Annals of Actuarial Science (2024), **18**, pp. 5–29
doi:10.1017/S1748499523000076

ORIGINAL RESEARCH PAPER

On Bayesian credibility mean for finite mixture distributions

Ehsan Jahanbani, Amir T. Payandeh Najafabadi* and Khaled Masoumifard

Department of Actuarial Science, Faculty of Mathematical Sciences, Shahid Beheshti University, Evin, 1983969411, Tehran, Iran
*Corresponding author. E-mail: amirtpayandeh@sbu.ac.ir

(Received 17 July 2022; revised 23 February 2023; accepted 25 February 2023; first published online 29 March 2023)

Abstract

Consider the problem of determining the Bayesian credibility mean $E(X_{n+1}|X_1, \cdots, X_n)$, whenever the random claims $X_1, \cdots, X_n$, given parameter vector $\boldsymbol{\Psi}$, are sampled from the K-component mixture family of distributions, whose members are the union of different families of distributions. This article begins by deriving a recursive formula for such a Bayesian credibility mean. Moreover, under the assumption that using additional information $Z_{i,1}, \cdots, Z_{i,m}$, one may probabilistically determine a random claim X_i belongs to a given population (or a distribution), the above recursive formula simplifies to an exact Bayesian credibility mean whenever all components of the mixture distribution belong to the exponential families of distributions. For a situation where a 2-component mixture family of distributions is an appropriate choice for data modelling, using the logistic regression model, it shows that: how one may employ such additional information to derive the Bayesian credibility model, say Logistic Regression Credibility model, for a finite mixture of distributions. A comparison between the Logistic Regression Credibility (LRC) model and its competitor, the Regression Tree Credibility (RTC) model, has been given. More precisely, it shows that under the squared error loss function, it shows the LRC's risk function dominates the RTC's risk function at least in an interval which about 0.5. Several examples have been given to illustrate the practical application of our findings.

Keywords: Finite mixture distributions; Bayesian credibility mean; Exponential family of distributions; Logistic regression; Regression tree credibility; Dominating estimators; Squared error loss function

2020 AMS Classifications: Primary: 62F15; 62E15; 91B05; 91G99

1. Introduction

Credibility theory is an experience rating method that combines information from the collective and individual risks to obtain an accurate estimation of the premium of an insurance contract. In a situation where exact credibility can be obtained, the credibility theory is determined how much weight should be assigned to the claim history of an individual. However, in the Bayesian credibility theory, we restate our belief on risk parameters in terms of the prior distribution. Then, given the past risk experience, our belief has been updated and restated in terms of the posterior distribution. Finally, using such a posterior distribution, we derive a predictive distribution to make inferences about the future claim. In cases where the measurable space $\mathcal{X}$, or alternatively say population, is heterogeneous and can be partitioned into some finite homogenous populations, the posterior distribution and consequently the predictive distribution cannot represent in a closed form. Therefore, any inferential statistics, including the Bayesian credibility mean, about future claims cannot derive explicitly.

The history of the credibility theory began with Mowbray (1914)'s and Whitney (1918)'s papers. They suggested a convex combination $P = \zeta\bar{X} + (1-\zeta)\mu$, of collective premium, μ, and individual premium, $\bar{X}$, with credibility factor ζ, as an appropriate premium of an insurance contract. In 1950, Bailey restated this premium (well-known as an exact credibility premium) in the language of parametric Bayesian statistics. Bühlmann (1967) and Bühlmann & Straub (1970) extended the idea of the exact credibility premium to the model-based approach. After the seminal works of Bühlmann (1967) and Bühlmann & Straub (1970), the credibility theory has become very popular in most actuarial aspects. For a comprehensive discussion on various developments and methodologies in credibility theory, see Bühlmann & Gisler (2005) and Payandeh (2010). The classical credibility theory provides a relatively simple, but inflexible to mean of that predictive distribution. Hong & Martin (2017, 2018) introduced a Dirichlet process mixture model as an alternative approach to the classical credibility theory. They studied several theoretical properties and the advantages of their approach. Moreover, they compared it with the classical credibility theory. The precise choice of prior distribution in the Bayesian credibility theory has been studied by Hong & Martin (2020, 2022).

The Bayesian credibility mean under mixture distributions has been studied by several authors such as Lau *et al.* (2006), Cai *et al.* (2015), Hong & Martin (2017, 2018), Zhang *et al.* (2018), Payandeh & Sakizadeh (2019, 2023), Li *et al.* (2021), among others. All of their approaches are derived based on an approximation. For instance, (1) Payandeh & Sakizadeh (2019) approximated the complicated posterior distribution by a mixture distribution, and then, they derived an approximation for the Bayesian credibility means. Unfortunately, their approximation error rises as the number of past experiences increases; (2) Lau *et al.* (2006) following Lo (1984) restated the predictive distribution of X_{n+1} given the past claim experience $X_1, \cdots, X_n$ as a finite sum over all possible partitions of the past claim experience. Then, using the credibility premium, which is a convex combination of the collective premium (the prior mean) and the sample average of the past claim experience, to derive the Bayesian credibility mean. Certainly, under the class of the exponential family of distributions such a credibility premium coincides with the Bayesian credibility mean, see Payandeh (2010) for more details.

This article considers a random sample observation $X_1, \cdots, X_n$ from a K-component mixture distribution whit the cdf $F_X(\cdot) = \sum_{l=1}^{K} \omega_l G_l(\cdot)$, where $\sum_{l=1}^{K} \omega_l = 1$. Moreover, it assumes that for a random variable X_i, $i = 1, \cdots, n$, there is additional information $Z_{i,1}, \cdots, Z_{i,m}$, such that given such additional information, one may probabilistically determine the random variable X_i belongs to which component of the K-component mixture distribution, i.e., $P(X_i \sim G_l(\cdot)|Z_{i,1}, \cdots, Z_{i,m}) = \omega_l$. Under these assumptions, this article provides (1) the Bayesian credibility premium for such a finite mixture distribution, (2) the exact credibility premium for such finite mixture distributions, whenever populations' claim distribution belongs to the exponential family of distributions and their corresponding prior distribution conjugates with such a claim distribution, (3) a Logistic Regression Credibility model for a situation that a 2-component mixture family of distributions is an appropriate choice for data modelling, and (4) a comparison between the Logistic Regression Credibility and well-known Regression Tree Credibility model.

The rest of this article develops as the following. Section 2 collects some useful preliminaries and provides technical notations and symbols that we will use hereafter now. The main results are represented in section 3. The exact credibility mean under the class of single-parameter exponential family of distributions along with several examples has been given in section 4. For a situation that a 2-component mixture family of distributions is an appropriate choice for data modelling, section 5 suggests a probabilistic model to formulate such additional information and derive the Bayesian credibility mean for a finite mixture of distributions. Moreover, a comparison between the LRC model and its competitor, the Regression Tree Credibility (RTC) model, has been given in section 5.1. Conclusion and suggestions are given in section 6.

https://doi.org/10.1017/S1748499523000076 Published online by Cambridge University Press

2. Preliminaries

A single-parameter exponential family is a family of probability distributions whose probability density/mass function can be restated as

$$f(x|\theta) = a(x)e^{\{\phi(\theta)t(x)\}}/c(\theta)\ \ \forall\, x \in S_X, \tag{1}$$

where $a(\cdot)$, $\phi(\cdot)$, and $t(\cdot)$ are given functions, and the normalising factor $c(\cdot)$ is defined based on the fact that $\int_{S_X} f(x|\theta)dx = 1$.

By setting $\eta = -\phi(\theta)$, Jewell (1974) showed that, based upon random sample $X_1, \cdots, X_n$, and under the conjugate prior distribution

$$\pi^{conj}(\eta) = [c(\eta)]^{-\alpha_0} e^{\{-\beta_0\eta\}}/d(\alpha_0, \beta_0),$$

the Bayesian credibility can be expressed based on the sufficient statistic $t(\cdot)$ as

$$E(t(X_{n+1})|X_1, \cdots, X_n) = \zeta_n \bar{t}_n + (1 - \zeta_n)\beta_0/\alpha_0, \tag{2}$$

where the credibility factor $\zeta_n = n/(n + \alpha_0)$ and $\bar{t}_n = \sum_i^n t(x_i)/n$.

For example, for the normal distribution with given mean μ_0 and unknown variance σ^2. To imply Jewell (1974)'s findings, one may define the precision θ as $\theta = 1/\sigma^2$ and $t(x) = (x - \mu_0)^2/2$. Now by considering the Gamma conjugate prior (with parameters α_0 and β_0) for θ then get

$$E\left(\frac{(X_{n+1} - \mu_0)^2}{2}|X_1, \cdots, X_n\right) = \zeta_n \frac{\sum_{i=1}^n (x_i - \mu_0)^2/2}{n} + (1 - \zeta_n)\frac{\beta_0}{\alpha_0}. \tag{3}$$

Therefore, the Bayesian credible prediction for the variance of X_{n+1} is the linear combination of sample variance and mean of the conjugate prior.

A random variable X, given parameter vector $\mathbf{\Psi}$, has a K-component finite mixture distribution if it's corresponding cdf can be reformulated as

$$F_X(x|\mathbf{\Psi}) = \sum_{l=1}^{K} \omega_l G_l(x|\mathbf{\Psi}), \tag{4}$$

where $G_l(x|\mathbf{\Psi})$-s are some given the cdfs, $\omega_l \in [0, 1]$, for $l = 1, \cdots, K$, $\sum_{l=1}^K \omega_l = 1$.

The finite mixture distributions have proved remarkably useful in modelling an enormous variety of phenomena in a wide range of branches in climatology, demographics, economics, actuarial science, statistics, healthcare, and a mixture of expert models and engineering. Indeed, the shape of a finite mixture distribution is flexible, being able to capture, many aspects of the collected data, such as multimodality, heavy-tailed, truncated, skewness, and kurtosis, see Miljkovic & Grün (2016), Blostein & Miljkovic (2019) and de Alencar *et al.* (2021), among others, for more details. Moreover, one of the most advantages of finite mixture distributions is that they illustrate most aspects of complex systems which cannot be done by a single distribution, see McLachlan & Peel (2004), among others, for more details on mixture models. A finite mixture distribution is a simple and elementary model, but unfortunately, such simplicity does not extend to the derivation of either the maximum likelihood estimator or Bayes estimators (Lee *et al.*, 2009). In fact, based upon a random sample observation $X_1, \cdots, X_n$, the likelihood function of a K-component mixture distribution is a product of a summation, which can be turned into a sum of K^n terms. Therefore, it will be computationally too expensive to be used for more than a few observations. To overcome this problem, several attractive approaches have been introduced by the authors. For instance, Keatinge (1999) used the Karush-Kuhn-Tucker theorem to provide a maximum likelihood estimator algorithm to estimate the weights of a finite mixture of exponential distributions. Other authors employed a demarginalisation argument (or missing data approach) to assign a random variable X_i to a subgroup, using a random latent variable. Then using an EM algorithm (Dempster *et al.*, 1977) or the data augmentation algorithm (Carvajal *et al.*, 2018) to derive an

https://doi.org/10.1017/S1748499523000076 Published online by Cambridge University Press

estimation. Some authors came up with an approximation technique; for instance, Payandeh & Sakizadeh (2019) approximated the Bayesian likelihood function for a mixture distribution by a practical and appropriate distribution. Unfortunately, the accuracy of their approximation technique dramatically reduces as the number of observations increases. All of these methods are time-consuming (Frühwirth-Schnatter, 2019) or suffer from low accuracy.

A class of K-component finite mixture distributions is said to be identifiable whenever the equality of any two members $F(\cdot)$ and $F^*(\cdot)$ of this class implies: (1) equality of their components, (2) theirs weights, and (3) their cdfs. Identifiability problems for finite and countable mixtures have been widely investigated. Teicher (1960, 1963) established a necessary and sufficient condition for the identifiability of the class of finite mixture distributions. Moreover, he proved the identifiability of a class of mixture Normal (or Gamma) distributions. Atienza *et al.* (2006) showed that a class of all finite mixtures distributions generated by a union of Lognormal, Gamma, and Weibull distributions is identifiable. Unfortunately, most mixture distributions are not identifiable because they are invariant under permutations of the indices of their components. This problem is well-known as the " *label-switching* problem." The posterior distribution may also inherit such the " *label-switching* problem" from its prior distribution (Rufo *et al.*, 2006 and 2007). Under the " *label-switching* problem," there is a positive probability that at least one of the components in a finite mixture distribution does not contribute to any of the observations. Therefore, the random sample $x_1, \cdots, x_n$ does not carry any information on this component. Consequently, unknown parameter(s) of such a component cannot be estimated under *either* classical *or* Bayesian frameworks. A naïve solution to the " *label-switching* problem" is to impose some constraint on the parameter space for the classical approach (Maroufy & Marriott, 2017), and for the Bayesian approach, some constraints have been added to the prior distribution that leads to a posterior distribution that does not suffer from the " *label-switching* problem" (Marin *et al.*, 2005). Unfortunately, insufficient care in the choice of suitable identifiability constraints can lead to other problems (Rufo *et al.*, 2006 and 2007).

It is worthwhile to mention that if random variable X, given $\mathbf{\Psi}$, has the cdf function (4), one may not conclude that $P(X \in PoP_k|\mathbf{\Psi}) = \omega_k$, where $X \in PoP_k$ stands for "$X|\mathbf{\Psi} \sim F_k$." To observe this fact, consider a 2-component mixture distribution $F(x) = \omega_1 G_1(x) + \omega_2 G_2(x)$. Now for an arbitrary density function $G_3(\cdot)$, set $G_1^*(x) = G_3(x)$ and $G_2^*(x) = (\omega_1 G_1(x) + \omega_2 G_2(x) - \omega_1 G_3(x))/\omega_2$. Now observe that $F^*(x) = \omega_1 G_1^*(x) + \omega_2 G_2^*(x) = F(x)$.

Note 1. *We should note that in this article, alternatively, we use $X \in PoP_k$ instead of $X \sim G_k$.*

Suppose parameter vector $\mathbf{\Psi}$ can be restated as $\mathbf{\Psi} = (\theta_1, \theta_2, \cdots, \theta_K)$, based upon random sample $\tilde{\mathbf{X}} = (X_1, X_2, \cdots, X_n)$, the likelihood function and the posterior distribution, respectively, can be restated as

$$L\left(\mathbf{\Psi}|\tilde{\mathbf{X}} = \tilde{\mathbf{x}}\right) = \prod_{i=1}^{n} \left(\sum_{l=1}^{K} w_l g_l(x_i|\theta_l) \right) \tag{5}$$

$$\pi\left(\mathbf{\Psi}|\tilde{\mathbf{X}} = \tilde{\mathbf{x}}\right) \propto \left(\prod_{i=1}^{n} \sum_{l=1}^{K} w_l g_l(x_i|\theta_l) \right) \pi(\mathbf{\Psi}), \tag{6}$$

where $\pi(\mathbf{\Psi})$ stands for prior distribution on $\mathbf{\Psi}$ and $g_k(\cdot)$ is density function of the k^{th} component.

To derive a maximum likelihood estimation (resp. a Bayesian estimator) using Equation (5) (resp. Equation (6)), the missing data approach is the most popular method.

The following explain such an approach.

Note 2. *Suppose random variables $X_1, X_2, \cdots, X_n$ corresponding to the observed sample $x_1, x_2, \cdots, x_n$ are accompanied with latent binary random vector $\tilde{\mathbf{H}} = \left(H_{1,l}, H_{2,l}, \cdots, H_{n,l}\right)'$, for*

https://doi.org/10.1017/S1748499523000076 Published online by Cambridge University Press

$l = 1, 2, \cdots, K$, which indicating each observation arrives from which component/population, i.e., $P(X_i \in PoP_k | H_{i,k} = 1) = 1$ and $P(X_i \notin PoP_k | H_{i,k} = 0) = 1$. The likelihood function (5) and posterior distribution (6), respectively, can be restated as

$$L(\boldsymbol{\Psi}, \tilde{\boldsymbol{H}} | \tilde{\boldsymbol{x}}) = \prod_{i=1}^{n} \prod_{l=1}^{K} \left(w_l g_l(x_i | \theta_l)\right)^{H_{il}}$$

$$\pi(\boldsymbol{\Psi}, \tilde{\boldsymbol{H}} | \tilde{\boldsymbol{x}}) \propto \prod_{i=1}^{n} \prod_{l=1}^{K} \left(w_l g_l(x_i | \theta_l)\right)^{H_{il}} \pi(\boldsymbol{\Psi}).$$

Now in sth iteration of the E-step, one takes expectation with respect to conditional posterior distribution of the binary latent variable H_{il}, given observed data and update parameters at $(s-1)$th iteration.

Diebolt & Robert (1994) and Zhang *et al.* (2004) showed that such a missing data approach is very expensive from computational viewpoint.

Directly using the Likelihood function (5) or the posterior distribution (6), well-known as a combinatorial approach, see Marin *et al.* (2005) for a brief review. The combinatorial approach restates such product-summations equations as K^n summation terms. To avoid a long presentation, we use some notations or symbols which defined in Table 1.

Before we go further, we provide a simple example.

Consider a 2-component mixture distribution function with density function $\omega_1 f_1(x|\theta_1) + \omega_2 f_2(x|\theta_2)$. Moreover, suppose that we have sample observation X_1, X_2, X_3. Using Table 1's symbols, we have

$$\mathcal{S}^3 = \{1, 2, 3\};\ \mathcal{S}_0^3 = \{\emptyset\};\ \mathcal{S}_1^3 = \{\{1\}, \{2\}, \{3\}\};\ \mathcal{S}_2^3 = \{\{1, 2\}, \{1, 3\}, \{2, 3\}\};\ \mathcal{S}_3^3 = \{\{1, 2, 3\}\};$$

$$B_{11} = \{\{1\}\};\ B_{12} = \{\{2\}\};\ B_{13} = \{\{3\}\};\ B_{11}^c = \{\{2, 3\}\};\ B_{12}^c = \{\{1, 3\}\};\ B_{13}^c = \{\{1, 2\}\};\ B_{21} = \{\{1, 2\}\};$$

$$B_{22} = \{\{1, 3\}\};\ B_{23} = \{\{2, 3\}\};\ B_{21}^c = \{\{3\}\};\ B_{22}^c = \{\{2\}\};\ B_{23}^c = \{\{1\}\};\ B_{31} = \{\{1, 2, 3\}\};\ B_{31}^c = \{\emptyset\}.$$

The likelihood function can be restated as

$$\begin{aligned} L(\boldsymbol{\Psi} | \tilde{\boldsymbol{x}}) &= \Phi_3^1(0) + \Phi_3^1(1) + \Phi_3^1(2) + \Phi_3^1(3) \\ &= \omega_2^3 f_2(x_1|\theta_2) f_2(x_2|\theta_2) f_2(x_3|\theta_2) \\ &\quad + \omega_1 \omega_2^2 \big[f_1(x_1|\theta_1) f_2(x_2|\theta_2) f_2(x_3|\theta_2) + f_1(x_2|\theta_1) f_2(x_1|\theta_2) f_2(x_3|\theta_2) \\ &\quad + f_1(x_3|\theta_1) f_2(x_1|\theta_2) f_2(x_1|\theta_2))\big] \\ &\quad + \omega_1^2 \omega_2 \big[f_1(x_1|\theta_1) f_1(x_2|\theta_1) f_2(x_3|\theta_2) + f_1(x_1|\theta_1) f_1(x_3|\theta_1) f_2(x_2|\theta_2) \\ &\quad + f_1(x_2|\theta_1) f_1(x_3|\theta_1) f_2(x_1|\theta_2))\big] + \omega_1^3 f_1(x_1|\theta_1) f_1(x_2|\theta_1) f_1(x_3|\theta_1) \end{aligned}$$

It would worthwhile to mention that a given K-component mixture distribution can be reformulated as

$$f(x|\boldsymbol{\Psi}) = \omega_l g_l(x|\theta_l) + (1 - \omega_l) g^*(x|\boldsymbol{\Psi}(-l)),$$

where

$$\begin{aligned} g^*(x|\boldsymbol{\Psi}(-l)) &= \frac{\omega_1}{1-\omega_l} f_1(x|\theta_1) + \cdots + \frac{\omega_{l-1}}{1-\omega_l} f_{l-1}(x|\theta_{l-1}) + \frac{\omega_{l+1}}{1-\omega_l} f_{l+1}(x|\theta_{l+1}) \\ &\quad + \cdots + \frac{\omega_K}{1-\omega_l} f_K(x|\theta_K). \end{aligned} \tag{7}$$

This type of presentation will be employed whenever we like to just estimate the parameter of the lth component.

https://doi.org/10.1017/S1748499523000076 Published online by Cambridge University Press

Table 1. Notations and symbols.

Symbol	Definition
K	Number of components
n	Number of observations
$\tilde{\boldsymbol{X}}$	$=(X_1, X_2, ..., X_n)$
$\mathcal{S}^n$	$=\{1, 2, ..., n\}$
B_{ir}	The rth subset S^n which has exactly i elements
B_{ir}^c	Complement of B_{ir}
$\mathcal{S}_i^n$	$=\left\{B_{ir} : r=1, 2, \cdots, \binom{n}{i}\right\}$
$\tilde{\boldsymbol{X}}_{B_{ir}}$	$=(X_{k_1}, X_{k_2}, ..., X_{k_i})$, where $k_1, k_2, \cdots, k_i \in B_{ir}$
$\bar{\boldsymbol{X}}_{B_{ir}}$	$=\sum_{k\in B_{ir}} X_k/i$
$\tilde{\boldsymbol{X}}_{B_{ir}^c}$	$=(X_{k_1}, X_{k_2} \cdots, X_{k_{n-i}})$, where $k_1, k_2, \cdots, k_{n-i} \in B_{ir}^c$
$\bar{\boldsymbol{X}}_{B_{ir}^c}$	$=\sum_{k\in B_{ir}^c} X_k/(n-i)$
$\boldsymbol{\Psi}$	$=(\theta_1, \theta_2, \cdots, \theta_K)$
$\boldsymbol{\Psi}(-l)$	$=(\theta_1, \theta_2, \cdots, \theta_{l-1}, \theta_{l+1}, \cdots, \theta_K)$
ω_l	Weight of the lth component
$F_l(\cdot\|\cdot)$	The cdf for the lth component
θ_l	Parameters of the lth component
$\Phi_n^l(i)$	The likelihood function based n observations, whenever i
	of n observations belong to the lth component.
$\tilde{X}_{B_{ir}} \in PoP_l\|\theta_l$	For $k_1, k_2, \cdots, k_i \in B_{ir}$, given parameter θ_l, random variables
	$X_{k_1}, X_{k_2} \cdots, X_{k_i}$ are i.i.d with common cdf $F_l(\cdot\|\cdot)$.
$\mathbf{E}_K(\cdot)$	The expectation with respect to a $K-$component mixture distribution.

Note: $\mathcal{S}^n$, $\mathcal{S}_i^n$, B_{ir} and B_{ir}^c define on the index of observations rather than their values.

Hereafter now, we assume the K-component mixture distribution (4) is an identifiable model.

The following used the combinatorial method, to restate the likelihood function for the K-component mixture distribution (4).

Lemma 1. *Suppose that random sample $X_1, \cdots, X_n$ comes from the K-component mixture distribution (4). The likelihood function for mixtures of distributions can be restated in the following recursive manner*

$$L_K(\boldsymbol{\Psi}|\tilde{\boldsymbol{X}}) = \sum_{i=0}^{n} \omega_K^i (1-\omega_K)^{n-i} \sum_{r=1}^{\binom{n}{i}} L_{K-1}\left(\boldsymbol{\Psi}(-K)|\tilde{\boldsymbol{X}}_{B_{ir}^c}\right) \prod_{k\in B_{ir}} f_K(x_k|\theta_K),$$

where $L_{K-1}\left(\boldsymbol{\Psi}(-K)|\tilde{\boldsymbol{X}}_{B_{ir}^c}\right)$ stands for the likelihood function, based upon the density function $g^*(\cdot)$ (given by Equation (7) and random sample $\tilde{\boldsymbol{X}}_{B_{ir}^c}$.

Proof. Using the fact that

$$\tilde{\boldsymbol{X}} \in \bigcup_{l=1}^{K} PoP_l = \bigcup_{i=0}^{n} \bigcup_{r=1}^{\binom{n}{i}} \left(\tilde{\boldsymbol{X}}_{B_{ir}} \in PoP_K \,\&\, \tilde{\boldsymbol{X}}_{B_{ir}^c} \notin PoP_K\right)$$

and such partitions are disjoint, one may restate the likelihood function as

$$\begin{aligned} L_K(\boldsymbol{\Psi}|\tilde{\boldsymbol{X}}) &= \sum_{i=0}^{n}\sum_{r=1}^{\binom{n}{i}} P\Big(\tilde{\boldsymbol{X}}_{B_{ir}} \in PoP_K, \tilde{\boldsymbol{X}}_{B^c_{ir}} \notin PoP_K|\theta_K, \boldsymbol{\Psi}(-K)\Big) \\ &= \sum_{i=0}^{n}\sum_{r=1}^{\binom{n}{i}} P\big(\tilde{\boldsymbol{X}}_{B_{ir}} \in PoP_K|\theta_K\big)\, P\Big(\tilde{\boldsymbol{X}}_{B^c_{ir}} \notin PoP_K|\boldsymbol{\Psi}(-K)\Big) \\ &= \sum_{i=0}^{n}\sum_{r=1}^{\binom{n}{i}} \prod_{k\in B_{ir}} \left[\omega_K f_K(x_k|\theta_K)\right] \prod_{k\in B^c_{ir}} \left[(1-\omega_K)g^*(x_k|\boldsymbol{\Psi}(-K))\right] \\ &= \sum_{i=0}^{n} \omega_K^i (1-\omega_K)^{n-i} \sum_{r=1}^{\binom{n}{i}} L_{K-1}\Big(\boldsymbol{\Psi}(-K)|\tilde{\boldsymbol{X}}_{B^c_{ir}}\Big) \prod_{k\in B_{ir}} f_K(x_k|\theta_K). \end{aligned}$$

The second equation arrives from the assumption that given parameter vector $\boldsymbol{\Psi}$, two random samples $\tilde{\boldsymbol{X}}_{B_{ir}}$ and $\tilde{\boldsymbol{X}}_{B^c_{ir}}$ are independent. □

The Bayes estimator for a given parameter of the K-component mixture distribution (4) under the squared error loss function is given as follows.

Lemma 2. *Assume that random sample* $X_1, \cdots, X_n$ *comes from the K-component mixture distribution (4). Moreover assume that* $\pi(\theta_1, \theta_2, \cdots, \theta_K) = \prod_{j=1}^{K} \pi(\theta_j)$. *Then, the Bayesian estimator, under the square error loss function, for parameter* θ_l *is*

$$E(\Theta_l|\tilde{\boldsymbol{X}}) = \sum_{i=0}^{n}\sum_{r=1}^{\binom{n}{i}} C^{(l)}_{ir} E\big(\Theta_l|\tilde{\boldsymbol{X}}_{B_{ir}} \in PoP_l\big)\,,$$

where

$$C^{(l)}_{ir} = P\left(\tilde{\boldsymbol{X}}_{B_{ir}} \in PoP_l, \tilde{\boldsymbol{X}}_{B^c_{ir}} \notin PoP_l|\tilde{\boldsymbol{X}} \in \bigcup_{l=1}^{K} PoP_l\right).$$

Proof. The posterior distribution of $\Theta_l|\tilde{\boldsymbol{X}}$ can be restated as

$$\begin{aligned} \pi\left(\theta_l|\tilde{\boldsymbol{X}}\right) &= \sum_{i=0}^{n}\sum_{r=1}^{\binom{n}{i}} \pi\left(\theta_l|\tilde{\boldsymbol{X}}_{B_{ir}} \in PoP_l, \tilde{\boldsymbol{X}}_{B^c_{ir}} \notin PoP_l\right) P\left(\tilde{\boldsymbol{X}}_{B_{ir}} \in PoP_l, \tilde{\boldsymbol{X}}_{B^c_{ir}} \notin PoP_l|\tilde{\boldsymbol{X}} \in \bigcup_{k=1}^{K} PoP_k\right) \\ &= \sum_{i=0}^{n}\sum_{r=1}^{\binom{n}{i}} C^{(l)}_{ir}\pi\left(\theta_l|\tilde{\boldsymbol{X}}_{B_{ir}} \in PoP_l, \tilde{\boldsymbol{X}}_{B^c_{ir}} \notin PoP_l\right) \\ &= \sum_{i=0}^{n}\sum_{r=1}^{\binom{n}{i}} C^{(l)}_{ir} \frac{\int_{\psi(-l)} P(\tilde{\boldsymbol{X}}_{B_{ir}} \in PoP_l \,\&\, \tilde{\boldsymbol{X}}_{B^c_{ir}} \notin PoP_l|\theta_l, \psi(-l))\pi(\theta_l)\pi(\psi(-l))\mathbf{d}\psi(-l)}{\int_{\theta_l}\int_{\psi(-l)} P(\tilde{\boldsymbol{X}}_{B_{ir}} \in PoP_l \,\&\, \tilde{\boldsymbol{X}}_{B^c_{ir}} \notin PoP_l|\theta_l, \psi(-l))\pi(\theta_l)\pi(\psi(-l))d\theta_l\mathbf{d}\psi(-l)} \end{aligned}$$

https://doi.org/10.1017/S1748499523000076 Published online by Cambridge University Press

$$= \sum_{i=0}^{n} \sum_{r=1}^{\binom{n}{i}} C_{ir}^{(l)} \frac{\pi(\theta_l) P\big(\tilde{\mathbf{X}}_{B_{ir}} \in PoP_l | \theta_l\big) \int_{\psi(-l)} P\big(\tilde{\mathbf{X}}_{B_{ir}^c} \notin PoP_l | \psi(-l)\big)\, \pi(\psi(-l)) \mathbf{d}\psi(-l)}{\int_{\theta_l} \pi(\theta_l) P\big(\tilde{\mathbf{X}}_{B_{ir}} \in PoP_l | \theta_l\big)\, d\theta_l \int_{\psi(-l)} P\big(\tilde{\mathbf{X}}_{B_{ir}^c} \notin PoP_l | \psi(-l)\big)\, \pi(\psi(-l)) \mathbf{d}\psi(-l)}$$

$$= \sum_{i=0}^{n} \sum_{r=1}^{\binom{n}{i}} C_{ir}^{(l)} \pi\left(\theta_l | \tilde{\mathbf{X}}_{B_{ir}} \in PoP_l\right),$$

where $\int_{\psi(-l)}$ stands for $\int_{\theta_1} \cdots \int_{\theta_{l-1}} \int_{\theta_{l+1}} \cdots \int_{\theta_K}$, $\mathbf{d}\psi(-l) = d\theta_1 \cdots d\theta_{l-1} d\theta_{l+1} \cdots d\theta_K$ and $C_{ir}^{(l)} = P\left(\tilde{\mathbf{X}}_{B_{ir}} \in PoP_l, \tilde{\mathbf{X}}_{B_{ir}^c} \notin PoP_l | \tilde{\mathbf{X}} \in \bigcup_{l=1}^{K} PoP_l\right)$.

Since the Bayes estimator under the squared error loss function is the posterior expectation, we obtain the desired result. □

Now, we concentrate on the Bayesian credibility mean for the K-component mixture distribution (4).

3. A Recursive Formula for the Bayesian Credibility Mean

The Bayesian credibility mean of X_{n+1} based upon the past information $X_1, X_2, \cdots, X_n$ is

$$E(X_{n+1} | X_1, X_2, \cdots, X_n). \tag{8}$$

The following represents a recursive statement for the Bayesian credibility mean under the K-component mixture distribution (4).

Theorem 1. *Assume that the observations $X_1, \cdots, X_n$ come from the K-component mixture distribution (4). Moreover, suppose that the prior distribution $\pi(\theta_1, \theta_2, \cdots, \theta_K)$ is independent, i.e., $\pi(\theta_1, \theta_2, \cdots, \theta_K) = \prod_{k=1}^{K} \pi_k(\theta_k)$. The Bayesian credibility mean based upon such random sample and the K-component mixture distribution is*

$$\mathbf{E}_K\left(X_{n+1} | \tilde{\mathbf{X}}\right) = \sum_{i=0}^{n} \sum_{r=1}^{\binom{n}{i}} C_{ir}^{(K)} \left[\omega_K \mathbf{E}_1\left(X_{n+1} | \tilde{\mathbf{X}}_{B_{ir}} \in PoP_K\right) + (1-\omega_K) \mathbf{E}_{K-1}\left(X_{n+1} | \tilde{\mathbf{X}}_{B_{ir}^c} \notin PoP_K\right)\right],$$

where $C_{ir}^{(K)} = P\left(\tilde{\mathbf{X}}_{B_{ir}} \in PoP_K, \tilde{\mathbf{X}}_{B_{ir}^c} \notin PoP_K | \tilde{\mathbf{X}} \in \bigcup_{k=1}^{K} PoP_k\right)$.

Proof. Using the definition of $C_{ir}^{(K)}$, one may conclude that

$$\mathbf{E}_K(X_{n+1} | \tilde{\mathbf{X}}) = E\left(E\left(X_{n+1} | \theta_K, \psi(-K) | \tilde{\mathbf{X}} \in \bigcup_{l=1}^{K} PoP_l\right)\right)$$

$$= \sum_{i=0}^{n} \sum_{r=1}^{\binom{n}{i}} E\big(X_{n+1} | \left\{\tilde{\mathbf{X}}_{B_{ir}} \in PoP_K, \tilde{\mathbf{X}}_{B_{ir}^c} \notin PoP_K\right\}\big)\, P\left(\tilde{\mathbf{X}}_{B_{ir}} \in PoP_K, \tilde{\mathbf{X}}_{B_{ir}^c} \notin PoP_K | \tilde{\mathbf{X}} \in \bigcup_{l=1}^{K} PoP_l\right)$$

$$= \sum_{i=0}^{n} \sum_{r=1}^{\binom{n}{i}} C_{ir}^{(K)} E\big(E[X_{n+1} | \theta_K, \psi(-K)] \mid \left\{\tilde{\mathbf{X}}_{B_{ir}} \in PoP_K, \tilde{\mathbf{X}}_{B_{ir}^c} \notin PoP_K\right\}\big)$$

https://doi.org/10.1017/S1748499523000076 Published online by Cambridge University Press

$$= \sum_{i=0}^{n} \sum_{r=1}^{\binom{n}{i}} C_{ir}^{(K)} \left[\omega_K E\big(\mu(\Theta_K) | \tilde{\mathbf{X}}_{B_{ir}} \in PoP_K\big) + (1-\omega_K) E\big(\mu(\psi(-K)) | \tilde{\mathbf{X}}_{B_{ir}^c} \notin PoP_K\big) \right]$$

$$= \sum_{i=0}^{n} \sum_{r=1}^{\binom{n}{i}} C_{ir}^{(K)} \left[\omega_K \mathbf{E}_1 \left(X_{n+1} | \tilde{\mathbf{X}}_{B_{ir}} \in PoP_K \right) + (1-\omega_K) \mathbf{E}_{K-1} \left(X_{n+1} | \tilde{\mathbf{X}}_{B_{ir}^c} \notin PoP_K \right) \right]. \qquad \square$$

Theorem 1 provides a recursive formula to evaluate the Bayesian credibility mean. A practical application of this theorem is very expensive, to see that, please see the following example.

Teicher (1960, 1963) established the identifiability of a class of mixture Gamma distribution, using this fact, the following example provides the Bayesian credibility mean (or premium) for a class of 2-component exponential distribution with Gamma conjugate prior distributions.

Example 1. Suppose given parameter vector $\mathbf{\Psi} = (\theta_1, \theta_2)$, random sample $X_1, X_2, \cdots, X_n$ obtained from a 2-component exponential distribution with density function

$$\omega_1 Exp(\theta_1) + \omega_2 Exp(\theta_2),$$

where $\omega_1, \omega_2 \in [0, 1]$ and $\omega_1 + \omega_2 = 1$. Moreover, consider the conjugate prior $Gamma(\alpha_i, \beta_i)$ for parameter θ_i, for $i = 1, 2$. Now, we are interested in the Bayesian credibility premium under this setting.

To obtain the desired Bayesian credibility premium, we employ the result of Theorem 1. Application of this theorem arrives under the following two steps:

Step 1) $C_{ir}^{(2)} = P\Big(\tilde{\mathbf{X}}_{B_{ir}} \in PoP_2 \,\&\, \tilde{\mathbf{X}}_{B_{ir}^c} \notin PoP_2 | \tilde{\mathbf{X}} \in \bigcup_{l=1}^{2} PoP_l\Big)$,

Step 2) $\mathbf{E}_1\left(X_{n+1} | \tilde{\mathbf{X}}_{B_{ir}^c} \in PoP_1 \right)$ and $\mathbf{E}_1\left(X_{n+1} | \tilde{\mathbf{X}}_{B_{ir}} \in PoP_2 \right)$.

For Step 1 observe that:

$$C_{ir}^{(2)} = P\left(\tilde{\mathbf{X}}_{B_{ir}} \in PoP_2 \,\&\, \tilde{\mathbf{X}}_{B_{ir}^c} \notin PoP_2 | \tilde{\mathbf{X}} \in \bigcup_{l=1}^{2} PoP_l \right)$$

$$= \frac{P\Big(\tilde{\mathbf{X}}_{B_{ir}} \in PoP_2 \,\&\, \tilde{\mathbf{X}}_{B_{ir}^c} \notin PoP_2\Big)}{\sum_{i=0}^{n} \sum_{r=1}^{\binom{n}{i}} P\Big(\tilde{\mathbf{X}}_{B_{ir}} \in PoP_2 \,\&\, \tilde{\mathbf{X}}_{B_{ir}^c} \notin PoP_2\Big)}.$$

Therefore, one has to calculate

$$P\Big(\tilde{\mathbf{X}}_{B_{ir}} \in PoP_2 \,\&\, \tilde{\mathbf{X}}_{B_{ir}^c} \notin PoP_2\Big) = \int_{\theta_1} \int_{\theta_2} P\Big(\tilde{\mathbf{X}}_{B_{ir}} \in PoP_2 \,\&\, \tilde{\mathbf{X}}_{B_{ir}^c} \notin PoP_2 | \theta_1, \theta_2\Big)\, \pi(\theta_1)\pi(\theta_2) d\theta_2 d\theta_1$$

$$= \int_{\theta_1} \int_{\theta_2} \omega_2^i (1-\omega_2)^{n-i} \prod_{k \in B_{ir}} f_2(x_k|\theta_2) \prod_{k \in B_{ir}^c} f_1(x_k|\theta_1) \pi(\theta_1)\pi(\theta_2) d\theta_2 d\theta_1$$

$$= \omega_2^i (1-\omega_2)^{n-i} \Bigg[\int_0^\infty \prod_{k \in B_{ir}} \left(\theta_2 e^{-\theta_2 x_k} \right) \left(\frac{\beta_2^{\alpha_2}}{\Gamma(\alpha_2)} \theta_2^{\alpha_2 - 1} e^{-\beta_2 \theta_2} \right) d\theta_2$$

$$\times \int_0^\infty \prod_{k \in B_{ir}^c} \left(\theta_1 e^{-\theta_1 x_k} \right) \left(\frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)} \theta_1^{\alpha_1 - 1} e^{-\beta_1 \theta_1} \right) d\theta_1 \Bigg]$$

https://doi.org/10.1017/S1748499523000076 Published online by Cambridge University Press

$$= \omega_2^i (1-\omega_2)^{n-i} \left[\frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)} \frac{\Gamma(n-i+\alpha_1)}{\left((n-i)\bar{\boldsymbol{x}}_{B_{ir}^c} + \beta_1\right)^{(n-i+\alpha_1)}} \times \frac{\beta_2^{\alpha_2}}{\Gamma(\alpha_2)} \frac{\Gamma(i+\alpha_2)}{\left(i\bar{\boldsymbol{x}}_{B_{ir}} + \beta_2\right)^{(i+\alpha_2)}} \right].$$

Using the above findings, we have

$$C_{ir}^{(2)} = \frac{\omega_2^i (1-\omega_2)^{n-i} \left[\frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)} \frac{\Gamma(n-i+\alpha_1)}{\left((n-i)\bar{\boldsymbol{x}}_{B_{ir}^c} + \beta_1\right)^{(n-i+\alpha_1)}} \frac{\beta_2^{\alpha_2}}{\Gamma(\alpha_2)} \frac{\Gamma(i+\alpha_2)}{\left(i\bar{\boldsymbol{x}}_{B_{ir}} + \beta_2\right)^{(i+\alpha_2)}} \right]}{\sum_{j=0}^{n} \sum_{r=1}^{\binom{n}{j}} \omega_2^j (1-\omega_2)^{n-j} \left[\frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)} \frac{\Gamma(n-j+\alpha_1)}{\left((n-j)\bar{\boldsymbol{x}}_{B_{jr}^c} + \beta_1\right)^{(n-j+\alpha_1)}} \frac{\beta_2^{\alpha_2}}{\Gamma(\alpha_2)} \frac{\Gamma(j+\alpha_2)}{\left(j\bar{\boldsymbol{x}}_{B_{jr}} + \beta_2\right)^{(j+\alpha_2)}} \right]}.$$

Now observe that:

$$\pi\left(\theta_2 | \tilde{\boldsymbol{X}}_{B_{ir}} \in PoP_2\right) = \frac{P\left(\tilde{\boldsymbol{X}}_{B_{ir}} \in PoP_2 | \theta_2\right) \pi(\theta_2)}{\int_{\theta_2} P\left(\tilde{\boldsymbol{X}}_{B_{ir}} \in PoP_2 | \theta_2\right) \pi(\theta_2) d\theta_2}$$

$$= \frac{\prod_{k \in B_{ir}} \left(\theta_2 e^{-\theta_2 x_k}\right) \left(\frac{\beta_2^{\alpha_2}}{\Gamma(\alpha_2)} \theta_2^{\alpha_2 - 1} e^{-\beta_2 \theta_2}\right)}{\int_0^\infty \prod_{k \in B_{ir}} \left(\theta_2 e^{-\theta_2 x_k}\right) \left(\frac{\beta_2^{\alpha_2}}{\Gamma(\alpha_2)} \theta_2^{\alpha_2 - 1} e^{-\beta_2 \theta_2}\right) d\theta_2}$$

$$= \frac{\frac{\beta_2^{\alpha_2}}{\Gamma(\alpha_2)} \theta_2^{i+\alpha_2-1} e^{-\theta_2\left(\sum_{k \in B_{ir}} x_k + \beta_2\right)}}{\int_0^\infty \frac{\beta_2^{\alpha_2}}{\Gamma(\alpha_2)} \theta_2^{i+\alpha_2-1} e^{-\theta_2\left(\sum_{k \in B_{ir}} x_k + \beta_2\right)} d\theta_2}$$

$$= \frac{\left(i\bar{\boldsymbol{x}}_{B_{ir}} + \beta_2\right)^{(i+\alpha_2)}}{\Gamma(i+\alpha_2)} \theta_2^{(i+\alpha_2-1)} e^{-\theta_2\left(i\bar{\boldsymbol{x}}_{B_{ir}} + \beta_2\right)}.$$

One may similarly calculate $\pi(\theta_1 | \left(\tilde{\boldsymbol{X}}_{B_{ir}^c} \in PoP_1\right)$.

Now, we move to Step 2.

$$\mathbf{E}_1\left(X_{n+1} | \tilde{\boldsymbol{X}}_{B_{ir}} \in PoP_2\right) = E\left(E(X_{n+1} | \theta_2) | \tilde{\boldsymbol{X}}_{B_{ir}} \in PoP_2\right)$$

$$= E\left(\frac{1}{\Theta_2} | \tilde{\boldsymbol{X}}_{B_{ir}} \in PoP_2\right)$$

$$= \int_0^\infty \frac{1}{\theta_2} \pi\left(\theta_2 | \tilde{\boldsymbol{X}}_{B_{ir}} \in PoP_2\right) d\theta_2$$

$$= \frac{i\bar{\boldsymbol{x}}_{B_{ir}} + \beta_2}{i + \alpha_2}$$

$$= \left[\frac{i}{i+\alpha_2} \bar{\boldsymbol{x}}_{B_{ir}} + \left(1 - \frac{i}{i+\alpha_2}\right) \frac{\beta_2}{\alpha_2}\right].$$

https://doi.org/10.1017/S1748499523000076 Published online by Cambridge University Press

Table 2. Number of combinators that one has to calculate for Equation (9).

Sample size	Number of combinators
5	31
10	1,023
20	1,048,575
30	1,073,741,823
50	1.1259e+15
100	1.267651e+30

Similarly,

$$\mathbf{E}_1\left(X_{n+1}|\tilde{\boldsymbol{X}}_{B^c_{ir}} \notin PoP_2\right) = \left[\frac{n-i}{n-i+\alpha_1}\bar{\boldsymbol{x}}_{B^c_{ir}} + \left(1-\frac{n-i}{n-i+\alpha_1}\right)\frac{\beta_1}{\alpha_1}\right].$$

Therefore, using Theorem 1 the Bayesian credibility premium is

$$\begin{aligned} E(X_{n+1}|X_1,\cdots,X_n) = &\sum_{i=0}^{n}\sum_{r=1}^{\binom{n}{i}} C^{(2)}_{ir}\omega_2\left[\frac{i}{i+\alpha_2}\bar{\boldsymbol{x}}_{B_{ir}} + \left(1-\frac{i}{i+\alpha_2}\right)\frac{\beta_2}{\alpha_2}\right] \\ &+\sum_{i=0}^{n}\sum_{r=1}^{\binom{n}{i}} C^{(2)}_{ir}(1-\omega_2)\left[\frac{n-i}{n-i+\alpha_1}\bar{\boldsymbol{x}}_{B^c_{ir}} + \left(1-\frac{n-i}{n-i+\alpha_1}\right)\frac{\beta_1}{\alpha_1}\right]. \quad (9)\end{aligned}$$

It is worthwhile mentioning that, in a situation that $\omega_1 = 1$ (or $\omega_2 = 0$), the summation $\sum_{i=0}^{n}\sum_{r=1}^{\binom{n}{i}} C^{(2)}_{ir}$ just valid for $i = 0$ which $C^{(2)}_{01} = 1$. Therefore, under this setting the Bayesian credibility premium, given by Equation (9), is the well-known Bayesian credibility premium under the Exponential-Gamma assumption, in the other words,

$$E(X_{n+1}|X_1,\cdots,X_n) = \frac{n}{n+\alpha_1}\bar{\boldsymbol{x}}_{\mathcal{S}^n} + \left(1-\frac{n}{n+\alpha_1}\right)\frac{\beta_1}{\alpha_1},$$

where $\bar{\boldsymbol{x}}_{\mathcal{S}^n} = \bar{\boldsymbol{x}} = \sum_{k=1}^{n} x_k/n$.

The combinatorial object in the Bayesian credibility mean (see Equation (9)) for Example 1 makes it very hard to use. Table 2 represents the number of combinators one has to be considered, whenever he/she would like to use Equation (9). As one may observe, implementation of Equation (9) even for sample size $n = 30$ is very expensive and cannot be done with a regular computer.

To remove such barrier, we have two possibilities:

- Approximate $C^{(l)}_{ir}$ by a function which just depends on i an l
- Impose some restriction on our problem such that $C^{(l)}_{ir}$ does not depend on r.

Somehow, the first approach has been employed by Lau *et al.* (2006). They employed the sampling scheme based on a weighted Chinese Restaurant algorithm to estimate the Bayesian credibility for the infinite mixture model from observed data.

The next section considers a situation where the above recursive formula is simplified and the exact Bayesian credibility mean is obtained.

https://doi.org/10.1017/S1748499523000076 Published online by Cambridge University Press

4. Exact Bayesian Credibility Mean

Hereafter now, we follow the second approach. Therefore, we consider the following model assumption.

Model Assumption 1. *Suppose given parameter vector* $\boldsymbol{\Psi}$, *random variables* $X_1, \cdots, X_n$ *are i.i.d. Moreover suppose that there is an additional information* $Z_{i,1}, \cdots, Z_{i,m}$ *where given such information random variable* X_i, *with probability* ω_l, *has the cdf* $G_l(\cdot)$, *for* $l = 1, 2, \cdots, K$, *where* $\sum_{l=1}^{K} \omega_l = 1$.

The following lemma shows that, under the above model assumption, random variables $X_1, \cdots, X_n$ are a member of the K-component mixture distributions (4).

Lemma 3. *Under Model Assumption 1, given* $\boldsymbol{\Psi}$, *random variables* $X_1, \cdots, X_n$ *are a member of the K-component mixture distributions (4).*

Proof. Under Model Assumption 1 given $\boldsymbol{\Psi}$, random variables $X_1, \cdots, X_n$ are i.i.d. Therefore, we just need to find distribution of the random variable X_1

$$\begin{aligned} F_{X_1|\boldsymbol{\Psi}}(t) &= P(X_1 \leq t|\boldsymbol{\Psi}) \\ &= \sum_{j=1}^{K} P\big(X_1 \leq t|X_1 \in PoP_j, \boldsymbol{\Psi}\big)\, \omega_j \\ &= \sum_{j=1}^{K} \omega_j G_j(t). \end{aligned}$$

□

Another useful property of Model Assumption 1 has been given by the following.

Lemma 4. *Under Model Assumption 1, the* $C_{ir}^{(l)}$ *defined in Theorem 1 can be simplified as*

$$C_{ir}^{(l)} = \omega_l^i (1 - \omega_l)^{n-i}.$$

Proof. Conditioning the $C_{ir}^{(l)}$ on $\boldsymbol{\Psi}$, one may restate

$$\begin{aligned} C_{ir}^{(l)} &= P\left(\tilde{\mathbf{X}}_{B_{ir}} \in PoP_l \,\&\, \tilde{\mathbf{X}}_{B_{ir}^c} \notin PoP_l | \tilde{\mathbf{X}} \in \bigcup_{k=1}^{K} PoP_k\right) \\ &= \int_{\boldsymbol{\Psi}} P\Big(\tilde{\mathbf{X}}_{B_{ir}} \in PoP_l, \tilde{\mathbf{X}}_{B_{ir}^c} \notin PoP_l | \boldsymbol{\Psi}, X_1, X_2, \cdots, X_n\Big)\, \pi(\boldsymbol{\Psi}|X_1, X_2, \cdots, X_n)\, d\boldsymbol{\Psi} \\ &= \int_{\boldsymbol{\Psi}} P\big(\tilde{\mathbf{X}}_{B_{ir}} \in PoP_l|\theta_l\big)\, P\Big(\tilde{\mathbf{X}}_{B_{ir}^c} \notin PoP_l|\boldsymbol{\Psi}(-l)\Big)\, \pi(\boldsymbol{\Psi}|X_1, X_2, \cdots, X_n)\, d\boldsymbol{\Psi} \\ &= \int_{\boldsymbol{\Psi}} [P(X_1 \in PoP_l|\theta_l)]^i \big[P\big(X_1 \notin PoP_l|\boldsymbol{\Psi}(-l)\big)\big]^{n-i}\, \pi(\boldsymbol{\Psi}|X_1, X_2, \cdots, X_n)\, d\boldsymbol{\Psi} \\ &= \omega_l^i\, (1 - \omega_l)^{n-i} \int_{\boldsymbol{\Psi}} \pi(\boldsymbol{\Psi}|X_1, X_2, \cdots, X_n)\, d\boldsymbol{\Psi} \\ &= \omega_l^i (1 - \omega_l)^{n-i}. \end{aligned}$$

The last two equations arrive from the fact that $P(X_1 \in PoP_j|\boldsymbol{\Psi}) = \omega_j$ and the posterior distribution $\pi(\boldsymbol{\Psi}|X_1, X_2, \cdots, X_n)$ is a proper distribution. □

https://doi.org/10.1017/S1748499523000076 Published online by Cambridge University Press

Under Model Assumption, 1's result of Theorem 1 can be simplified as follows.

Corollary 1. *Under Model Assumption 1, the Bayesian credibility mean is*

$$\mathbf{E}_K(X_{n+1}|\tilde{\mathbf{X}}) = \sum_{i=0}^{n}\sum_{r=1}^{\binom{n}{i}} \omega_K^i (1-\omega_K)^{n-i}\left[\omega_K \mathbf{E}_1\left(X_{n+1}|\tilde{\mathbf{X}}_{B_{ir}} \in PoP_K\right)\right.$$
$$\left. +(1-\omega_K)\mathbf{E}_{K-1}\left(X_{n+1}|\tilde{\mathbf{X}}_{B^c_{ir}} \notin PoP_K\right)\right].$$

Now, by several examples, we develop the Bayesian credibility mean under the single-parameter exponential family of distributions.

For simplicity in presentation, hereafter now, we just consider the single-parameter exponential family of distributions, given by Equation (1), with $\phi(\theta) = -\theta$ for some possible extension of our finding see section 5.

Before move further, it would be useful to observe that

$$\begin{aligned}
\sum_{r=1}^{\binom{n}{i}} i\bar{\boldsymbol{x}}_{B_{ir}} &= i\left[\bar{x}_{B_{i,1}} + \bar{x}_{B_{i,2}} + \cdots + \bar{x}_{B_{i,\binom{n}{i}}}\right] \\
&= i\left[\frac{\overbrace{x_1 + x_2 + \cdots + x_k}^{i \text{ observations}}}{i} + \cdots + \frac{\overbrace{x_1 + x_2 + \cdots + x_{k'}}^{i \text{ observations}}}{i}\right] \\
&= \binom{n-1}{i-1}\sum_{i=1}^{n} x_i \\
&= \binom{n}{i} i\bar{x}.
\end{aligned} \tag{10}$$

Identifiability of a class of mixture of normal distributions has been established by Teicher (1960, 1963). Therefore, we may consider the following example.

Example 2. Suppose that under Model Assumption 1, the random sample $X_1, X_2, \cdots, X_n$, given parameter vector $\mathbf{\Psi} = (\theta_1, \theta_2, \theta_3)'$, has been distributed according the following 3-component normal mixture distribution

$$\omega_1 N\left(\theta_1, \sigma_1^2\right) + \omega_2 N\left(\theta_2, \sigma_2^2\right) + \omega_3 N\left(\theta_3, \sigma_3^2\right),$$

where $\sigma_1^2, \sigma_2^2\ \sigma_3^2$ are given, $\omega_1, \omega_2, \omega_3 \in [0, 1]$ and $\omega_1 + \omega_2 + \omega_3 = 1$.

Moreover, suppose that, for $l = 1, 2, 3$, θ_l has the conjugate prior distribution $N\left(\mu_l, b_l^2\right)$.

Now using result of Corollary 1, the recursive Bayesian credibility mean/premium is

$$\mathbf{E}_3(X_{n+1}|\tilde{\mathbf{X}}) = \sum_{i=0}^{n}\sum_{r=1}^{\binom{n}{i}} \omega_3^i (1-\omega_3)^{n-i}\left[\omega_3 \mathbf{E}_1\left(X_{n+1}|\tilde{\mathbf{X}}_{B_{ir}} \in PoP_3\right)\right.$$
$$\left. + (1-\omega_3)\mathbf{E}_2\left(X_{n+1}|\tilde{\mathbf{X}}_{B^c_{ir}} \notin PoP_3\right)\right]. \tag{11}$$

https://doi.org/10.1017/S1748499523000076 Published online by Cambridge University Press

Another application of Theorem 1 leads to

$$\begin{aligned}\mathbf{E}_2\left(X_{n+1}|\tilde{\mathbf{X}}_{B^c_{ir}}\notin PoP_3\right)&=\sum_{d=0}^{n-i}\sum_{e=1}^{\binom{n-i}{d}}C^{(2)}_{de}\left(\frac{\omega_2}{1-\omega_3}\mathbf{E}_1\left(X_{n+1}|\tilde{\mathbf{X}}_{B_{de}}\in PoP_2\right)\right.\\&\quad\left.+\frac{\omega_1}{1-\omega_3}\mathbf{E}_1\left(X_{n+1}|\tilde{\mathbf{X}}_{B^c_{de}}\in PoP_1\right)\right)\\&=\sum_{d=0}^{n-i}\sum_{e=1}^{\binom{n-i}{d}}\left(\frac{\omega_2}{1-\omega_3}\right)^{d}\left(\frac{\omega_1}{1-\omega_3}\right)^{n-i-d}\frac{\omega_2}{1-\omega_3}\mathbf{E}_1\left(X_{n+1}|\tilde{\mathbf{X}}_{B_{de}}\in PoP_2\right)\\&\quad+\sum_{d=0}^{n-i}\sum_{e=1}^{\binom{n-i}{d}}\left(\frac{\omega_2}{1-\omega_3}\right)^{d}\left(\frac{\omega_1}{1-\omega_3}\right)^{n-i-d}\frac{\omega_1}{1-\omega_3}\mathbf{E}_1\left(X_{n+1}|\tilde{\mathbf{X}}_{B^c_{de}}\in PoP_1\right).\end{aligned}$$

The exact Bayesian credibility mean under a 1-component normal mixture distribution helps us to conclude

$$\mathbf{E}_1\left(X_{n+1}|\tilde{\mathbf{X}}_{B_{ir}}\in PoP_3\right)=\frac{ib_3^2}{ib_3^2+\sigma_3^2}\bar{x}_{B_{ir}}+\frac{\sigma_3^2}{ib_3^2+\sigma_3^2}\mu_3,\ \textit{for } i=0,\cdots,n,$$

$$\mathbf{E}_1\left(X_{n+1}|\tilde{\mathbf{X}}_{B_{de}}\in PoP_2\right)=\frac{db_2^2}{db_2^2+\sigma_2^2}\bar{x}_{B_{de}}+\frac{\sigma_2^2}{db_2^2+\sigma_2^2}\mu_2,\ \textit{for } d=0,\cdots,n-i,$$

$$\mathbf{E}_1\left(X_{n+1}|\tilde{\mathbf{X}}_{B^c_{de}}\in PoP_1\right)=\frac{(n-i-d)b_1^2}{(n-i-d)b_1^2+\sigma_1^2}\bar{x}_{B^c_{de}}+\frac{\sigma_1^2}{(n-i-d)b_1^2+\sigma_1^2}\mu_1,\ \textit{for } d=0,\cdots,n-i,$$

see Bühlmann & Gisler (2005), among others for more details.

Substituting the above findings in Equation (11) and an application of Equation (10), the Bayesian credibility mean $\mathbf{E}_3(X_{n+1}|\tilde{\mathbf{X}})$ can be restated as

$$\begin{aligned}&\sum_{i=0}^{n}\sum_{r=1}^{\binom{n}{i}}\omega_3^i(1-\omega_3)^{n-i}\omega_3\left[\frac{ib_3^2}{ib_3^2+\sigma_3^2}\bar{x}_{B_{ir}}+\frac{\sigma_3^2}{ib_3^2+\sigma_3^2}\mu_3\right]\\&\quad+\sum_{i=0}^{n}\sum_{r=1}^{\binom{n}{i}}\omega_3^i(1-\omega_3)^{n-i}(1-\omega_3)\sum_{d=0}^{n-i}\sum_{e=1}^{\binom{n-i}{d}}\left(\frac{\omega_2}{1-\omega_3}\right)^{d}\left(\frac{\omega_1}{1-\omega_3}\right)^{n-i-d}\\&\qquad\times\frac{\omega_2}{1-\omega_3}\left[\frac{db_2^2}{db_2^2+\sigma_2^2}\bar{x}_{B_{de}}+\frac{\sigma_2^2}{db_2^2+\sigma_2^2}\mu_2\right]\\&\quad+\sum_{i=0}^{n}\sum_{r=1}^{\binom{n}{i}}\omega_3^i(1-\omega_3)^{n-i}(1-\omega_3)\sum_{d=0}^{n-i}\sum_{e=1}^{\binom{n-i}{d}}\left(\frac{\omega_2}{1-\omega_3}\right)^{d}\left(\frac{\omega_1}{1-\omega_3}\right)^{n-i-d}\\&\qquad\times\frac{\omega_1}{1-\omega_3}\left[\frac{(n-i-d)b_1^2}{(n-i-d)b_1^2+\sigma_1^2}\bar{x}_{B^c_{de}}\frac{\sigma_1^2}{(n-i-d)b_1^2+\sigma_1^2}\mu_1\right]\\&=\omega_3\sum_{i=0}^{n}\omega_3^i(1-\omega_3)^{n-i}\left[\frac{ib_3^2}{ib_3^2+\sigma_3^2}\binom{n}{i}\bar{x}+\frac{\sigma_3^2}{ib_3^2+\sigma_3^2}\binom{n}{i}\mu_3\right]\end{aligned}$$

https://doi.org/10.1017/S1748499523000076 Published online by Cambridge University Press

$$+\omega_2\sum_{i=0}^{n}\omega_3^i(1-\omega_3)^{n-i}\sum_{d=0}^{n-i}\left(\frac{\omega_2}{1-\omega_3}\right)^d\left(\frac{\omega_1}{1-\omega_3}\right)^{n-i-d}\left[\frac{db_2^2}{db_2^2+\sigma_2^2}\binom{n-i}{d}\binom{n}{i}\bar{x}\right.$$
$$\left.+\frac{\sigma_2^2}{db_2^2+\sigma_2^2}\binom{n-i}{d}\binom{n}{i}\mu_2\right]$$
$$+\omega_1\sum_{i=0}^{n}\omega_3^i(1-\omega_3)^{n-i}\sum_{d=0}^{n-i}\left(\frac{\omega_2}{1-\omega_3}\right)^d\left(\frac{\omega_1}{1-\omega_3}\right)^{n-i-d}\left[\frac{(n-i-d)b_1^2}{(n-i-d)b_1^2+\sigma_1^2}\binom{n-i}{d}\binom{n}{i}\bar{x}\right.$$
$$\left.+\frac{\sigma_1^2}{(n-i-d)b_1^2+\sigma_1^2}\binom{n-i}{d}\binom{n}{i}\mu_1\right]$$
$$=\omega_3[\zeta_3\bar{x}+(1-\zeta_3)\mu_3]+\omega_2[\zeta_2\bar{x}+(1-\zeta_2)\mu_2]+\omega_1[\zeta_1\bar{x}+(1-\zeta_1)\mu_1],$$

where

$$\zeta_1=\sum_{i=0}^{n}\omega_3^i(1-\omega_3)^{n-i}\sum_{d=0}^{n-i}\left(\frac{\omega_2}{1-\omega_3}\right)^d\left(\frac{\omega_1}{1-\omega_3}\right)^{n-i-d}\frac{(n-i-d)b_1^2}{(n-i-d)b_1^2+\sigma_1^2}\binom{n-i}{d}\binom{n}{i}$$
$$\zeta_2=\sum_{i=0}^{n}\omega_3^i(1-\omega_3)^{n-i}\sum_{d=0}^{n-i}\left(\frac{\omega_2}{1-\omega_3}\right)^d\left(\frac{\omega_1}{1-\omega_3}\right)^{n-i-d}\frac{db_2^2}{db_2^2+\sigma_2^2}\binom{n-i}{d}\binom{n}{i}$$
$$\zeta_3=\sum_{i=0}^{n}\omega_3^i(1-\omega_3)^{n-i}\binom{n}{i}\frac{ib_3^2}{ib_3^2+\sigma_3^2}.$$

To show application of recursive formula represented in Theorem 1, the following considers a 4-component mixture distribution.

Example 3. Suppose that under Model Assumption 1, the random sample $X_1, X_2, \cdots, X_n$, given parameter vector $\boldsymbol{\Psi}=(\theta_1,\theta_2,\theta_3,\theta_4)'$, has been distributed according the following 4-component normal mixture distribution

$$\omega_1N\big(\theta_1,\sigma_1^2\big)+\omega_2N\big(\theta_2,\sigma_2^2\big)+\omega_3N\big(\theta_3,\sigma_3^2\big)+\omega_4N\big(\theta_4,\sigma_4^2\big),$$

where for $l=1,2,3,4$, variance σ_l^2, are given, $\omega_l\in[0,1]$ and $\omega_1+\omega_2+\omega_3+\omega_4=1$.

Moreover, suppose that, for $l=1,2,3,4$, θ_l has the conjugate prior distribution $N\big(\mu_l,b_l^2\big)$.

Now an application of Corollary 1 leads to the following Bayesian credibility mean

$$\mathbf{E}_4(X_{n+1}|\tilde{\mathbf{X}})=\sum_{i=0}^{n}\sum_{r=1}^{\binom{n}{i}}\omega_4^i(1-\omega_4)^{n-i}\left[\omega_4\mathbf{E}_1(X_{n+1}|\tilde{\mathbf{X}}_{B_{ir}}\in PoP_4)\right.$$
$$\left.+(1-\omega_4)\mathbf{E}_3\left(X_{n+1}|\tilde{\mathbf{X}}_{B_{ir}^c}\notin PoP_4\right)\right]. \tag{12}$$

And again, application of Corollary 1 leads to

$$\mathbf{E}_3\left(X_{n+1}|\tilde{\mathbf{X}}_{B_{ir}^c}\notin PoP_4\right)=\sum_{d=0}^{n-i}\sum_{e=1}^{\binom{n-i}{d}}\left(\frac{\omega_3}{1-\omega_4}\right)^d\left(\frac{1-\omega_3-\omega_4}{1-\omega_4}\right)^{n-i-d}\frac{\omega_3}{1-\omega_4}\mathbf{E}_1\left(X_{n+1}|\tilde{\mathbf{X}}_{B_{de}}\in PoP_3\right)$$
$$+\sum_{d=0}^{n-i}\sum_{e=1}^{\binom{n-i}{d}}\left(\frac{\omega_3}{1-\omega_4}\right)^d\left(\frac{1-\omega_3-\omega_4}{1-\omega_4}\right)^{n-i-d}$$
$$\frac{1-\omega_3-\omega_4}{1-\omega_4}\mathbf{E}_2\left(X_{n+1}|\tilde{\mathbf{X}}_{B_{de}^c}\notin PoP_3\right)$$

https://doi.org/10.1017/S1748499523000076 Published online by Cambridge University Press

$$\mathbf{E}_2\left(X_{n+1}|\tilde{\mathbf{X}}_{B^c_{de}} \notin PoP_3\right) = \sum_{h=0}^{n-i-d} \sum_{o=1}^{\binom{n-i-d}{h}} \left(\frac{\omega_2}{1-\omega_3-\omega_4}\right)^h \left(\frac{\omega_1}{1-\omega_3-\omega_4}\right)^{n-i-d-h} \frac{\omega_2}{1-\omega_3-\omega_4}\mathbf{E}_1\left(X_{n+1}|\tilde{\mathbf{X}}_{B_{ho}} \in PoP_2\right)$$
$$+ \sum_{h=0}^{n-i-d} \sum_{e=1}^{\binom{n-i-d}{h}} \left(\frac{\omega_2}{1-\omega_3-\omega_4}\right)^h \left(\frac{\omega_1}{1-\omega_3-\omega_4}\right)^{n-i-d-h} \frac{\omega_1}{1-\omega_3-\omega_4}\mathbf{E}_1\left(X_{n+1}|\tilde{\mathbf{X}}_{B^c_{ho}} \in PoP_1\right).$$

The exact credibility mean is well-known for a 1-component normal mixture distribution (Bühlmann & Gisler, 2005), using this knowledge, we may have

$$\mathbf{E}_1\left(X_{n+1}|\tilde{\mathbf{X}}_{B_{ir}} \in PoP_4\right) = \frac{ib_4^2}{ib_4^2+\sigma_4^2}\bar{x}_{B_{ir}} + \frac{\sigma_4^2}{ib_4^2+\sigma_4^2}\mu_4, \ \ for\ i = 0, \cdots, n,$$
$$\mathbf{E}_1\left(X_{n+1}|\tilde{\mathbf{X}}_{B_{de}} \in PoP_3\right) = \frac{db_3^2}{db_3^2+\sigma_3^2}\bar{x}_{B_{de}} + \frac{\sigma_3^2}{db_3^2+\sigma_3^2}\mu_3, \ \ for\ d = 0, \cdots, n-i,$$
$$\mathbf{E}_1\left(X_{n+1}|\tilde{\mathbf{X}}_{B_{ho}} \in PoP_2\right) = \frac{hb_2^2}{hb_2^2+\sigma_2^2}\bar{x}_{B_{ho}} + \frac{\sigma_2^2}{hb_2^2+\sigma_2^2}\mu_2, \ \ for\ h = 0, \cdots, n-i-d,$$
$$\mathbf{E}_1\left(X_{n+1}|\tilde{\mathbf{X}}_{B^c_{ho}} \in PoP_1\right) = \frac{(n-i-d-h)b_1^2}{(n-i-d-h)b_1^2+\sigma_1^2}\bar{x}_{B^c_{de}} + \frac{\sigma_1^2}{(n-i-d-h)b_1^2+\sigma_1^2}\mu_1, \ \ for\ h = 0, \cdots, n-i-d.$$

Putting the above findings in Equation (11), the Bayesian credibility mean is

$$\mathbf{E}_4\left(X_{n+1}|\tilde{\mathbf{X}}\right) = \omega_4[\zeta_4\bar{x} + (1-\zeta_4)\mu_4] + \omega_3[\zeta_3\bar{x} + (1-\zeta_3)\mu_3] + \omega_2[\zeta_2\bar{x} + (1-\zeta_2)\mu_2] + \omega_1[\zeta_1\bar{x} + (1-\zeta_1)\mu_1],$$

where

$$\zeta_1 = \sum_{i=0}^{n} \omega_4^i(1-\omega_4)^{n-i} \sum_{d=0}^{n-i} \left(\frac{\omega_3}{1-\omega_4}\right)^d \left(\frac{1-\omega_3-\omega_4}{1-\omega_4}\right)^{n-i-d} \times \sum_{h=0}^{n-i-d} \left(\frac{\omega_2}{1-\omega_3-\omega_4}\right)^h \left(\frac{\omega_1}{1-\omega_3-\omega_4}\right)^{n-i-d-h} \times \frac{(n-i-d-h)b_1^2}{(n-i-d-h)b_1^2+\sigma_1^2}\binom{n}{i}\binom{n-i}{d}\binom{n-i-d}{h}$$
$$\zeta_2 = \sum_{i=0}^{n} \omega_4^i(1-\omega_4)^{n-i} \sum_{d=0}^{n-i} \left(\frac{\omega_3}{1-\omega_4}\right)^d \left(\frac{1-\omega_3-\omega_4}{1-\omega_4}\right)^{n-i-d}$$

https://doi.org/10.1017/S1748499523000076 Published online by Cambridge University Press

$$\times \sum_{h=0}^{n-i-d} \left(\frac{\omega_2}{1-\omega_3-\omega_4}\right)^h \left(\frac{\omega_1}{1-\omega_3-\omega_4}\right)^{n-i-d-h} \frac{hb_2^2}{hb_2^2+\sigma_2^2} \binom{n}{i} \binom{n-i}{d} \binom{n-i-d}{h}$$

$$\zeta_3 = \sum_{i=0}^{n} \omega_4^i (1-\omega_4)^{n-i} \sum_{d=0}^{n-i} \left(\frac{\omega_3}{1-\omega_4}\right)^d \left(\frac{1-\omega_3-\omega_4}{1-\omega_4}\right)^{n-i-d} \frac{db_3^2}{db_3^2+\sigma_3^2} \binom{n}{i} \binom{n-i}{d}$$

$$\zeta_4 = \sum_{i=0}^{n} \omega_4^i (1-\omega_4)^{n-i} \frac{ib_4^2}{ib_4^2+\sigma_4^2} \binom{n}{i}.$$

In the above two examples, we just consider a situation, in which all elements of the mixture distributions belong to a family of distribution. The following example considers a case that the mixture distributions are the union of different distributions. Using Atienza *et al.* (2006)'s method, we established (but for briefness we eliminate its proof) that a mixture union of Gamma, Lognormal and Weibull distributions constructs a class of identifiable distributions. Therefore, without any concern about identifiability, we may consider the following example.

Example 4. Suppose that under Model Assumption 1, the random sample $X_1, X_2, \cdots, X_n$, given parameter vector $\boldsymbol{\Psi} = (\theta_1, \theta_2, \theta_3)$, has been distributed according the following distribution

$$\omega_1 Gamma(\alpha, \theta_1) + \omega_2 LN\left(\theta_2, \sigma_0^2\right) + \omega_3 Weibull(\theta_3, \lambda),$$

where parameters $\alpha, \sigma_0^2, \lambda$ are given and given weights ω_1, ω_2 and ω_3 satisfy $\omega_1 + \omega_2 + \omega_3 = 1$.

Moreover, suppose that θ_1, θ_2 and $1/\theta_3$, respectively, have the conjugate prior distribution $Gamma(\alpha_1, \beta_1)$, $N\left(\mu_2, b_2^2\right)$ and $Gamma(\alpha_3, \beta_3)$.

It is well-known that the exact Bayesian credibility mean for a 1-component Gamma mixture, a 1-component Lognormal mixture and a 1-component Weibull mixture distributions, respectively, are

$$\mathbf{E}_1\left(X_{n+1}|\tilde{\mathbf{X}}_{B_{ir}} \in PoP_3\right) = \frac{\left(\sum_{k\in B_{ir}} x_k^\lambda + \beta_2\right)}{(i+r_2-1)}, \ for\ i = 0, \cdots, n,$$

$$\mathbf{E}_1\left(X_{n+1}|\tilde{\mathbf{X}}_{B_{de}} \in PoP_2\right) = \frac{b_2^2 \sum_{k\in B_{de}} Ln(x_k) + \mu_2\sigma_2^2}{db_2^2+\sigma_2^2}, \ for\ d = 0, \cdots, n-i,$$

$$\mathbf{E}_1\left(X_{n+1}|\tilde{\mathbf{X}}_{B_{de}^c} \in PoP_2\right) = \frac{\sum_{k\in B_{de}^c} x_k + \beta}{(n-i-d)\alpha_1 + r_1} \ for\ d = 0, \cdots, n-i,$$

see Bühlmann & Gisler (2005), among others, for more details.

Using the above results along with double applications of Corollary 1, the Bayesian credibility mean is

$$\begin{aligned}
\mathbf{E}_3(X_{n+1}|\tilde{\mathbf{X}}) &= \omega_3 \sum_{i=0}^{n} \sum_{r=1}^{\binom{n}{i}} \omega_3^i (1-\omega_3)^{n-i} \mathbf{E}_1\left(X_{n+1}|\tilde{\mathbf{X}}_{B_{ir}} \in PoP_3\right) \\
&\quad + (1-\omega_3) \sum_{i=0}^{n} \sum_{r=1}^{\binom{n}{i}} \omega_3^i (1-\omega_3)^{n-i} \mathbf{E}_2\left(X_{n+1}|\tilde{\mathbf{X}}_{B_{ir}^c} \notin PoP_3\right) \\
&= \omega_3 \sum_{i=0}^{n} \sum_{r=1}^{\binom{n}{i}} \omega_3^i (1-\omega_3)^{n-i} \mathbf{E}_1\left(X_{n+1}|\tilde{\mathbf{X}}_{B_{ir}} \in PoP_3\right)
\end{aligned}$$

https://doi.org/10.1017/S1748499523000076 Published online by Cambridge University Press

$$+(1-\omega_3)\sum_{i=0}^{n}\sum_{r=1}^{\binom{n}{i}}\omega_3^i(1-\omega_3)^{n-i}\sum_{d=0}^{n-i}\sum_{e=1}^{\binom{n-i}{d}}\left(\frac{\omega_2}{1-\omega_3}\right)^d\left(\frac{\omega_1}{1-\omega_3}\right)^{n-i-d}$$

$$\times\frac{\omega_2}{1-\omega_3}\mathbf{E}_1\left(X_{n+1}|\tilde{\mathbf{X}}_{B_{de}}\in PoP_2\right)$$

$$+(1-\omega_3)\sum_{i=0}^{n}\sum_{r=1}^{\binom{n}{i}}\omega_3^i(1-\omega_3)^{n-i}\sum_{d=0}^{n-i}\sum_{e=1}^{\binom{n-i}{d}}\left(\frac{\omega_2}{1-\omega_3}\right)^d\left(\frac{\omega_1}{1-\omega_3}\right)^{n-i-d}$$

$$\times\frac{\omega_1}{1-\omega_3}\mathbf{E}_1\left(X_{n+1}|\tilde{\mathbf{X}}_{B^c_{de}}\in PoP_1\right)$$

$$=\omega_1\left[(1-\zeta_1)\frac{\beta_1}{\alpha_1}+\zeta_1\frac{\sum_i^n x_i}{n}\right]+\omega_2\left[(1-\zeta_2)\mu_2+\zeta_2\frac{\sum_i^n ln(x_i)}{n}\right]$$

$$+\omega_3\left[(1-\zeta_3)\frac{\beta_3}{\alpha_3}+\zeta_3\frac{\sum_i^n x_i^\lambda}{n}\right],$$

where

$$\zeta_1=\sum_{i=0}^{n}\omega_3^i(1-\omega_3)^{n-i+1}\sum_{d=0}^{n-i}\left(\frac{\omega_2}{1-\omega_3}\right)^d\left(\frac{\omega_1}{1-\omega_3}\right)^{n-i-d}\frac{(n-i-d)\alpha_1}{(n-i-d)\alpha_1+r_1}\binom{n-i}{d}\binom{n}{i}$$

$$\zeta_2=\sum_{i=0}^{n}\omega_3^i(1-\omega_3)^{n-i+1}\sum_{d=0}^{n-i}\left(\frac{\omega_2}{1-\omega_3}\right)^d\left(\frac{\omega_1}{1-\omega_3}\right)^{n-i-d}\frac{db_2^2}{db_2^2+\sigma_2^2}\binom{n-i}{d}\binom{n}{i}$$

$$\zeta_3=\sum_{i=0}^{n}\omega_3^i(1-\omega_3)^{n-i}\binom{n}{i}\frac{i}{i+r_2-1}.$$

The next section develops a practical idea based on the logistic regression to derive a probabilistic model to use the additional information $Z_{i,1}\cdots,Z_{i,m}$ to assign population's weight whenever, we partition measurable space $\mathcal{X}$ into two populations.

5. Logistic Regression Credibility for Two Populations

Consider a situation that the measurable space $\mathcal{X}$ can be partitioned into two populations. Moreover, suppose that for each random variable X_i there is some additional information $Z_{i,1}\cdots,Z_{i,m}$, are available. Now using the logistic regression, one may evaluate the first Population's weight by

$$\omega=P\left(X_i\in PoP_1|z_{i,1}\cdots,z_{i,m}\right)$$
$$=\frac{\exp\left\{\beta_0+\sum_{l=1}^m\beta_l z_{i,l}\right\}}{1+\exp\left\{\beta_0+\sum_{l=1}^m\beta_l z_{i,l}\right\}}. \tag{13}$$

Therefore, the result of Corollary 1 can be simplified by the following. Since this result initiated from the logistic regression, hereafter now, we call it "Logistic Regression Credibility."

Remark 1. Suppose that measurable space $\mathcal{X}$ can be partitioned into two populations, then, under Model Assumption 1, the Bayesian credibility mean is

$$E_2[X_{n+1}|X_1, X_2, ..., X_n] = \sum_{i=0}^{n}\sum_{r=1}^{\binom{n}{i}} \omega^i(1-\omega)^{n-i}\left[\omega E_1\left(X_{n+1}|\tilde{X}_{B_{ir}} \in PoP_1\right)\right.$$
$$\left. + (1-\omega)E_1\left(X_{n+1}|\tilde{X}_{B^c_{ir}} \notin PoP_1\right)\right],$$

where ω is given by Equation (13).

The following example represents a practical application of the Logistic Regression Credibility (given by Remark 1)

Example 5. Suppose an insurance company based upon its past experience classified its policyholders in two homogenous groups, labelled "Group 1" and "Group 2," where claim size distribution for the Group 1 is a Normal distribution (with mean θ_1 and variance 0.36) and for the Group 2 is a Normal distribution (with mean θ_2 and variance 0.40) where θ_1 and θ_2, respectively, have been distributed according $N(9, 0.25)$ and $N(10, 0.25)$. Moreover, suppose that the insurance company developed the following logistic regression model to assign its policyholder to the "Group 1"

$$P(Y=1|\boldsymbol{z}) = \frac{\exp\{20.33 + 1.37z_1 - 2.87z_2 + 0.10z_3 - 10.06z_4 + 0.50z_5\}}{1 + \exp\{20.33 + 1.37z_1 - 2.87z_2 + 0.10z_3 - 10.06z_4 + 0.50z_5\}}, \tag{14}$$

where $z_1, \cdots, z_5$, respectively, stand for Gender (0=male and 1=female), Marital Status (0=single and 1=Married), Age (ranges from 20 to 80), Occupation class (distinct values 1,2, 3, and 4) and location (distinct values 1 to 30).

Now consider a 40 years single man who lives in location labelled 9 and his job is labelled 3. Moreover, suppose that his 10 years loss reports are 16.19502, 13.92823, 15.69760, 15.00515, 15.30293, 16.54005, 16.03626, 16.84823, 14.49716, 14.75258.

Using the Equation (14), the policyholder with probability $\omega = 0.2378$ $(1-\omega = 0.7622)$ belongs to "Group 1" ("Group 2"), and his next year Bayesian credibility premium is

$$E_2[X_{11}|X_1, X_2, ..., X_{10}] = \sum_{i=0}^{10}\sum_{r=1}^{\binom{10}{i}} \omega^i(1-\omega)^{10-i}\left[\omega E_1\left(X_{11}|\tilde{X}_{B_{ir}} \in PoP_1\right)\right.$$
$$\left. + (1-\omega)E_1\left(X_{11}|\tilde{X}_{B^c_{ir}} \notin PoP_1\right)\right] = 16.79856.$$

The Logistic Regression Credibility, say LRC, and the Regression Tree Credibility, say RTC, share a same idea. Both of them use a statistical model to partition the measurable space $\mathcal{X}$ into some populations. But, the RTC method develops a credibility prediction for each population while the LRC provides just one credibility prediction for all populations with different weight. The following subsection shows that at least for some cases the LRC has a lower risk function.

5.1 Logistic regression credibility versus the regression tree credibility

Diao & Weng (2019) introduced the RTC model. Their model, in the first step, employs some statistical techniques (such as logistic regression) to partition the measurable space $\mathcal{X}$ into some small regions in which a simple model provides a good fit. Then, in the second step, for each region they applied the Bühlmann-Straub credibility premium formula for each region to predict credibility premium prediction. More precisely, given observed data X_i and its associated information $Z_{i,1} \cdots, Z_{i,m}$, for $i = 1, \cdots, n$. Using a statistical model, such as logistic regression

https://doi.org/10.1017/S1748499523000076 Published online by Cambridge University Press

given by Equation (13), it determines the probability that such claim experience $X_1, X_2, \cdots, X_n$ arrives from the Population 1. If such a probability passes 1/2, the credibility premium predicts using the model which developed for Population 1, otherwise, the model developed for Population 2.

Under the squared error loss function, the RTC method decreases risk function of prediction compared to the regular credibility method. Diao & Weng (2019) presented its theoretical proof for the situation where measurable space $\mathcal{X}$ has been partitioned into two distinguished classes.

The following lemma shows that at least for an interval about 1/2, the LRC's risk function dominates the RTC's risk function.

Lemma 5. *Under Model Assumption 1, consider two following different scenarios to predict the credibility mean based upon the i.i.d. random claim experience $X_1, X_2, \cdots, X_n$..*

***Scenario* 1 (the LRC approach):** *The claim experience $X_1, X_2, \cdots, X_n$, given parameter vector $\mathbf{\Psi} = (\theta_1, \theta_2)'$, has been distributed according the following 2-component normal mixture distribution $\omega N\left(\theta_1, \sigma_1^2\right) + (1-\omega) N\left(\theta_2, \sigma_2^2\right)$, where σ_1^2, σ_2^2 are given, $\omega \in [0, 1]$ and for $j = 1, 2$, θ_j has the conjugate prior distribution $N\left(\mu_j, \tau_j^2\right)$.*

***Scenario* 2 (the RTC approach):** *The measurable space $\mathcal{X}$ partitions into two populations in which if the i.i.d. random claim experience $X_1, X_2, \cdots, X_n$ are belong to Population $j = 1, 2$, then $E(X_i) = \mu_j$, $Var(E(X_i|\Theta)) = \tau_j^2$ and $E(Var(X_i|\Theta)) = \sigma_j^2$.*

Then, at least for the situation that the population's weight, ω, (given by Equation (13)) locates in an interval $I = \left[\left(\tau_2^2 - R_2\right) / \left(R_1 + \tau_2^2\right),\ (R_1 + R_2) / \left(R_2 + \tau_1^2\right)\right]$ under the squared error loss function, the LRC's risk function dominates the RTC's risk function, where $R_l = \sigma_l^2 \tau_l^2 / \left(n\tau_l^2 + \sigma_l^2\right)$ for $l = 1, 2$.

Proof. Similar to Example 2, one may show that under the Scenario 1, the Bayesian credibility premium is $\omega[\xi_1 \bar{X} + (1-\xi_1)\mu_1] + (1-\omega)[\xi_2 \bar{X} + (1-\xi_2)\mu_2]$ and its corresponding risk function under the squared error loss function is

$$L_{LRC}(\omega) = \omega^2 \left[\xi_1^2 \frac{\sigma_1^2}{n} + (1-\xi_1)^2 \tau_1^2\right] + (1-\omega)^2 \left[\xi_2^2 \frac{\sigma_2^2}{n} + (1-\xi_2)^2 \tau_2^2\right],$$

where $\xi_1 = \sum_{i=0}^n \omega^i (1-\omega)^{n-i} \binom{n}{i} \frac{i\tau_1^2}{i\tau_1^2 + \sigma_1^2}$ and $\xi_2 = \sum_{i=0}^n \omega^i (1-\omega)^{n-i} \binom{n}{i} \frac{(n-i)\tau_2^2}{(n-i)\tau_2^2 + \sigma_2^2}$.

However under the Scenario 2, since the RTC method employs the Bühlmann-Straub credibility premium formula, its credibility premium is $\alpha_j \bar{X} + \left(1-\alpha_j\right) \mu_j$, where $\alpha_j = \frac{n}{n + \sigma_j^2/\tau_j^2}$, whenever Population $j = 1, 2$ has been chosen. Therefore, its corresponding risk function under the squared error loss function is

$$L_{RTC}(\omega) = \omega \left[\alpha_1^2 \frac{\sigma_1^2}{n} + (1-\alpha_1)^2 \tau_1^2\right] + (1-\omega) \left[\alpha_2^2 \frac{\sigma_2^2}{n} + (1-\alpha_2)^2 \tau_2^2\right]$$

where the probability that the past claim experience $X_1, X_2, \cdots, X_n$ belong to Population 1, ω, derived from Equation (13).

Now observe that, difference between the above two risk functions, $L_{LRC}(\omega) - L_{RTC}(\omega)$, can be restated as

$$= \omega \frac{\sigma_1^2 \tau_1^2}{n} \sum_{i=0}^n \omega^i (1-\omega)^{n-i} \binom{n}{i} \tau_1^2 \left[\frac{i\omega^{\frac{1}{2}}}{i\tau_1^2 + \sigma_1^2} + \frac{n}{n\tau_1^2 + \sigma_1^2}\right] \left[\frac{i\omega^{\frac{1}{2}}}{i\tau_1^2 + \sigma_1^2} - \frac{n}{n\tau_1^2 + \sigma_1^2}\right]$$

$$+ \omega \frac{\sigma_1^2 \tau_1^2}{n} \sum_{i=0}^n \omega^i (1-\omega)^{n-i} \binom{n}{i} n\sigma_1^2 \left[\frac{\omega^{\frac{1}{2}}}{i\tau_1^2 + \sigma_1^2} + \frac{1}{n\tau_1^2 + \sigma_1^2}\right] \left[\frac{\omega^{\frac{1}{2}}}{i\tau_1^2 + \sigma_1^2} - \frac{1}{n\tau_1^2 + \sigma_1^2}\right]$$

https://doi.org/10.1017/S1748499523000076 Published online by Cambridge University Press

$$+(1-\omega)\frac{\sigma_2^2\tau_2^2}{n}\sum_{i=0}^{n}\omega^i(1-\omega)^{n-i}\binom{n}{i}\tau_2^2\left[\frac{(n-i)(1-\omega)^{\frac{1}{2}}}{(n-i)\tau_2^2+\sigma_2^2}+\frac{n}{n\tau_2^2+\sigma_2^2}\right]\left[\frac{(n-i)(1-\omega)^{\frac{1}{2}}}{(n-i)\tau_2^2+\sigma_2^2}-\frac{n}{n\tau_2^2+\sigma_2^2}\right]$$

$$+(1-\omega)\frac{\sigma_2^2\tau_2^2}{n}\sum_{i=0}^{n}\omega^i(1-\omega)^{n-i}\binom{n}{i}n\sigma_2^2\left[\frac{(1-\omega)^{\frac{1}{2}}}{(n-i)\tau_2^2+\sigma_2^2}+\frac{1}{n\tau_2^2+\sigma_2^2}\right]\left[\frac{(1-\omega)^{\frac{1}{2}}}{(n-i)\tau_2^2+\sigma_2^2}-\frac{1}{n\tau_2^2+\sigma_2^2}\right]$$

$$=\omega\frac{\sigma_1^2\tau_1^2}{n}\sum_{i=0}^{n}\omega^i(1-\omega)^{n-i}\binom{n}{i}\left[\tau_1^2\left(\left(\frac{i\omega^{\frac{1}{2}}}{i\tau_1^2+\sigma_1^2}\right)^2-\left(\frac{n}{n\tau_1^2+\sigma_1^2}\right)^2\right)+n\sigma_1^2\left(\left(\frac{\omega^{\frac{1}{2}}}{i\tau_1^2+\sigma_1^2}\right)^2-\left(\frac{1}{n\tau_1^2+\sigma_1^2}\right)^2\right)\right]$$

$$+(1-\omega)\frac{\sigma_2^2\tau_2^2}{n}\sum_{i=0}^{n}\omega^i(1-\omega)^{n-i}\binom{n}{i}\left[\tau_2^2\left(\left(\frac{(n-i)(1-\omega)^{\frac{1}{2}}}{(n-i)\tau_2^2+\sigma_2^2}\right)^2-\left(\frac{n}{n\tau_2^2+\sigma_2^2}\right)^2\right)\right.$$
$$\left.+n\sigma_2^2\left(\left(\frac{(1-\omega)^{\frac{1}{2}}}{(n-i)\tau_2^2+\sigma_2^2}\right)^2-\left(\frac{1}{n\tau_2^2+\sigma_2^2}\right)^2\right)\right]$$

$$=\omega\frac{\sigma_1^2\tau_1^2}{n}\sum_{i=0}^{n}\omega^i(1-\omega)^{n-i}\binom{n}{n-i}\left[\frac{(n-i)^2\tau_1^2\omega+n\sigma_1^2\omega}{\left((n-i)\tau_1^2+\sigma_1^2\right)^2}-\frac{n^2\tau_1^2+n\sigma_1^2}{\left(n\tau_1^2+\sigma_1^2\right)^2}\right]$$
$$+(1-\omega)\frac{\sigma_2^2\tau_2^2}{n}\sum_{i=0}^{n}\omega^i(1-\omega)^{n-i}\binom{n}{i}\left[(1-\omega)\frac{(n-i)^2\tau_2^2+n\sigma_2^2}{\left((n-i)\tau_2^2+\sigma_2^2\right)^2}-\frac{n^2\tau_2^2+n\sigma_2^2}{\left(n\tau_2^2+\sigma_2^2\right)^2}\right]$$
$$=\sum_{i=0}^{n}\omega^i(1-\omega)^{n-i}\binom{n}{i}H_i(\omega),$$

where

$$H_i(\omega)=\omega\frac{\sigma_1^2\tau_1^2}{n}\left[\frac{i^2\tau_1^2\omega+n\sigma_1^2\omega}{\left(i\tau_1^2+\sigma_1^2\right)^2}-\frac{n^2\tau_1^2+n\sigma_1^2}{\left(n\tau_1^2+\sigma_1^2\right)^2}\right]$$
$$+(1-\omega)\frac{\sigma_2^2\tau_2^2}{n}\left[\frac{(n-i)^2\tau_2^2(1-\omega)+n\sigma_2^2(1-\omega)}{\left((n-i)\tau_2^2+\sigma_2^2\right)^2}-\frac{n^2\tau_2^2+n\sigma_2^2}{\left(n\tau_2^2+\sigma_2^2\right)^2}\right].$$

Now without losing the generality, assume that we can take a derivative with respect to i and observe

$$M_i:=\frac{\partial H_i(\omega)}{\partial i}=-\omega^2\frac{\sigma_1^2\tau_1^2}{n}\frac{2\tau_1^2\sigma_1^2(n-i)}{\left(i\tau_1^2+\sigma_1^2\right)^3}+(1-\omega)^2\frac{\sigma_2^2\tau_2^2}{n}\frac{2i\tau_2^2\sigma_2^2}{\left((n-i)\tau_2^2+\sigma_2^2\right)^3}$$

https://doi.org/10.1017/S1748499523000076 Published online by Cambridge University Press

at $i=0$, $M_0<0$, at $i=n$, $M_n>0$, and $\partial^2 H_i(\omega)/\partial^2 i>0$. This means $H_i(\omega)$ is a concave function with respect to i that attains its maximum at $i=0$ and $i=n$.

Therefore,

$$L_{LRC}(\omega)-L_{RTC}(\omega)\leq\sum_{i=0}^{n}\omega^i(1-\omega)^{n-i}\binom{n}{i}H_0(\omega)=H_0(\omega) \tag{15}$$

$$L_{LRC}(\omega)-L_{RTC}(\omega)\leq\sum_{i=0}^{n}\omega^i(1-\omega)^{n-i}\binom{n}{i}H_n(\omega)=H_n(\omega). \tag{16}$$

Imposing negativity on Equations (15) and (16), perspectively, lead to $\omega\leq(R_1+R_2)/\left(\tau_1^2+R_2\right)$ and $1-\omega\leq(R_1+R_2)/\left(\tau_2^2+R_1\right)$. This observation completes the desired result. □

One should note that, the above $H_i(\omega)$ also can be stated as

$$\begin{aligned}H_i(\omega)=&\left(A_i^{(1)}C^{(1)}+A_i^{(2)}C^{(2)}\right)\omega^2+\left(-B^{(1)}C^{(1)}-2A_i^{(2)}C*(2)+B^{(2)}C^{(2)}\right)\omega\\&+\left(A_i^{(2)}C^{(2)}-B^{(2)}C^{(2)}\right),\end{aligned}$$

where $A_i^{(l)}=\frac{(n-i)^2\tau_l^2+n\sigma_l^2}{\left((n-i)\tau_l^2+\sigma_l^2\right)^2}$, $B^{(l)}=\frac{n^2\tau_l^2+n\sigma_l^2}{\left(n\tau_l^2+\sigma_l^2\right)^2}$ and $C^{(l)}=\frac{\sigma_l^2\tau_l^2}{n}$ for $l=1,2$.

Since $A_i^{(l)}$ is an increasing function with respect to i, one may observe that $A_i^{(l)}\leq A_n^{(l)}$ and $A_0^{(l)}=B^{(l)}$. This fact allows one to conclude that

$$\begin{aligned}4H_i(\omega=0.5)&=A_i^{(1)}C_i^{(1)}-7A_i^{(2)}C^{(2)}-2B^{(1)}C^{(1)}-2B^{(2)}C^{(2)}\\&=C^{(1)}\left[A_i^{(1)}-2B^{(1)}\right]-7A_i^{(2)}C^{(2)}-2B^{(2)}C^{(2)}\\&\leq C^{(1)}\left[A_n^{(1)}-2B^{(1)}\right]-7A_i^{(2)}C^{(2)}-2B^{(2)}C^{(2)}\\&=-C^{(1)}B^{(1)}-7A_i^{(2)}C^{(2)}-2B^{(2)}C^{(2)}\\&\leq 0\end{aligned}$$

and consequently $L_{LRC}(\omega=0.5)-L_{RTC}(\omega=0.5)\leq 0$. The continuation of $L_{LRC}(\omega)-L_{RTC}(\omega)$ in ω shows that at least in an interval about $\omega=0.5$ the LRC's risk function dominates the RTC's risk function. This means that at least in a situation where one with probability a bit more than 50% is going to assign the past claim experience to one of the population and using the RTC's method derives the credibility mean for the future claim. We suggest him/her to use the LRC's method.

Figure 1 illustrates behaviour of $L_{LRC}(\omega)-L_{RTC}(\omega)$ with respect to ω, for some values of $(n,\sigma_1,\sigma_2,\tau_1,\tau_2)$.

6. Discussion and Suggestions

This article considered the Bayesian credibility prediction for the mean of X_{n+1} under a finite class of mixture distributions. In the first step, it developed a recursive formula for the Bayesian credibility mean under such a class of distributions. Since the implementation of the recursive formula is very expensive (see Example 1), therefore, it imposed some additional conditions on the problem. More precisely, it assumed random variables X_i, for $i=1,\cdots,n$, corresponding to the observed sample x_i accompanied with additional information $Z_{i,1},\cdots,Z_{i,m}$, where under a probabilistic model one may use such observable information to determine the population of random variables X_i, see Model Assumption 1 for more details. Under this new assumption, it developed an exact Bayesian credibility mean whenever all members of such a class of mixture distributions belong to the single-parameter exponential family of distributions. Finally for a situation that the

https://doi.org/10.1017/S1748499523000076 Published online by Cambridge University Press

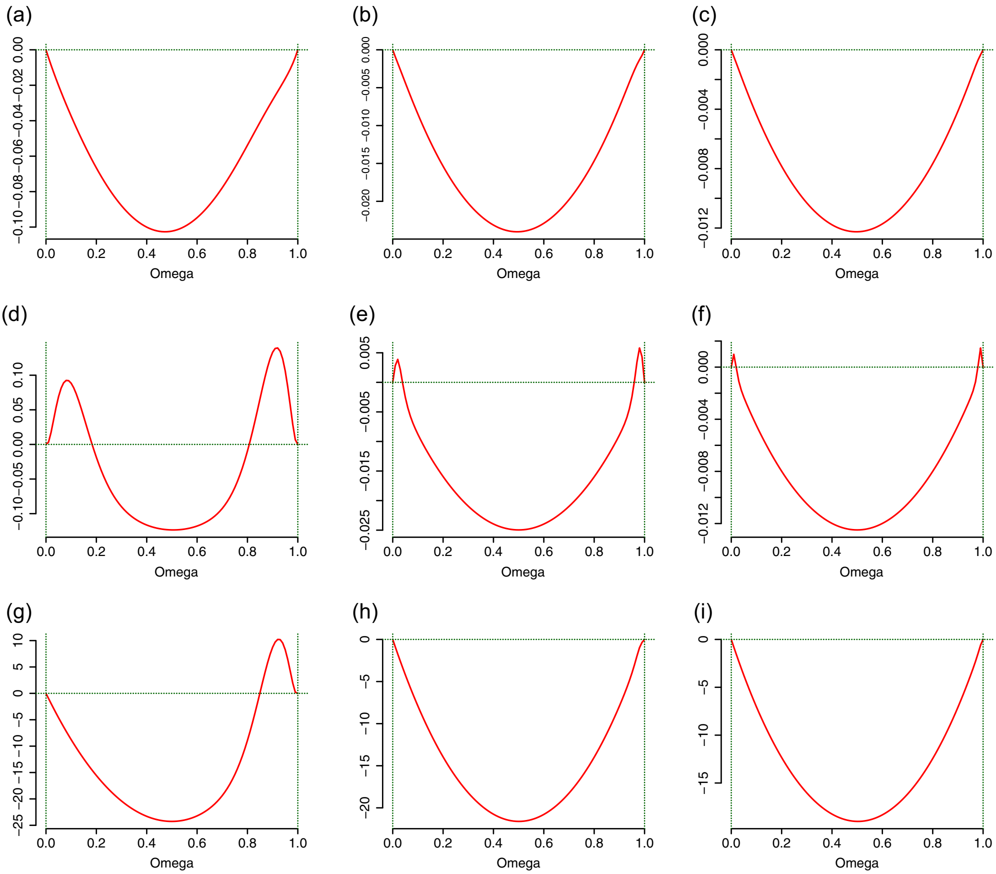

Figure 1. Behaviour of $L_{LRC}(\omega) - L_{RTC}(\omega)$ with respect to ω, under Lemma 5's assumptions, whenever $(n, \sigma_1, \sigma_2, \tau_1, \tau_2) = (10, 1, 2, 1, 2)$ (Panel, a); (50, 1, 2, 1, 2) (Panel, b); (100, 1, 2, 1, 2) (Panel, c); (10, 2, 1, 10, 12) (Panel, d); (50, 2, 1, 10, 12) (Panel, e); (100, 2, 1, 10, 12) (Panel, f); (10, 200, 1, 10, 120) (Panel, g); (50, 200, 1, 10, 120) (Panel, h); and (100, 200, 1, 10, 120) (Panel, i).

measurable space $\mathcal{X}$ can be partitioned into two populations, it employed the logistic regression and introduced the Logistic Regression Credibility which in the sense of the risk function in some specific population's weight dominates the Regression Tree Credibility.

We should note that assumption on the additional information $Z_{i,1}, \cdots, Z_{i,m}$, has a slight difference by assumption on latent variable Z_{ij} in the EM algorithm (see Note 2). More precisely, under Model Assumption 1, $Z_{i,1}, \cdots, Z_{i,m}$ are observable and give a probabilistic information about distribution of random variable X_i, say population's weight. While under the missing data approach, Z_{ij} is a latent variable which provides certain information about distribution X_i. This fact persuades us to claim assumptions in Model Assumption 1 are available and practicable in many cases, see Example 5 as an evidence.

Our finding can be extended for (1) other indices of X_{n+1}, such as the variance of X_{n+1}, as represented in Equation (3), (2) the M-parameter exponential family of distributions, and (3) the Bayesian non-parametric credibility under the Dirichlet process mixture models, which introduced by Fellingham *et al.* (2015) and enriched by Hong & Martin (2017, 2018).

To see the second possible extension, the following recalls Jewell (1974)'s findings for the M-parameter exponential family of distributions with probability density/mass function

$$f(x|\theta) = a(x)e^{\sum_{m=1}^{M} \phi_m(\theta)t_m(x)}/c(\theta) \;\; \forall\, x \in S_X, \tag{17}$$

https://doi.org/10.1017/S1748499523000076 Published online by Cambridge University Press

where $a(\cdot)$, $\phi_m(\cdot)$, $t_m(\cdot)$, for $m=1,2,\cdots,M$, are given functions and the normalising factor $c(\cdot)$ is defined based on the fact that $\int_{S_X} f(x|\theta)dx=1$. To derive the Bayesian credibility prediction for a given index of X_{n+1} under the M-parameter exponential family of distributions, he set $\eta_m=-\phi_m(\theta)$, and considered the conjugate prior distribution

$$\pi^{conj}(\Delta)=[c(\Delta)]^{-\alpha_0}e^{\sum_{m=1}^{M}\{-\beta_{0m}\eta_m\}}/d(\alpha_0,\beta_0),$$

where $\Delta=(\eta_1,\eta_2,\cdots,\eta_M)'$. Then, he showed the Bayesian credibility can be expressed based on the sufficient statistics $t_m(\cdot)$ as

$$E(t_m(X_{n+1})|X_1,\cdots,X_n)=\zeta_n\bar{t}_{m,n}(\tilde{\boldsymbol{x}})+(1-\zeta_n)\frac{\beta_{0m}}{\alpha_0},\ for\ m=1,2,\cdots,M, \tag{18}$$

where the credibility factor $\zeta_n=n/(n+\alpha_0)$ and $\bar{t}_{m,n}(\tilde{\boldsymbol{x}})=\sum_i^n b_m(x_i)/n$.

Acknowledgements. The authors would like to appreciate anonymous reviewers for their constructive comments which improve theoretical foundation and presentation of this article.

References

Atienza, N., **Garcia-Heras, J.** & **Munoz-Pichardo, J.M.** (2006). A new condition for identifiability of finite mixture distributions. *Metrika*, **63**(2), 215–221.

Bailey, A. (1950). A generalized theory of credibility. *Proceedings of the Casualty Actuarial Society*, **13**, 13–20.

Blostein, M. & **Miljkovic, T.** (2019). On modeling left-truncated loss data using mixtures of distributions. *Insurance: Mathematics and Economics*, **85**, 35–46.

Bühlmann, H. (1967). Experience rating and credibility. *Astin Bulletin*, **4**(3), 199–207.

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

Bülmann, H. & **Straub, E.** (1970). Glaubwüdigkeit für Schadensže. *Bulletin of the Swiss Association of Actuaries*, **70**, 111–133.

Carvajal, R., **Orellana, R.**, **Katselis, D.**, **Escárate, P.** & **Agüero, J.C.** (2018). A data augmentation approach for a class of statistical inference problems. *PloS One*, **13**(12), 1–24.

Cai, X., **Wen, L.**, **Wu, X.** & **Zhou, X.** (2015). Credibility estimation of distribution functions with applications to experience rating in general insurance. *North American Actuarial Journal*, **19**(4), 311–335.

de Alencar, F.H., **Galarza, C.E.**, **Matos, L.A.** & **Lachos, V. H.** (2022). Finite mixture modeling of censored and missing data using the multivariate skew-normal distribution. *Advances in Data Analysis and Classification*, **16**(3), 521–557.

Dempster, A.P., **Laird, N.M.** & **Rubin, D.B.** (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, **39**(1), 1–22.

Diao, L. & **Weng, C.** (2019). Regression tree credibility model. *North American Actuarial Journal*, **23**(2), 169–196.

Diebolt, J. & **Robert, C.P.** (1994). Estimation of finite mixture distributions through Bayesian sampling. *Journal of the Royal Statistical Society: Series B (Methodological)*, **56**(2), 363–375

Fellingham, G.W., **Kottas, A.** & **Hartman, B.M.** (2015). Bayesian nonparametric predictive modeling of group health claims. *Insurance: Mathematics and Economics*, **60**, 1–10.

Frühwirth-Schnatter, S., **Celeux, G.** & **Robert, C.P.** (Eds.) (2019). *Handbook of Mixture Analysis*. CRC Press, New York.

Hong, L. & **Martin, R.** (2017). A flexible Bayesian nonparametric model for predicting future insurance claims. *North American Actuarial Journal*, **21**(2), 228–241.

Hong, L. & **Martin, R.** (2018). Dirichlet process mixture models for insurance loss data. *Scandinavian Actuarial Journal*, **2018**(6), 545–554.

Hong, L. & **Martin, R.** (2020). Model misspecification, Bayesian versus credibility estimation, and Gibbs posteriors. *Scandinavian Actuarial Journal*, **2020**(7), 634–649.

Hong, L. & **Martin, R.** (2022). Imprecise credibility theory. *Annals of Actuarial Science*, **16**(1), 136–150.

Jewell, W.S. (1974). Credible means are exact Bayesian for exponential families. *ASTIN Bulletin: The Journal of the IAA*, **8**(1), 77–90.

Keatinge, C.L. (1999). Modeling losses with the mixed exponential distribution. *In Proceedings of the Casualty Actuarial Society*, **86**, 654–698.

Lau, J.W., **Siu, T.K.** & **Yang, H.** (2006). On Bayesian mixture credibility. *ASTIN Bulletin: The Journal of the IAA*, **36**(2), 573–588.

Lee, K., **Marin, J.M.**, **Mengersen, K.** & **Robert, C.** (2009). Bayesian inference on finite mixtures of distributions. In *Perspectives in Mathematical Sciences I: Probability and Statistics* (pp. 165–202).

Li, **H.**, **Lu**, **Y.** & **Zhu**, **W.** (2021). Dynamic Bayesian ratemaking: a Markov chain approximation approach. *North American Actuarial Journal*, **25**(2), 186–205.

Lo, **A.Y.** (1984). On a class of Bayesian nonparametric estimates: I. Density estimates. *The Annals of Statistics*, **12**(1), 351–357.

Marin, **J.M.**, **Mengersen**, **K.** & **Robert**, **C.P.** (2005). Bayesian modelling and inference on mixtures of distributions. *Handbook of Statistics*, **25**, 459–507.

Maroufy, **V.** & **Marriott**, **P.** (2017). Mixture models: building a parameter space. *Statistics and Computing*, **27**(3), 591–597.

McLachlan, **G.** & **Peel**, **D.** (2004). *Finite Mixture Models*. John Wiley & Sons, New York.

Miljkovic, **T.** & **Grün**, **B.** (2016). Modeling loss data using mixtures of distributions. *Insurance: Mathematics and Economics*, **70**, 387–396.

Mowbray, **A.** (1914). How extensive a payroll is necessary to give dependable pure premium? *Proceedings of the Casualty Actuarial Society*, **1**, 24–30.

Payandeh Najafabadi, **A.T.** (2010). A new approach to the credibility formula. *Insurance: Mathematics and Economics*, **46**(2), 334–338.

Payandeh Najafabadi, **A.T.** & **Sakizadeh**, **M.** (2019). Designing an optimal bonus-malus system using the number of reported claims, steady-state distribution, and mixture claim size distribution. *International Journal of Industrial and Systems Engineering*, **32**(3), 304–331.

Payandeh Najafabadi, **A.T.** & **Sakizadeh**, **M.** (2023). Bayesian Estimation for ij-Inflated Mixture Power Series Distributions using an EM Algorithm. Accepted for publication by *Thailand Statistician Journal*.

Rufo, **M.J.**, **Pérez**, **C.J.** & **Martín**, **J.** (2006). Bayesian analysis of finite mixture models of distributions from exponential families. *Computational Statistics*, **21**(3), 621–637.

Rufo, **M.J.**, **Pérez**, **C.J.** & **Martín**, **J.** (2007). Bayesian analysis of finite mixtures of multinomial and negative-multinomial distributions. *Computational Statistics & Data Analysis*, **51**(11), 5452–5466.

Teicher, **H.** (1960). On the mixture of distributions. *The Annals of Mathematical Statistics*, **31**(1), 55–73.

Teicher, **H.** (1963). Identifiability of finite mixtures. *The Annals of Mathematical Statistics*, **34**(4) 1265–1269.

Whitney, **A.** (1918). The theory of experience rating. *Proceedings of the Casualty Actuarial Society*, **4**, 274–292.

Zhang, **B.**, **Zhang**, **C.** & **Yi**, **X.** (2004). Competitive EM algorithm for finite mixture models. *Pattern Recognition*, **37**(1), 131–144.

Zhang, **J.**, **Qiu**, **C.** & **Wu**, **X.** (2018). Bayesian ratemaking with common effects modeled by mixture of Polya tree processes. *Insurance: Mathematics and Economics*, **82**, 87–94.

Cite this article: Jahanbani E, Payandeh Najafabadi AT and Masoumifard K (2024). On Bayesian credibility mean for finite mixture distributions, *Annals of Actuarial Science*, **18**, 5–29. https://doi.org/10.1017/S1748499523000076