilon : \varepsilon \in E^m, m = 1, 2 \dots M - 1\}$ (see Eq. (1)) are mutually independent and $Y_\varepsilon \sim \mathrm{Beta}(\alpha_{\varepsilon 0}, \alpha_{\varepsilon 1})$ for all $\varepsilon \in E^m$, $m = 1, 2 \dots M - 1$, and
(b) on all the partition sets in level M, G follows G_0 in the sense that $G(A|B_{\epsilon_1 \dots \epsilon_M}) = \frac{G_0(A)}{G_0(B_{\epsilon_1 \dots \epsilon_M})}$ for any $A \subset B_{\epsilon_1 \dots \epsilon_M}$ for the case $M < \infty$.

This definition covers both partially and fully specified Polya tree processes. The fully specified corresponds to the setting $M = \infty$ for which G_0 takes no effect and thus is denoted by $\mathrm{PT}(\Pi, \mathscr{A})$; the case is not excluded that Π can be derived from G_0 (see below for an example).

On the one hand, fully specified Polya tree processes are applied more extensively than tail free priors because they are conjugate with respect to nonparametric models so as to give tractable posteriors. On the other hand, they are preferable over Dirichlet process priors in that they are capable of sitting on the set of continuous, or even absolutely continuous probability measures with respect to certain probability (or σ-finite) measure on $(\Omega, \mathscr{B})$ in probability 1, under suitably specified super parameters, whereas Dirichlet process priors always give probability 1 to the set of discrete probability measures on Ω (see, e.g.,Theorem 3.2.3 of Ghosh and Ramamoorthi, 2003 or Sethuraman, 1994's representation).

In this paper, we work only with the canonical case in which $\Omega = \mathbb{R}$, G_0 is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$ and every partition π_m consists of the subsets $B_\varepsilon = G_0^{-1}\left(\frac{\sum_{j=1}^{m} \epsilon_j 2^{j-1}}{2^m}, \frac{\sum_{j=1}^{m} \epsilon_j 2^{j-1} + 1}{2^m}\right]$, $\varepsilon = \epsilon_1 \dots \epsilon_m \in E^m$, so that the Polya tree process is simplified as $\mathrm{PT}^M(G_0)$.

In this canonical form, no matter $M = \infty$ or finite, the centering distributions G_0 play an important role and thus need to be cautiously treated (Walker and Mallick, 1999 and Barron et al., 1999). For a canonical Polya tree, it often causes much more spread out to use a fixed centering distribution than what the data warrant. A preferable way is to introduce mixtures of Polya trees (MPT) as in Hanson and Johnson (2002), who found that, in modeling regression errors, the mixture procedure increased efficiency of the Monte Carlo algorithm over a subjectively assigned centering distribution that was quite different from the true error distribution. According to Hanson and Johnson (2002), this simple mixture can smooth out the partitioning effects of a single Polya tree and, at the same time, avoid choosing a centering distribution. A mixture of Polya trees, denoted by $\mathrm{MPT}(G_\eta, H)$, is induced for G by allowing the centering distribution to be random, as defined below.

Definition 2.2. Let G_η stand for a family of parametric distributions with density g_η and indexed by a parameter η that is also a random variable distributed as a probability measure $\mathrm{H}(\eta)$. Then, $G \sim \mathrm{MPT}(G_\eta, H)$ if $G|\eta \sim \mathrm{PT}^M(G_\eta)$, $\eta \sim H$.

With an appropriately selected distribution H of η, MPTs can accommodate any features of the data, such as skewness and multi-modality, etc., even in the most simplest case of canonical Polya trees (Lavine, 1992).

The following presented are two general remarks.

1. The values of $\mathscr{A}$ determine how smoothly the probability measure G changes. For example, the realizations of G are mutually singular with probability 1 if $\alpha_\varepsilon = 1$ for every $\varepsilon \in E$ and mutually absolutely continuous with probability 1 if $\alpha_\varepsilon = m^2$ for each $\varepsilon \in E^m$ (Ferguson, 1974). Moreover, the predictive distribution under $\mathrm{PT}(\Pi, \mathscr{A})$ prior is closer to $\mathbb{E}[G]$ for larger values of α_ε and to the sample distribution for smaller values of α_ε. Typical choice of $\mathscr{A}$ for $\mathrm{PT}(\Pi, \mathscr{A})$ processes to sit on the collection of absolutely continuous probability measures is $\alpha_\varepsilon = c\rho(m)$ with a constant $c > 0$ and function $\rho(m) = m^2, m^3, 2^m, 4^m$ or 8^m (Berger and Guglielmi, 2001). Johnson (2006) declared that $\rho(m) = m^2$ was sufficient to capture interesting features of most distributions used in practice. In this paper, we take the convenience $\alpha_{\epsilon_1 \cdots \epsilon_m} = m^2$, as in Walker and Mallick (1999) and Paddock (1999).
2. In contrast to $\mathrm{PT}(\Pi, \mathscr{A})$, Lavine (1992) argued that $G \sim \mathrm{PT}(\Pi^M, \mathscr{A}^M; G_0)$ simplifies models and calculations and a $\mathrm{PT}(\Pi, \mathscr{A})$ can be more and more accurately approximated by a $G \sim \mathrm{PT}(\Pi^M, \mathscr{A}^M; G_0)$ with increasing depth M and this approximation is exact for M large enough (Proposition 1 of Hanson and Johnson, 2002). Hanson (2006) noted that high value of M often does little to improve inference but makes a model more time consuming to fit. Lavine (1994) suggested to chose the maximum level M by placing bounds on the posterior predictive density at a point. Paddock et al. (2003) used a pre-specified mediate integer $M = 10$. Hanson and Johnson (2002) and Hanson (2006) empirically determined M by the rule of thumb $M \approx \log_2\{n/N\}$, where N was a pre-assigned "typical" number of observations falling into each set at level M and n the sample size, so that the maximum

level was $M \approx \log_2 n$ for the assignment of "typical" number $N = 1$.

2.2. Bayesian premiums and their computational difficult

We work with a portfolio of n insureds $i = 1, 2, \ldots, n$. The risk feature of individual i is characterized by a latent random variable Θ_i and individual i has contributed a sequence of empirical claims $\boldsymbol{X}_i = (X_{i1}, X_{i2}, \ldots, X_{im_i})$. For mathematical ease, we sometimes are involved with a potential claims history $X_{i1}, X_{i2}, \ldots, X_{im_i} \ldots$ in which the first m_i claims are assumed having been observed. The purpose is to determine appropriate future premiums for the individuals in terms of the claims data and the distributional feature of the quantities involved. The observations and latent parameters from the whole portfolio are denoted respectively by $\mathcal{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n)$ and $\boldsymbol{\Theta} = (\Theta_1, \Theta_2, \ldots, \Theta_n)$. The experience ratemaking is discussed under the following distributional setting in which the significant feature is that the priors follow mixture of Polya trees.

Assumption 2.1 (*Model Structure*)**.**

(a) G is a mixture of Polya trees $\text{MPT}(G_\eta, H)$ on the parameter space $(\Omega_\Theta, \mathscr{B}_\Theta)$, i.e., $G|\eta \sim \text{PT}^M(G_\eta), \eta \sim H$, as defined in Definition 2.2.

(b) Given G, the sequences $(\Theta_i; X_{i1}, X_{i2} \ldots, X_{im_i}, \ldots)$ are mutually independent and identically distributed over $i = 1, 2, \ldots, n$ such that, $\Theta_i \sim G$ and $X_{i1}, X_{i2}, \ldots, X_{i,m_i}, \ldots |\Theta_i \overset{i.i.d}{\sim} f(x|\Theta_i), i = 1, 2, \ldots, n$ for certain completely specified probability measure f (generally represented by a density).

Note that the data $(\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n)$ possess a hierarchical structure and thus are associated with a type of equa-correlation, as what is typical in Bayesian statistics, over both laterally the risks and longitudinally the claims from any individuals. For example, the covariance matrix of $(\boldsymbol{X}_1, \boldsymbol{X}_2)$, for a special case $m_1 = m_2 = m$, can be expressed as

$$C_1 \begin{pmatrix} \boldsymbol{I}_m & 0 \\ 0 & \boldsymbol{I}_m \end{pmatrix} + C_2 \begin{pmatrix} \boldsymbol{1}_m \boldsymbol{1}'_m & 0 \\ 0 & \boldsymbol{1}_m \boldsymbol{1}'_m \end{pmatrix} + C_3 \begin{pmatrix} \boldsymbol{1}_m \boldsymbol{1}'_m & \boldsymbol{1}_m \boldsymbol{1}'_m \\ \boldsymbol{1}_m \boldsymbol{1}'_m & \boldsymbol{1}_m \boldsymbol{1}'_m \end{pmatrix} \tag{2}$$

for some positive constants C_1, C_2 and C_3, where $\boldsymbol{1}_m$ stands for the m-vector of 1s. However, explicit computation of the constants C_1, C_2 and C_3 is not an easy exercise due to the complex elements involved, including the choices of $f(x|\Theta)$, the depth M of the binary partitions in defining the Polya tree processes, the mixing distribution H and the absence of a closed form for the (random and marginal) density of G_η.

Premiums for future claims are set up under the general decision-making framework. To this end, use $L(a, b)$ to denote a generic bivariate function indicating the error of using b to predict a, referred to as an error function. For every individual $i = 1, 2, \ldots, n$, the objective of experience ratemaking is to find a statistic $\hat{P}_i$ (referred to as the Bayes premium), such that

$$\mathbb{E}L(X_{i,m_i+1}, \hat{P}_i) = \min_Q \mathbb{E}L(X_{i,m_i+1}, Q(\mathcal{X})), \tag{3}$$

where the minimization is taken over all measurable function Q of $\mathcal{X}$.

To figure out the quantity that we need to generally compute so as to resolve the optimization problem (3), we discuss the following three subcategories:

(1) There exists a closed form solution to (3) and $\hat{P}_i$ can be determined by computing $\mathbb{E}[g(\Theta_i)|\mathcal{X}]$ for certain function g. Due to its simplicity in mathematics, the most popular and most extensively applied example is the square error $L(a, b) = (b - a)^2$ under which the Bayes premiums are simply $\hat{P}_i = \mathbb{E}(X_{i,m_i+1}|\mathcal{X})$. By the conditional independence among individuals given G and among $X_{i1}, X_{i2}, \ldots, X_{im_i}, \ldots$ given G and Θ_i (part (b) of Assumption 2.1), it follows that

$$\hat{P}_i = \mathbb{E}[(X_{i,m_{i+1}}|\Theta_1, \ldots, \Theta_n, G, \mathcal{X})|\mathcal{X}] = \mathbb{E}[\mathbb{E}(X_{i1}|\Theta_i)|\mathcal{X}].$$

In other words, we need to compute a conditional expectation $\mathbb{E}[g(\Theta_i)|\mathcal{X}]$ with $g(\Theta_i) = \mathbb{E}(X_{i1}|\Theta_i)$. This particular example is almost the same as the so-called risk models with common effects analyzed by Yeo and Valdez (2006) and Wen et al. (2009).

(2) $\hat{P}_i$ is a solution of an equation that can be expressed as the conditional expectation of certain function of parameters given the data.
A frequently discussed example is the asymmetric absolute error function $L(a, b) = |a - b|(\tau I_{\{a>b\}} + (1 - \tau) I_{\{a<b\}})$. Denote by $F^i_{\mathcal{X}}(x)$ the conditional distribution function of X_{i,m_i+1} given $\mathcal{X}$. Then, $\hat{P}_i$ is the solution of equation $F^i_{\mathcal{X}}(x) - \tau = 0$, i.e., $\hat{P}_i = \text{VaR}_\tau(F^i_{\mathcal{X}})$, the quantity best known in risk management community as Value-at-Risk (VaR). This equation does generally not allow for analytical solution and needs to be solved by means of *Newton iteration* provided $F^i_{\mathcal{X}}(x)$ can be numerically computed. By again the conditional independence in part (b) of Assumption 2.1, it follows that

$$\begin{aligned} F^i_{\mathcal{X}}(x) &= \mathbb{E}[\Pr(X_{i,m_i+1} < x|\Theta_1, \ldots, \Theta_n, G, \mathcal{X})|\mathcal{X}] \\ &= \mathbb{E}[F(x|\Theta_i)|\mathcal{X}] = \mathbb{E}[g(\Theta_i)|\mathcal{X}], \end{aligned}$$

where $g(\Theta_i) = F(x|\Theta_i)$ is the conditional distribution function of X_{i,m_i+1} given Θ_i.

(3) General cases.
Because $\mathbb{E}[L(X_{i,m_i+1}, Q(\mathcal{X}))] = \mathbb{E}[\mathbb{E}[L(X_{i,m_i+1}, Q(\mathcal{X}))|\mathcal{X}]]$, $\hat{P}_i$ is generally a solution to $\min_b \mathbb{E}[L(X_{i,m_i+1}, b)|\mathcal{X}]$, in which, by still the conditional independence, the objective function to be minimized can be expressed as

$$\begin{aligned} \mathbb{E}[L(X_{i,m_i+1}, b)|\mathcal{X}] &= \mathbb{E}[\mathbb{E}[L(X_{i,m_i+1}, b)|\Theta_1, \ldots, \Theta_n, G, \mathcal{X}]|\mathcal{X}] \\ &= \mathbb{E}[g(b, \Theta_i)|\mathcal{X}] \end{aligned}$$

with $g(b, \Theta_i) = \mathbb{E}[L(X_{i,m_i+1}, b)|\Theta_i]$. The minimization $\min_b \mathbb{E}[g(b, \Theta_i)|\mathcal{X}]$ can be numerically retrieved in certain particular cases.

To summarize, in order to tackle the problem of experience ratemaking in decision-making framework, it is of crucial importance to compute posterior expectation $\mathbb{E}(h(\Theta_i)|\mathcal{X})$ for a general function h.

2.3. The difficulty in direct computation

In order to conceptually demonstrate how difficult it is to directly compute the desired posterior expectations by means of integration, we first introduce a few necessary notation at this point, which can also serve later use.

(1) For any $\theta \in \Omega_\Theta$, denote by $\varepsilon(m, \eta, \theta)$ the binary expansion $\varepsilon \in E_m$ such that $\theta \in B^\eta_\varepsilon$, i.e., the binary expansion of the level-m set containing θ.

(2) For any n-vector $\boldsymbol{\theta} = (\theta_1 \cdots \theta_n)$, denote by $\boldsymbol{\theta}^{(i)}$ the $(n-1)$-subvector of $\boldsymbol{\theta}$ by dropping θ_i and $n^i_{\varepsilon(m,\eta,\theta_i)}$ the number of the components of $\boldsymbol{\theta}^{(i)}$ lying in $B^\eta_{\varepsilon(m,\eta,\theta_i)}$. Because $n^i_{\varepsilon(m,\eta,\theta_i)}$ is clearly nonincreasing in m, we can let $\text{M}(\eta, \theta_i)$ be the smallest level m such that $n^i_{\varepsilon(m,\eta,\theta_i)} = 0$, i.e., no elements of $\boldsymbol{\theta}^{(i)}$ is contained in $B^\eta_{\varepsilon(m,\eta,\theta_i)}$.

(3) For any $m = 1, 2, \ldots$, denote by $n_{\varepsilon(m,\eta,\theta_j)}$ the number of the elements of $(\theta_1, \ldots, \theta_{j-1})$ that lie in $B^{\eta}_{\varepsilon(m,\eta,\theta_j)}$ and $\alpha'_{\varepsilon(m,\eta,\theta_j)} = \alpha_{\varepsilon(m,\eta,\theta_j)} + n_{\varepsilon(m,\eta,\theta_j)}$ the parameter updated by the observed count of $(\theta_1, \ldots, \theta_{j-1})$ in $B_{\varepsilon(m,\eta,\theta_j)}$. Note that $n_{\varepsilon(m,\eta,\theta_n)} = \varepsilon^n(m, \eta, \theta_n)$.

By Theorem 2 of Lavine (1992), the joint density of $(\Theta_1 \cdots \Theta_n)$ for given η is expressed as

$$\pi_\eta(\boldsymbol{\theta}) = g_\eta(\theta_1) \prod_{j=2}^{n} \prod_{m=1}^{\mathrm{M}(\eta,\theta_j)} \frac{\alpha'_{\varepsilon(m,\eta,\theta_j)}(\alpha_{\varepsilon(m-1,\eta,\theta_j)0} + \alpha_{\varepsilon(m-1,\eta,\theta_j)1})}{\alpha_{\varepsilon(m,\eta,\theta_j)}(\alpha'_{\varepsilon(m-1,\eta,\theta_j)0} + \alpha'_{\varepsilon(m-1,\eta,\theta_j)1})} \times g_\eta(\theta_j) \tag{4}$$

where $g_\eta(\theta)$ is the density of the base measure G_η with respect to Lebesgue measure; the formula for $n = 1$ corresponds to the marginal density $g_\eta(\theta)$ of Θ_i for given η. For the case $\alpha_{\varepsilon_m} = m^2, m = 1, 2, \ldots$ that we are considering,

$$\pi_\eta(\boldsymbol{\theta}) = g_\eta(\theta_1) \prod_{j=2}^{n} \prod_{m=1}^{\mathrm{M}(\eta,\theta_j)} \frac{2m^2 + 2n_{\varepsilon(m,\eta,\theta_j)}}{2m^2 + n_{\varepsilon(m-1,\eta,\theta_j)}} g_\eta(\theta_j). \tag{5}$$

Consequently, for any measurable function $h(\Theta_t)$ on $(\Omega_\Theta, \mathscr{B})$, $t \in \{1, 2, \ldots, n\}$,

$$\mathbb{E}[h(\Theta_t)|\mathcal{X}, \eta] = \frac{\int_{\Omega_\Theta^n} h(\theta_t) \prod_{i=1}^{n} f(\mathbf{X}_i|\theta_i)\pi_\eta(\boldsymbol{\theta})\mathrm{d}\theta_1 \ldots \mathrm{d}\theta_n}{\int_{\Omega_\Theta^n} \prod_{i=1}^{n} f(\mathbf{X}_i|\theta_i)\pi_\eta(\boldsymbol{\theta})\mathrm{d}\theta_1 \ldots \mathrm{d}\theta_n}. \tag{6}$$

Formulas (4) through (6) clearly state the intricacy and generally block any attempt to generally compute the posterior expectation by means of direct integration, even for a mediate number of individuals. Hence, it is inevitable to resort to the contemporarily developed MCMC methodologies, especially the Gibbs sampling, for posterior distributions, so as to numerically approximate the posterior expectations involved.

3. Posterior expectation via MCMC

The notation introduced in previous section are followed here. In what follows, we introduce the MCMC procedure for the Bayesian ratemaking discussed in the previous section under the model setting defined by Assumption 2.1. It involves conditionally sampling Θ_i given $(\boldsymbol{\Theta}^{(i)}, \mathcal{X}, \eta)$ and sampling η given $\mathcal{X}$, as depicted separately Sections 3.1 and 3.2. Section 3.3 presents the integrated algorithm that computes the posterior expectation of $g(\Theta_i)$.

To proceed, however, we need in turn the conditional distributions of Θ_i given $(\boldsymbol{\Theta}^{(i)} = \boldsymbol{\theta}^{(i)}, \eta)$ and $(\Theta_1 = \theta_1, \ldots, \Theta_{i-1} = \theta_{i-1}, \eta)$, respectively, as expressed by

$$\pi(\theta_i|\boldsymbol{\theta}^{(i)}, \eta) = \prod_{m=1}^{\mathrm{M}} \frac{2m^2 + 2n^i_{\varepsilon(m,\eta,\theta_i)}}{2m^2 + n^i_{\varepsilon(m-1,\eta,\theta_i)}} g_\eta(\theta_i), \tag{7}$$

$$\pi(\theta_i|\theta_1 \ldots, \theta_{i-1}, \eta) = \prod_{m=1}^{\mathrm{M}} \frac{2m^2 + 2n_{\varepsilon(m,\eta,\theta_i)}}{2m^2 + n_{\varepsilon(m-1,\eta,\theta_i)}} g_\eta(\theta_i), \tag{8}$$

which are similar to Eq. (4) in Hanson and Johnson (2002) for mixture of Polya tree processes.

3.1. Sampling Θ_i given $(\boldsymbol{\Theta}^{(i)}, \mathcal{X}, \eta)$

The following theorem establishes the conditional distribution of Θ_i given $(\boldsymbol{\Theta}^{(i)}, \mathcal{X}, \eta)$ under the data structure described in Section 2.2.

Theorem 3.1. *The conditional distribution density of Θ_i given $(\boldsymbol{\Theta}^{(i)} = \boldsymbol{\theta}^{(i)}, \mathcal{X}, \eta)$ is*

$$\pi(\theta_i|\boldsymbol{\theta}^{(i)}, \mathcal{X}, \eta) \propto \prod_{m=1}^{\mathrm{M}} \frac{2m^2 + 2n^i_{\varepsilon(m,\eta,\theta_i)}}{2m^2 + n^i_{\varepsilon(m-1,\eta,\theta_i)}} f(\boldsymbol{X}_i|\theta_i)g_\eta(\theta_i)$$
$$\propto \prod_{m=1}^{\mathrm{M}} \frac{2m^2 + 2n^i_{\varepsilon(m,\eta,\theta_i)}}{2m^2 + n^i_{\varepsilon(m-1,\eta,\theta_i)}} g_\eta(\theta_i|\boldsymbol{X}_i), \tag{9}$$

where the meaning of $g_\eta(\theta_i|\boldsymbol{X}_i)$ is self-explanatory.

Proof. By Eq. (7),

$$\begin{aligned}\pi(\theta_i|\boldsymbol{\theta}^{(i)}, \mathcal{X}, \eta) &= \pi(\theta_i|\boldsymbol{\theta}^{(i)}, \boldsymbol{X}_i, \eta) \\ &= \frac{f(\boldsymbol{X}_i|\theta_i)\pi(\theta_i|\boldsymbol{\theta}^{(i)}, \eta)}{\int f(\boldsymbol{X}_i|\theta_i)\pi(\theta_i|\boldsymbol{\theta}^{(i)}, \eta)\mathrm{d}\theta_i} \\ &\propto \prod_{m=1}^{\mathrm{M}} \frac{2m^2 + 2n^i_{\varepsilon(m,\eta,\theta_i)}}{2m^2 + n^i_{\varepsilon(m-1,\eta,\theta_i)}} f(\boldsymbol{X}_i|\theta_i)g_\eta(\theta_i),\end{aligned}$$

where the first equation holds because Θ_i is conditionally independent of $\{\mathbf{X}_j, j \neq i\}$ given $(\boldsymbol{\Theta}^{(i)}, \eta)$. The theorem thus proves. ■

In order to draw Θ_i given $(\boldsymbol{\Theta}^{(i)}, \mathcal{X}, \eta)$, we need to differentiate two cases:

First, if G_η is a conjugate with respect to $f(X_{ij}|\theta_i)$, then Θ_i given $(\boldsymbol{\Theta}^{(i)}, \mathcal{X}, \eta)$ can be sampled from the conditional distribution given in Theorem 3.1 by means of an acceptance–rejection sampling method because now the conditional density of Θ_i given $(\Theta^{(i)}, \mathcal{X}, \eta)$ can be bounded by that of $g_\eta(\theta_i|\boldsymbol{X}_i)$ multiplied by a positive constant, as presented in the following lemma.

Lemma 3.1. *For each $\Theta_i \in \mathbb{R}$, the conditional distribution of Θ_i given $(\boldsymbol{\Theta}^{(i)} = \boldsymbol{\theta}^{(i)}, \mathcal{X}, \eta)$ satisfies $\pi(\theta_i|\boldsymbol{\theta}^{(i)}, \mathcal{X}, \eta) < Bg_\eta(\theta_i|\boldsymbol{X}_i)$ and here B is a θ_i-free quantity.*

Proof. By Theorem 3.1, we have

$$\begin{aligned}\pi(\theta_i|\boldsymbol{\theta}^{(i)}, \mathcal{X}, \eta) &= C^{-1} \prod_{m=1}^{\mathrm{M}} \frac{2m^2 + 2n^i_{\varepsilon(m,\eta,\theta_i)}}{2m^2 + n^i_{\varepsilon(m-1,\eta,\theta_i)}} f(\boldsymbol{X}_i|\theta_i)g_\eta(\theta_i) \\ &< C^{-1} \prod_{m=1}^{\mathrm{M}} \frac{m^2 + n - 1}{m^2} f(\boldsymbol{X}_i|\theta_i)g_\eta(\theta_i) \\ &= Bg_\eta(\theta_i|\boldsymbol{X}_i),\end{aligned}$$

where C is the normalizing constant and $B = C^{-1}\prod_{m=1}^{\mathrm{M}} \frac{m^2+n-1}{m^2} \int f(\boldsymbol{X}_i|\theta_i)g_\eta(\theta_i)\mathrm{d}\theta_i$. The lemma is thus proved by noting the fact $\prod_{m=1}^{\mathrm{M}} \frac{m^2+n-1}{m^2} \leq \prod_{m=1}^{\infty} \frac{m^2+n-1}{m^2} < \infty$. ■

Second, for general (maybe non-conjugate) case, Metropolis–Hastings (Hastings, 1970) algorithm is applied to produce samples of Θ_i given $(\boldsymbol{\Theta}^{(i)} = \boldsymbol{\theta}^{(i)}, \mathcal{X}, \eta)$. To be specific, Θ_i can be updated by sampling first $\theta_i^* \sim N(\theta_i, s^2)$ and then accepted with probability

$$\min\left\{1, \frac{q_\eta(\theta_i^*, \boldsymbol{\theta}^{(i)})}{q_\eta(\theta_i, \boldsymbol{\theta}^{(i)})}\right\},$$

where the parameter s^2 is tuning parameter and

$$q_\eta(\theta_i^*, \boldsymbol{\theta}^{(i)}) = \prod_{m=1}^{\mathrm{M}} \frac{m^2 + n_{\varepsilon(m,\eta,\theta_i^*)}}{2m^2 + n_{\varepsilon(m-1,\eta,\theta_i^*)}} f(\boldsymbol{X}_i|\theta_i^*) \exp(-\frac{1}{2}\sigma^{-2}(\theta_i^* - \eta)^2).$$

3.2. Sampling η given $\mathcal{X}$

Similarly, η can be sampled by MCMC from the full conditional distribution $h(\eta|\mathcal{X})$. By Eq. (8), the posterior conditional distribution of η given $\mathcal{X}$ can be computed by

$$\begin{aligned} h(\eta|\mathcal{X}) \propto& h(\eta) \int \prod_{i=1}^{n} f(\boldsymbol{X}_i|\theta_i)\pi(\theta_i|\theta_1,\ldots,\theta_{i-1},\eta)\mathrm{d}\theta_1\cdots\mathrm{d}\theta_n \\ =& h(\eta) \int \prod_{i=1}^{n} f(\boldsymbol{X}_i|\theta_i) \prod_{m=1}^{\mathrm{M}} \frac{2m^2+2n_{\varepsilon(m,\eta,\theta_i)}}{2m^2+n_{\varepsilon(m,\eta,\theta_i)}} g_\eta(\theta_i)\mathrm{d}\theta_1\cdots\mathrm{d}\theta_n, \end{aligned}$$

which can be approximated with Monte Carlo method by sampling a set of n-vector $(u_1^j, u_2^j, \ldots, u_n^j)$, $j = 1, 2, \ldots, A$ such that $u_1^j, u_2^j, \ldots, u_n^j \overset{\text{i.i.d.}}{\sim} g_\eta$ for every j and then averaging, where A is a sufficiently large integer. Then, sample η by means of Metropolis–Hastings algorithm with an updating rule characterized by the acceptance probability

$$\min\left\{1, \frac{R(\eta^*)}{R(\eta)}\right\}, \tag{10}$$

where

$$R(\eta) = h(\eta) \sum_{j=1}^{A} \prod_{k=1}^{n} f(\boldsymbol{X}_k|u_k^j) \prod_{m=1}^{\mathrm{M}} \frac{2m^2+2n_{j,\varepsilon(m,\eta,u_k^j)}}{2m^2+n_{j,\varepsilon(m,\eta,u_k^j)}} g_\eta(u_k^j) \tag{11}$$

and $n_{j,\varepsilon(m,\eta,u_k^j)}$ is the value of $n_{\varepsilon(m,\eta,u_k^j)}$ computed for the n-vector $(u_1^j, u_2^j \ldots, u_n^j)$.

3.3. The integrated algorithm

With an integration of the algorithms depicted above, the Gibbs samples of $\boldsymbol{\Theta}$ and η can be generated iteratively and then approximate $\mathbb{E}[h(\Theta_i)|\mathcal{X}]$ by the means of $h(\Theta_i)$ as what was depicted in the following Algorithm 3.1. Note that, in the algorithm the values of A, L and B need to be specified in advance.

Algorithm 3.1 (*The Integrated Algorithm*).

(1) Draw a number A of n-vectors $(u_1^j, u_2^j, \ldots, u_n^j)$ such that $u_1^j, u_2^j, \ldots, u_n^j \overset{\text{i.i.d.}}{\sim} g_\eta, j = 1, 2, \ldots, A$.
(2) Initiate with an arbitrary value $\eta^{(0)}$.
(3) For each iteration $l = 1, \ldots, L$, generate $\eta^{(l)}$ from the conditional density $h(\eta|\mathcal{X})$ using an MH algorithm characterized by Eqs. (10) and (11).

 (i) Then, with the given value $\eta^{(l)}$ and beginning at any initial value $\boldsymbol{\theta}_{(l)}^{(0)}$, use Eq. (9) to generate $\boldsymbol{\theta}_{(l)}^{(1)}$ by:

$$\begin{aligned} &\text{Sample } \theta_{1(l)}^{(1)} \text{ from} && \Theta_1|\Theta_2 = \theta_2^{(0)}, \Theta_3 = \theta_{3(l)}^{(0)}, \\ & && \ldots, \Theta_n = \theta_{n(l)}^{(0)}, \eta^{(l)}, \mathcal{X}; \\ &\text{Sample } \theta_{2(l)}^{(1)} \text{ from} && \Theta_2|\Theta_1 = \theta_{1(l)}^{(1)}, \Theta_3 = \theta_{3(l)}^{(0)}, \\ & && \ldots, \Theta_n = \theta_{n(l)}^{(0)}, \eta^{(l)}, \mathcal{X}; \\ &\vdots && \vdots \\ &\text{Sample } \theta_{n(l)}^{(1)} \text{ from} && \Theta_n|\Theta_1 = \theta_{1(l)}^{(1)}, \Theta_2 = \theta_{2(l)}^{(1)}, \\ & && \ldots, \Theta_{n-1} = \theta_{n-1,(l)}^{(1)}, \eta^{(l)}, \mathcal{X}. \end{aligned}$$

 (ii) Iterate procedure (i) to generate $\boldsymbol{\theta}_{(l)}^{(b)} = (\theta_{1(l)}^{(b)}, \theta_{2(l)}^{(b)}, \ldots, \theta_{n(l)}^{(b)})$ from $\boldsymbol{\theta}_{(l)}^{(b-1)}$ for any $b = 2, 3, \cdots, B$ in the same way as generating $\boldsymbol{\theta}_{(l)}^{(1)}$ from $\boldsymbol{\theta}_{(l)}^{(0)}$.

(4) Generate $\eta_1, \eta_2, \ldots, \eta_L$ by iterating this procedure L times and then approximate the posterior expectation $\mathbb{E}[h(\Theta_i)|\mathcal{X}]$ for general function g by

$$\hat{\mathbb{E}}[h(\Theta_i)|\mathcal{X}] = \frac{1}{\mathrm{LB}} \sum_{l=1}^{L} \sum_{b=1}^{B} h(\theta_{i(l)}^{(b)}). \tag{12}$$

4. Simulation

For numerical illustration, we conducted a small simulation study so as to compare the experience premiums under priors defined by mixture of Polya Trees and a popular parametric Bayesian models.

The claims data were produced from a log-normal distribution with the risk parameters following a mixture of Polya tree processes, i.e.,

$$\begin{aligned} &X_{ij} \sim \mathrm{LN}(\Theta_i, 1), j = 1, 2, \ldots, m_i = 30, \Theta_i \overset{iid}{\sim} \mathrm{MTP}(G_\eta, H), \\ &\quad i = 1, 2\ldots, n = 30, \\ &G_\eta = \mathrm{N}(\eta, 1), H(\eta) = \mathrm{N}(0, 100) \text{ and } \mathrm{M} = 5, \end{aligned}$$

by the following procedure:

(1) Draw an η from distribution H.
(2) Compute the level-M partition $B_{\epsilon_1\ldots\epsilon_M}(\eta)$.
(3) Draw a set $B_{\epsilon_1\ldots\epsilon_M}$ in level M with probability law

$$\begin{aligned} P(B_{\epsilon_1\ldots\epsilon_M}) &= \frac{\alpha_{\varepsilon_1}}{\alpha_0+\alpha_1} \times \cdots \times \frac{\alpha_{\varepsilon_1\cdots\varepsilon_M}}{\alpha_{\varepsilon_1\cdots\varepsilon_{M-1}0}+\alpha_{\varepsilon_1\cdots\varepsilon_{M-1}1}} \\ &= \frac{1}{2^M}, \quad \epsilon_1\ldots\epsilon_M \in E^M. \end{aligned}$$

(4) With the $B_{\epsilon_1\ldots\epsilon_M}$ selected, draw risk parameters θ_i from G_η restricted on the set $B_{\epsilon_1\ldots\epsilon_M}$.

The Bayesian premiums $\hat{P}_i$ for individual i in year $m_i + 1$ were computed under the following three error functions:

(1) Mean square error (MSE) $L(a, b) = (b - a)^2$.
(2) Asymmetric absolute error (AAE), $L(a, b) = \mathbb{E}[|a - b|(\tau I_{\{a>b\}} + (1-\tau)I_{\{a<b\}})]$, $\tau = 0.5$.
(3) Pseudo-Huber error (PHE), $L(a, b) = \sqrt{1 + (\frac{b-a}{\delta})^2} - 1$ with $\delta = 1$.

The premiums were computed in two methods:

(I) Method 1 used the true model and for each replicate, the posterior expectations were approximated by Eq. (12) with L = 100 and B = 200. The experience premiums $\hat{P}_i$s under the three error functions were computed according to Section 2.2, respectively, i.e., $\hat{P}_i$ is

(1) identical to $\mathbb{E}[g(\Theta_i)|\mathcal{X}]$ with $g(\Theta_i) = \mathbb{E}[X_{i1}|\Theta_i] = e^{\Theta_i+1/2}$ under MSE,

(2) the solution of $H_{\mathcal{X}}^i(x) - \tau = 0$, where $H_{\mathcal{X}}^i(x) = \Pr(X_{i,m_i+1} < x|\mathcal{X}) = \mathbb{E}[F_X(x|\Theta_i, 1)|\mathcal{X}]$ under AAE and

(3) the solution of $\mathbb{E}(g(x, \Theta_i)|\mathcal{X}) = 0$ under PHE, where $g(x, \Theta_i)$

$$= \mathbb{E}\left\{ \left. \frac{x - X_{i,m_{i+1}}}{\sqrt{1+\left\{\frac{x-X_{i,m_{i+1}}}{\delta}\right\}^2}} \right| \Theta_i \right\}.$$

(II) Method 2 factitiously supposed a parametric prior $\Theta_i \sim \mathrm{N}(\upsilon, \tau_0^2)$ with mean $\upsilon = 0$ and variance $\tau_0^2 = 100$, with a relatively large variance 100 to reflect the idea of noninformative prior for Θ_i. The corresponding posterior distribution of Θ_i was $\Theta_i|\mathcal{X} \sim \mathrm{N}(\upsilon^*, \tau_0^{*2})$ and the predictive distribution of X_{i,m_i+1} was also log-normally distributed: $X_{i,m_i+1}|\mathcal{X} \sim \mathrm{LN}(\upsilon^*, \tau_0^{*2} + \sigma^2)$ with

$$\upsilon^* = \frac{\upsilon + \tau_0^2 T}{1 + m_i\tau_0^2}, T = \sum_{j=1}^{m_i} \ln(X_{ij}) \text{ and } \tau_0^{*2} = \frac{\tau_0^2}{1 + m_i\tau_0^2}.$$

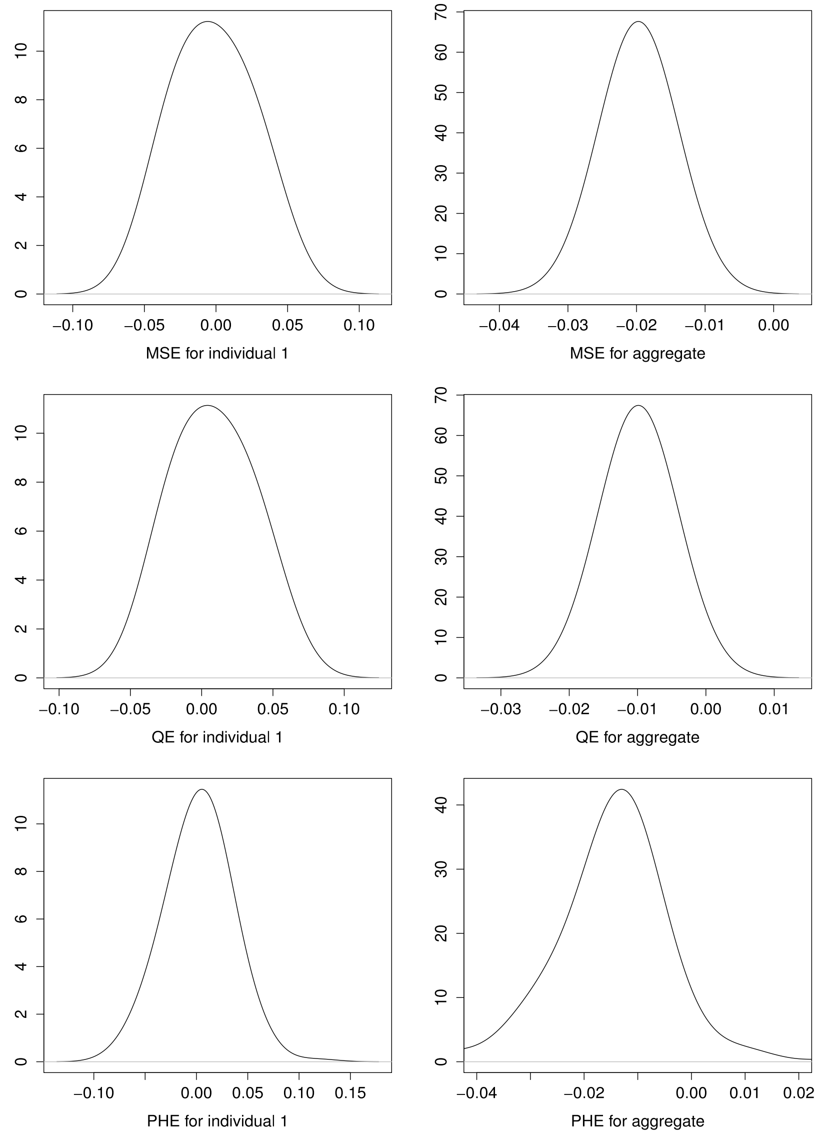

Fig. 1. The densities of the differences of individual 1 and aggregate.

Thus the Bayes premium $\hat{P}_i$ can be easily obtained through general steps under these three different error functions.

In order to compare the premiums, similar to Yeo and Valdez (2006), we computed the percentage difference in Bayes premiums for individual i between these two methods using

$$\Delta_i = \frac{\text{Method 1 premium} - \text{Method 2 premium}}{\text{Method 2 premium}} \times 100$$
$$= \frac{\hat{\mu}^{\text{Method 1}}_{i,m_i+1} - \hat{\mu}^{\text{Method 2}}_{i,m_i+1}}{\hat{\mu}^{\text{Method 2}}_{i,m_i+1}} \times 100,$$

where $\hat{\mu}^{Method1}_{i,m_i+1}$ denotes the Bayesian premiums using Method 1 and $\mu^{Method2}_{i,m_i+1}$ denotes that of Method 2. Moreover, the aggregate percentage difference

$$\Delta = \frac{\sum_{i=1}^{n} \mu^{\text{Method 1}}_{i,m_i+1} - \sum_{i=1}^{n} \mu^{\text{Method 2}}_{i,m_i+1}}{\sum_{i=1}^{n} \mu^{\text{Method 2}}_{i,m_i+1}}$$

was calculated also.

For each error function, a total of 200 duplicates (and thus the same number of percentage differences) are simulated and their empirical distributional features were examined for every individual $i = 1, 2, \ldots, 30$. Some descriptive statistics of the simulated data of Δ_1 (for individual 1) and Δ (the aggregate) are listed in Table 1.

Fig. 1 provides the estimated densities of the replicated differences (in original but not in the percentage scale) between these two methods produced by an R package, where the left panel is for individual 1 and the right is for the aggregate, under MSE, AAE and PHE, respectively.

According to these figures and statistics, the parametric Bayesian model in this simulation tended to give rise to biased approximates of the true Bayes premium calculated based on mixture of Polya tree model negative biases that were observed in this simulation.

5. Conclusion

Hierarchy or mixing distributions are powerful to model dependence among claims and risks. We have so far examined how to obtain Bayes premiums based on claims distributions completely specified in mathematical forms with priors modeled by mixture of Polya tree processes under general error functions. The computing strategy is to simulate posterior expectations of functions of

Table 1
Descriptive statistics for the differences (in percentage) of individual 1 and the aggregate.

Statistics	Individual 1			Aggregate		
	MSE	AAE	PHE	MSE	AAE	PHE
Mean (%)	−0.21	−0.79	−0.22	−1.97	−0.98	−1.50
Median (%)	−0.41	−0.59	−0.48	−1.98	−0.99	−1.37
Variance (%^{2})	6.60	6.71	9.06	0.10	0.10	1.19
Standard deviation (%)	2.57	2.59	3.01	0.31	0.31	1.09
Minimum (%)	−5.12	−4.16	−7.61	−2.83	−1.85	−5.69
Maximum (%)	5.36	6.42	11.77	−1.13	−0.14	2.76

the parameters Θ_i by means of Gibbs sampling method so as to overcome the difficulty in direct computation.

We believe this will possibly open a space for promising future research on Bayesian ratemaking with other nonparametric priors such as stick-breaking type mixtures, including, for example, general stick-breaking mixtures (Ishwaran and James, 2001), matrix stick-breaking process (Dunson et al., 2008), kernel stick-breaking process (Dunson and Park, 2008), and so on. Moreover, in credibility theory, association of covariates is important to characterize risk features of individuals (Bühlmann and Gisler, 2005), so as to assigning more precise personalized risk premiums in this area of big data. Hence, future promising research topics may include the corresponding theory for ratemaking associated with covariates and nonparametric priors.

References

Albrecher, H., Constantinescu, C., Loisel, S., 2011. Explicit ruin formulas for models with dependence among risks. Insurance Math. Econom. 48 (2), 265–270.

Albrecher, H., Constantinescu, C., Loisel, S., Kantor, J., 2002. Simulation of ruin probabilities for risk processes of Markovian type. Monte Carlo Methods Appl. 8 (2), 111–127.

Barron, A., Schervish, M.J., Wasserman, L., 1999. The consistency of posterior distributions in nonparametric problems. Ann. Statist. 27 (2), 536–561.

Berger, J.O., Guglielmi, A., 2001. Bayesian testing of a parametric model versus nonparametric alternatives. J. Amer. Statist. Assoc. 96, 174–184.

Bolancé, C., Guillén, M., Pinquet, J., 2008. Time-varying credibility for frequency risk models: estimation and tests for autoregressive specifications on the random effects. Insurance Math. Econom. 33, 273–282.

Bühlmann, H., 1967. Experience rating and credibility. Astin Bull. 4, 199–207.

Bühlmann, H., Gisler, A., 2005. A Course in Credibility Theory and its Applications. Springer, Netherlands.

Cai, X., Wen, L., Wu, X., Zhou, X., 2015. Credibility estimation of distribution functions with applications to experience rating in general insurance. N. Am. Actuar. J. 19 (4), 311–335.

Dunson, D.B., Park, J.H., 2008. Kernel stick-breaking processes. Biometrika 95, 307–323.

Dunson, D.B., Xue, Y., Carin, L., 2008. The matrix stick-Breaking process: flexible bayes meta-analysis. J. Amer. Statist. Assoc. 103, 317–327.

Fabius, J., 1964. Asymptotic behavior of bayes estimates. Ann. Statist. 35, 846–856.

Fellingham, G.W., Kottas, A., Hartman, B.M., 2015. Bayesian nonparametric predictive modeling of group health claims. Insurance Math. Econom. 60, 1–10.

Ferguson, T.S., 1973. A bayesian analysis of some nonparametric problems. Ann. Statist. 1, 209–230.

Ferguson, T.S., 1974. Prior distributions on spaces of probability measures. Ann. Statist. 2, 615–629.

Freedman, D.A., 1963. On the asymptotic behavior of bayes estimates in the discrete case. Ann. Statist. 34, 1386–1403.

Frees, E.W., Wang, P., 2005. Credibility using copulas. N. Am. Actuar. J. 9, 31–48.

Gelfand, A.E., Smith, A.F.M., 1990. Sampling-based approaches to calculating marginal densities. J. Amer. Statist. Assoc. 85, 398–409.

Gerber, H.U., 1979. An Introduction to Mathematical Risk Theory. SS Huebner Foundation for Insurance Education, Wharton School, University of Pennsylvania.

Ghosh, J.K., Ramamoorthi, R.V., 2003. Bayesian Nonparametrics. Springer-Verlag New York, Inc.

Hanson, T.E., 2006. Inferences for mixtures of finite polya tree models. J. Amer. Statist. Assoc. 101, 1548–1565.

Hanson, T., Johnson, W.O., 2002. Modeling regression error with a mixture of polya trees. J. Amer. Statist. Assoc. 97, 1020–1033.

Hastings, W.K., 1970. Monte carlo sampling methods using markov chains and their applications. Biometrika 57, 1–97.

Heilmann, W.R., 1986. On the impact of independence of risks on stop loss premiums. Insurance Math. Econom. 5 (3), 197–199.

Hurlimann, W.S., 1993. Bivariate distributions with diatomic conditionals and stop-loss transform of random sums. Statist. Probab. Lett. 17 (1), 329–335.

Ishwaran, H., James, L.F., 2001. Gibbs sampling methods for stick-breaking priors. J. Amer. Statist. Assoc. 96, 161–173.

Kass, R., Goovaerts, M., Dhaene, J., Denuit, M., 2001. Modern Actuarial Risk Theory. Kluwer, Boston.

Lau, J.W., Siu, T.K., Yang, H., 2006. On bayesian mixture credibility. Astin Bull. 36 (02), 573–588.

Lavine, M., 1992. Some aspects of polya tree distributions for statistical modeling. Ann. Statist. 20, 1222–1235.

Lavine, M., 1994. More aspects of polya tree distributions for statistical modeling. Ann. Statist. 22, 1161–1176.

Lo, A.Y., 1984. On a class of bayesian nonparametric estimates: I. density estimates. Ann. Statist. 12 (1), 351–357.

Makov, U.E., Smith, A.F., Liu, Y.H., 1996. Bayesian methods in actuarial science. J. Roy. Statist. Soc. Ser. D Statist. 45 (4), 503–515.

Mauldin, R.D., Sudderth, M.D., Williams, S.C., 1992. Polya trees and random distributions. Ann. Statist. 20, 1203–1221.

Müller, A., 1997. Stop-loss order for portfolios of dependent risks. Insurance Math. Econom. 21 (3), 219–233.

Norberg, R., 2004. Credibility theory. In: Encyclopedia of Actuarial Science. Wiley, Chichester, UK.

Valdez, E.A., Mo, K., 2002. Ruin probabilities with dependent claims. Working paper, University of New South Wales.

Paddock, S.M., 1999. Randomized Polya Trees: Bayesian Nonparamerics for Multivariate Data Analaysis. (Unpublished doctoral thesis), Duke University.

Paddock, S.M., Ruggeri, F., Lavine, M., West, M., 2003. Randomised polya tree models for nonparametric bayesian inference. Statist. Sinica 13, 443–460.

Pan, M., Wang, R., Wu, X., 2008. On the consistency of credibility premiums regarding Esscher principle. Insurance Math. Econom. 42 (1), 119–126.

Purcaru, O., Denuit, M., 2002. On the dependence induced by frequency credibility models. Belg. Actuar. Bull. 2 (1), 7379.

Purcaru, O., Denuit, M., 2003. Dependence in dynamic claim frequency credibility models. Astin Bull. 33 (1), 23–40.

Robert, C., Casella, G., 2011. A short history of Markov Chain Monte Carlo: subjective recollections from incomplete data. Statist. Sci. 26 (1), 102–115.

Sethuraman, J., 1994. A constructive definition of dirichlet priors. Statist. Sinica 4, 639–650.

Valdez, E.A., 2014. Empirical investigation of insurance claim dependencies using mixture models. Eur. Actuar. J. 4 (1), 155–179.

Walker, S.G., Mallick, B.K., 1999. Semiparametric accelerated life time model. Biometrics 55, 477–483.

Wen, L., Wu, X.Y., Zhou, X., 2009. The credibility premiums for models with dependence induced by common effects. Insurance Math. Econom. 44, 19–25.

Wu, X., Zhou, X., 2006. A new characterization of distortion premiums via countable additivity for comonotonic risks. Insurance Math. Econom. 38 (2), 324–334.

Yeo, K.L., Valdez, E.A., 2006. Claim dependence with common effects in credibility models. Insurance Math. Econom. 38, 609–629.

Zehnwirth, B., 1979. Credibility and the dirichlet process. Scand. Actuar. J. 1, 13–23.