

PAPER

Imprecise credibility theory

Liang Hong^{*1} and Ryan Martin²

¹Department of Mathematical Sciences, The University of Texas at Dallas, 800 West Campbell Road, Richardson, TX 75080, USA and ²Department of Statistics, North Carolina State University, 2311 Stinson Drive, Raleigh, NC 27695, USA

*Corresponding author. E-mail: liang.hong@utdallas.edu

(Received 23 October 2020; revised 06 February 2021; accepted 16 March 2021; first published online 15 April 2021)

Abstract

The classical credibility theory is a cornerstone of experience rating, especially in the field of property and casualty insurance. An obstacle to putting the credibility theory into practice is the conversion of available prior information into a precise choice of crucial hyperparameters. In most real-world applications, the information necessary to justify a precise choice is lacking, so we propose an *imprecise credibility estimator* that honestly acknowledges the imprecision in the hyperparameter specification. This results in an interval estimator that is doubly robust in the sense that it retains the credibility estimator's freedom from model specification and fast asymptotic concentration, while simultaneously being insensitive to prior hyperparameter specification.

Keywords: Imprecise probability; Credibility estimator; Model misspecification; Prior uncertainty; Robustness

1. Introduction

The distribution of insurance losses depends on risk characteristics of the policyholder, a feature that cannot be fully known during the underwriting process. However, it is possible to learn about the policyholders' underlying risk characteristics from observed losses, and a Bayesian approach is attractive to insurers for a number of reasons. In particular, this approach allows the insurer to seamlessly incorporate any available prior information about the risk parameter into the analysis and to construct full posterior and predictive distributions that quantify uncertainty for inference about the risk parameter and for future loss values, respectively. Beyond these foundational advantages, Hong & Martin (2017b) show that a Bayesian posterior distribution has several desirable asymptotic properties, and Hong & Martin (2017a, 2018), Huang & Meng (2020), Li & Liu (2018), and Richardson & Hartman (2018) demonstrate that a Bayesian non-parametric model is effective in predicting mortality in life insurance as well as predicting future insurance claims in both healthcare insurance and general insurance. However, despite its benefits, a full Bayesian analysis also has its drawbacks. First, the actuary is obligated to specify the joint distribution – model and prior – of losses and the risk parameter, and the quality of inferences and predictions he makes depends on his model specification being sound; that is, his posited model must closely mimic the true loss distribution, otherwise the inferences and predictions could be biased, potentially severely. Second, unless the model is of an especially simple form, computations of the posterior and predictive distributions will require Markov chain Monte Carlo (MCMC), which can be expensive, especially for non-parametric models. Moreover, when MCMC is used, it is not possible to take practical advantage of the theoretical property that the Bayesian posterior distribution can be updated online as new data becomes available; but see Hahn *et al.* (2018) and Hong & Martin (2019) for some recursive formulas/approximations.

Bühlmann (1967) recognised that, if only a point prediction for a future loss is required, instead of a full predictive distribution, then a very simple and robust approximation is available. Indeed, what is known as Bühlmann’s *credibility estimator* of the predictive distribution mean corresponds to the Bayes optimal linear estimator, and it boils down to a simple convex combination of the prior mean of the loss distribution and the sample mean of the observed losses. Thanks to its intuitive appeal, computational simplicity, and minimal modelling requirements, the credibility estimator has been embraced by the actuarial community since its origin, and it is still widely used in actuarial practice.

A non-trivial obstacle to putting the Bayesian theory into practice is the specification of a prior distribution; the same is true for credibility theory, though to a lesser extent. Asymptotic arguments suggest that the prior is largely irrelevant – and even that default or “non-informative” priors could be used – but this should not be a fully satisfactory answer to the actuary whose motivation to adopt a Bayesian approach in the first place was its ability to incorporate prior information. When one is unsure about the prior, then, at least implicitly, there is a class \mathcal{C} of plausible prior distributions under consideration; we will refer to this class as a *credal set*. There are a few options of how one can proceed with \mathcal{C} :

- *Hierarchical Bayes* (Berger, 2006, section 3.6; Ghosh *et al.*, 2006, section 9.1). Set $\mathcal{C} = \{\Pi_\lambda : \lambda \in \Lambda\}$ to be indexed by a *hyperparameter* λ , specify a prior Q for λ , and use the marginal prior $\int \Pi_\lambda Q(d\lambda)$ for the subsequent Bayesian analysis. The upside is that the posterior may be less sensitive to the choice of Q than to a direct choice of λ , but this only shifts the actuary’s responsibility from choosing λ to choosing Q .
- *Empirical Bayes* (Berger, 2006, section 4.5; Ghosh *et al.*, 2006, section 9.2). For \mathcal{C} indexed by a hyperparameter λ as above, let the data assist with the selection of a single prior from \mathcal{C} by letting $\hat{\lambda}$ be the maximum marginal likelihood estimator of λ , and then use prior $\Pi_{\hat{\lambda}}$ for the subsequent Bayesian analysis. The upside is that a data-assisted choice of prior is an attractive option, but it is ignoring the uncertainty in the prior that originally motivated this discussion, drastically exaggerating the precision.
- *Robust Bayes* (Berger, 2006, section 4.7; Ghosh *et al.*, 2006, section 3.8). Rather than averaging over \mathcal{C} or selecting an individual prior in \mathcal{C} , a third option is there to treat the entire credal set as the prior, i.e., to construct a posterior distribution for θ based on every prior in \mathcal{C} . The upside is that the uncertainty about the prior is acknowledged in the formulation and preserved in the posterior updates, but returning a range of answers to every question – in particular, an interval of predictive distribution means – can be computationally demanding and difficult to interpret.

The first two methods above are standard and have been widely used in the actuarial science literature and beyond. The third approach is more obscure, but it too has received some attention in the actuarial science literature; see section 2.2. Note, however, that robust Bayes is robust only with respect to priors in the credal set. In order to put the theory into practice, very specific assumptions about the model and/or credal set are needed, so the robust Bayes solution would still be sensitive to departures from these assumptions.

Our overall assessment of what is available in the literature is as follows. The classical Bayes solution is optimal, but puts a burden on the actuary to specify a sound model and precise prior; the robust Bayes solution relieves some of this burden by requiring only an imprecise prior but retains sensitivity to the model and increases the computational costs; and the credibility theory almost completely relieves the model specification and computational burdens but retains sensitivity to the prior. The question posed in the present paper is if the prior robustness afforded by the use of a credal set could be incorporated with the simplicity and model robustness of Bühlmann’s credibility framework. (Loss robustness is irrelevant in credibility theory where the squared-error loss is the standard loss function.) Indeed, the credibility estimator depends on the prior distribution through a small number of interpretable quantities, so specification of a credal set is

relatively straightforward. Moreover, the credibility estimator's simplicity makes its computation over the entire credal set almost trivial. Therefore, since there is inherent imprecision in the prior formulation – a consequence of having only limited *a priori* information about the phenomenon under investigation – we propose to directly acknowledge the imprecision via a suitable credal set, and return what we call the *imprecise credibility estimator*, which is an interval of credibility estimators corresponding to each prior in the credal set. We claim that the proposed imprecise credibility estimator is *doubly robust* in the sense that its performance is not affected by the actuary's choice of model or prior, in large part because the actuary is not required to make such specifications. Therefore, our imprecise credibility estimator gives the actuary the best of both worlds: an easy-to-compute estimator of the predictive mean that flexibly and honestly accounts for available prior information and is robust to biases resulting from model misspecification.

It is important to emphasise that, in our context, the word “imprecise” is not synonymous with “inaccurate,” “rough,” “unscientific,” etc. Instead, our proposed solution is imprecise in the sense that the actuary carefully considered the available *a priori* information and what assumptions he was willing to make, encoded all this in a well-defined credal set, and preserved that uncertainty all the way through his analysis.

The remainder of this article is organised as follows. In section 2 we give a brief review of Bayesian decision theory, imprecise probability and robust Bayes, and Bühlmann's credibility formula. Next, in section 3, we propose our imprecise credibility estimator and show several desirable properties to justify our “double-robustness” claims. In section 4, we give a brief numerical illustration to show how the level of imprecision can affect the imprecise credibility estimator compared to a typical robust Bayes solution. Section 5 describes some extensions of and alternative perspectives on the proposed imprecise credibility theory, and section 6 gives some concluding remarks.

2. Background

2.1 Bayesian decision theory

Let $X^n = (X_1, \dots, X_n)$ be the observable loss data where X_1, \dots, X_n are independent and identically distributed (iid) according to a true distribution P^* . In the Bayesian approach, the actuary starts by postulating a model $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ where Θ is the parameter space and assumes that losses are conditionally iid given θ , i.e.,

$$(X_1, \dots, X_n | \theta) \stackrel{\text{iid}}{\sim} P_\theta$$

Of course, it is possible that $P^* \notin \mathcal{P}$, this is out of the actuary's control. Next, the actuary accounts for uncertainty about the value of θ by assigning it a prior distribution Π . Let

$$\mu(\theta) = E_\theta(X) \quad \text{and} \quad \sigma^2(\theta) = V_\theta(X)$$

be the mean and variance of P_θ , respectively. Then, define

$$m_1(\Pi) = E_\Pi \{\mu(\theta)\}, \quad m_2(\Pi) = V_\Pi \{\mu(\theta)\}, \quad \text{and} \quad v(\Pi) = E_\Pi \{\sigma^2(\theta)\}$$

where expectation and variance are taken with respect to the prior distribution Π . In the insurance literature, $\mu(\theta)$, $\sigma^2(\theta)$, $m_1(\Pi)$, $m_2(\Pi)$, and $v(\Pi)$ are often referred to as the *hypothetical mean*, the *process variance*, the *collective premium*, the *variance of hypothetical mean*, and the *mean of the process variance*, respectively. Throughout, we assume that Π is such that $m_1(\Pi)$, $m_2(\Pi)$, and $v(\Pi)$ exist and are finite.

The usual goal of the actuary is to obtain a point prediction for the next loss X_{n+1} . In the Bayesian decision theory, this is often formulated as an optimisation problem: to find the estimator δ that minimises the expected squared error loss

$$E\{X_{n+1} - \delta(X^n)\}^2$$

where the expectation is taken with respect to the joint distribution of (X^n, X_{n+1}) under the aforementioned model. Once the optimal function δ_{opt} is obtained, and loss data $X^n = x^n$ are observed, then $\delta_{\text{opt}}(x^n)$ will be used to predict X_{n+1} . So the problem boils down to finding δ_{opt} . To this end, we apply the iterated expectation formula and Jensen’s inequality for conditional expectations to obtain

$$\begin{aligned} E\{X_{n+1} - \delta(X^n)\}^2 &= E\left(E\left[\{X_{n+1} - \delta(X^n)\}^2 \mid X^n, \theta\right]\right) \\ &\geq E\left(E\left[\{X_{n+1} - \delta(X^n)\} \mid X^n, \theta\right]^2\right) \\ &= r(\Pi, \delta) \end{aligned}$$

where

$$r(\Pi, \delta) = E\left[\{\mu(\theta) - \delta(X^n)\}^2\right] \tag{1}$$

and the expectation is taken with respect to the joint distribution of (θ, X^n) . The quantity $r(\Pi, \delta)$ is called the *Bayes risk* of the estimator δ under the prior Π relative to squared-error loss. On the other hand, the projection theory in the Hilbert space L^2 (e.g. Shiryaev, 1996, section II.11 or van der Vaart, 1998, section 11.2) implies that

$$r(\Pi, \delta) = E\left[\{\mu(\theta) - \delta(X^n)\}^2\right] \geq E\left[\{\mu(\theta) - E\{\mu(\theta) \mid X^n\}\}^2\right]$$

where equality holds if and only if $\delta(X^n) = E\{\mu(\theta) \mid X^n\}$. It follows that

$$\delta_{\text{opt}}(X^n) = E\{\mu(\theta) \mid X^n\} \tag{2}$$

Then δ_{opt} is the *Bayes rule*, the estimator that is optimal in the sense that it minimises the Bayes risk relative to the assumed prior Π .

2.2 Robustness to the prior: imprecise probability

An important point is that, in most cases, the unknown parameter θ is determined by the model specified \mathcal{P} . Therefore, questions about which priors for θ generally can be entertained only come after the model has been specified. For the present discussion, we assume that the model \mathcal{P} and corresponding model parameter have been determined, but we revisit this subtle point in section 3 below.

As discussed in section 1, if the application at hand lacks sufficient information to completely determine a (model and) prior distribution for the risk parameter, then there is a credal set consisting of candidate prior distributions. Of course, the credal set can take all sorts of forms. Two extreme cases are as follows:

- *Complete knowledge:* $\mathcal{C} = \{\Pi\}$;
- *Complete ignorance:* $\mathcal{C} = \{\text{all probability distributions } \Pi \text{ on } \Theta\}$.

Most real-world applications are somewhere between these two extremes. For example, it is not unreasonable that the actuary could specify ranges A_j for a few functionals $f_j, j = 1, \dots, J$, of the prior and, in that case, the credal set would be given by

$$\mathcal{C} = \{\Pi : f_j(\Pi) \in A_j, j = 1, \dots, J\}$$

These functionals could be moments, quantiles, or some other summaries; note that, especially in cases where θ is multidimensional, the summaries might be moments or quantiles associated with a scalar function of θ , e.g., the mean $\mu(\theta)$ under distribution P_θ .

To understand the role played by the credal set, and how it relates to the notions of precision and imprecision, a very brief jaunt into imprecise probability territory is necessary. Given a credal

set \mathcal{C} , whose generic element Π is a probability measure on Θ , we define the corresponding lower and upper probabilities:

$$\underline{\Pi}(A) = \inf_{\Pi \in \mathcal{C}} \Pi(A) \quad \text{and} \quad \overline{\Pi}(A) = \sup_{\Pi \in \mathcal{C}} \Pi(A), \quad A \subseteq \Theta \tag{3}$$

It should be immediately clear that the interpretation and mathematical properties of $\underline{\Pi}$ are very different from those of the individual $\Pi \in \mathcal{C}$. Indeed, there is a lot that can be said about this definition and the subsequent developments. Here, we will only mention what is essential for our present purposes; for more details, we refer the interested reader to the general introduction to imprecise probability in Augustin *et al.* (2014) and to the more comprehensive and technical works of Walley (1991) and Troffaes & de Cooman (2014). An obvious consequence of (3) is that

$$\underline{\Pi}(A) \leq \overline{\Pi}(A), \quad A \subseteq \Theta$$

Then the notion of (im)precision can be understood by looking at the gap between the lower and upper probabilities. In the “complete knowledge” extreme, the lower and upper probabilities are the same, hence gap is 0. This means there is no uncertainty about how to assign probability to A , so we say that the prior is *precise*. In the “complete ignorance” extreme, the lower probability for every A (except Θ) is 0 and the upper probability is 1. This means there is uncertainty about how to assign probability to A , so we say that the prior is *imprecise*. Most real-world examples fall somewhere in between these two extremes, i.e., where the difference $\overline{\Pi}(A) - \underline{\Pi}(A)$ is strictly between 0 and 1 for some A .

When an imprecise prior is specified, through a credal set \mathcal{C} , the posterior updates are based on the *generalised Bayes rule* (e.g. Walley, 1991, section 6.4), which boils down to applying the usual Bayes update to each $\Pi \in \mathcal{C}$, resulting in a posterior credal set

$$\mathcal{C}(X^n) = \{\Pi(\cdot | X^n) : \Pi \in \mathcal{C}\}$$

Since all the distributions contained in $\mathcal{C}(X^n)$ are plausible solutions, the most natural strategy is, for any relevant posterior summary, to return the range of that summary over $\mathcal{C}(X^n)$. For example, in insurance applications, the prior- Π posterior mean $\delta_{\text{opt}}^\Pi(X^n) = E\{\mu(\theta) | X^n\}$ is a relevant quantity, and the actuary could report the interval

$$\left[\inf_{\Pi \in \mathcal{C}} \delta_{\text{opt}}^\Pi(X^n), \sup_{\Pi \in \mathcal{C}} \delta_{\text{opt}}^\Pi(X^n) \right]$$

of plausible posterior mean values, given the assumed model and observed data X^n , which honestly accounts for the inherent imprecision in the prior specification. Computing the endpoints of this interval is non-trivial, but some general formulas and approximations are available for specific prior classes; see, e.g., Wasserman (1990), Berger (2006, section 4.7), and Ghosh *et al.* (2006, section 3.8).

Depending on the application, it may be necessary to report a single answer, rather than a range of answers. For such cases, a standard robust Bayes solution is the Γ -*minimax rule* (e.g. Berger, 2006; Vidakovic, 2000), where Γ the set of priors, what we are calling the credal set \mathcal{C} . The idea is to define the Bayes risk, $r(\Pi, \delta)$, for a given prior Π and decision rule δ . When the prior is uncertain, a robust solution is to use a rule that is “good” for all $\Pi \in \mathcal{C}$, so the Γ -minimax proposal is to find $\delta = \delta_{\text{opt}}^\mathcal{C}$ that *minimises the maximum risk*, i.e.,

$$\sup_{\Pi \in \mathcal{C}} r\left(\Pi, \delta_{\text{opt}}^\mathcal{C}\right) = \inf_{\delta} \sup_{\Pi \in \mathcal{C}} r(\Pi, \delta)$$

As one might expect, solving this optimisation problem is a practical challenge, and the available solutions make very specific assumptions about the model and prior; in the actuarial science literature, see Young (1998), Gómez-Déniz *et al.* (2006), and Gómez-Déniz (2009).

2.3 Robustness to the model: credibility theory

Despite the theoretical optimality of the Bayes rule δ_{opt} in (2), there are several disadvantages. First, it is rare that the Bayes rule would be available in closed-form; typically, MCMC would be required. Second, and perhaps more importantly, δ_{opt} depends heavily on the choice of model and the assumed prior distribution. If it happens that the model is misspecified, or if the prior distribution fails to adequately represent the *a priori* uncertainty, then the Bayes rule would be afflicted by model misspecification bias, rendering those theoretical optimality properties virtually meaningless. Bühlmann (1967) proposed a simple linear approximation to δ_{opt} that simultaneously overcomes both of these challenges.

Specifically, restrict attention in (1) to linear estimators of $\mu(\theta)$, i.e.,

$$\hat{\delta}(X^n) = a_0 + a_1X_1 + \dots + a_nX_n$$

where a_0, a_1, \dots, a_n are real numbers. Since the Bayes model assumes X^n are conditionally iid, de Finetti’s theorem (e.g. Kallenberg, 2002, Theorem 11.10) implies that X_1, \dots, X_n are exchangeable, i.e., the distributions of (X_1, \dots, X_n) and $(X_{k_1}, \dots, X_{k_n})$ are the same for any permutation (k_1, \dots, k_n) of $(1, \dots, n)$. It follows that the optimal $\hat{\delta}$ will have $a_1 = \dots = a_n$, so it suffices to consider a linear estimator of the form $\hat{\delta}(X^n) = \alpha + \beta\bar{X}_n$ where \bar{X}_n is the sample mean. Substituting $\hat{\delta}(X^n)$ for $\delta(X^n)$ in (1), breaking down the corresponding Bayes risk, and using the familiar mean and variance formulas for \bar{X}_n , we obtain

$$r(\Pi, \hat{\delta}) = \beta^2 n^{-1} v(\Pi) + (\beta - 1)^2 m_2(\Pi) + \{\alpha + (\beta - 1)m_1(\Pi)\}^2$$

It follows from routine calculus that

$$\hat{\delta}_{opt}(X^n) = \alpha_n + \beta_n \bar{X}_n \tag{4}$$

where

$$\alpha_n = \alpha_n(\Pi) = \frac{v(\Pi)m_1(\Pi)}{nm_2(\Pi) + v(\Pi)} \quad \text{and} \quad \beta_n = \beta_n(\Pi) = \frac{nm_2(\Pi)}{nm_2(\Pi) + v(\Pi)}$$

In the insurance literature, (4) is called the (Bühlmann) *credibility formula* and $\hat{\delta}_{opt}$ is referred to as the (Bühlmann) *credibility estimator*. The credibility estimator has a number of appealing properties: first, it is easy to implement numerically, no sophisticated Monte Carlo calculations are required; second, it has a nice interpretation as a weighted average of the individual risk and the group risk in the experience rating context (e.g. Bühlmann & Gisler, 2005, section 3.1; Klugman *et al.*, 2008, section 16.4.4); third, it depends only on a few low-dimensional features of the prior distribution Π for the full model parameter θ ; and, fourth, as $n \rightarrow \infty$, it is a strongly consistent estimator of the true mean, μ^* , independent of the model and prior (e.g. Hong & Martin, 2020). In view of the third and fourth properties, the credibility estimator can be applied without too much concern about the model misspecification risk, at least in cases where the sample size n is reasonably large.

3. Imprecise Credibility Estimation

Despite its flexibility, the credibility estimator does still depend, to a certain extent, on the posited model \mathcal{P} and the prior distribution Π . Indeed, the model defines the unknown model parameter θ in Θ and, in turn, the mean and variance functions $\mu(\theta)$ and $\sigma^2(\theta)$ under model P_θ . Then the prior Π , together with the model \mathcal{P} , determine the three key quantities,

$$Q(\Pi) = \{m_1(\Pi), m_2(\Pi), v(\Pi)\}$$

that appear in the credibility formula (4). An important observation is that the required prior information is in terms of the mean and variance functions, quantities directly related to the

observables. This simplifies the elicitation of meaningful prior information, either from experts, prior experience, or the observed data. Three things, however, are clear: first, there will be uncertainty about the choice of $Q(\Pi)$ in every practical application; second, the credibility formula requires these values to be specified; and third, the quality of the estimate depends, to some extent, on those specified values. How can the choice of these values be made in a principled way, one that reduces the burden on the actuary?

Because the quantities in $Q(\Pi)$ are directly related to information contained in the observables, it would be tempting to use the actual observed data to help guide the choice of $Q(\Pi)$ in an application. This boils down to an *empirical credibility estimator* as in, e.g., Klugman *et al.* (2008, section 20.4). There are advantages to letting the data assist with the choice of these prior features, resulting in an estimate $\hat{Q}(\Pi)$, directly plugging these values into the credibility formula amounts to ignoring the uncertainty in the choice of $Q(\Pi)$. To see this, suppose that Actuary A is *uncertain* about $Q(\Pi)$ and uses the data to obtain a plug-in estimator $\hat{Q}(\Pi) = (7, 2, 18)$. Now suppose that Actuary B has very detailed prior information at his disposal and is *certain* about the choice $Q(\Pi) = (7, 2, 18)$. The two actuaries would produce exactly the same credibility estimator despite having very different information and levels of certainty *a priori*. This issue is not the use of data to guide the choice of $Q(\Pi)$ – see section 5.1 below – but rather that the use of a plug-in estimator amounts to manufacturing precision/certainty when there is none.

To avoid manufacturing precision in the credibility estimator, we propose a variation on the robust Bayesian approach. Consider a credal set determined by a specified range for each of the three quantities in $Q(\Pi)$, i.e.,

$$\mathcal{C} = \{ \Pi : \underline{m}_1 \leq m_1(\Pi) \leq \bar{m}_1, \underline{m}_2 \leq m_2(\Pi) \leq \bar{m}_2, \underline{\nu} \leq \nu(\Pi) \leq \bar{\nu} \} \tag{5}$$

where $\underline{m}_1, \bar{m}_1, \underline{m}_2, \bar{m}_2, \underline{\nu}$, and $\bar{\nu}$ are positive numbers to be chosen at the actuary’s discretion. Since the credibility estimator only depends on the prior Π through the value of $Q(\Pi)$, we can equivalently formulate this with a set

$$\mathcal{Q} = \{ q = (m_1, m_2, \nu) : \underline{m}_1 \leq m_1 \leq \bar{m}_1, \underline{m}_2 \leq m_2 \leq \bar{m}_2, \underline{\nu} \leq \nu \leq \bar{\nu} \} \tag{6}$$

Now it is clear that the incorporation of imprecision in this credibility context is much more straightforward than in the robust Bayesian setting. That is, instead of fixing specific values for (m_1, m_2, ν) , the actuary now considers a range of such values. And only having to specify a range of values for (m_1, m_2, ν) reduces the burden on the actuary to make one “right” choice. Moreover, it is arguably easier to elicit information about (m_1, m_2, ν) in the form of intervals: between the two statements “the collective premium equals m_1 ” and “the collective premium is between \underline{m}_1 and \bar{m}_1 ,” an expert is much more likely to be confident in the latter.

It is clear that the credal set \mathcal{C} given by (5) or, equivalently, the set \mathcal{Q} of $q = (m_1, m_2, \nu)$ values in (6), corresponds to a range of credibility estimators. Indeed, for each $q = (m_1, m_2, \nu) \in \mathcal{Q}$, there is a corresponding credibility estimator:

$$\hat{\delta}_{\text{opt}}^q(X^n) = \frac{\nu m_1}{nm_2 + \nu} + \frac{nm_2}{nm_2 + \nu} \bar{X}_n$$

The next result shows that this range is, in fact, an interval in \mathbb{R} .

Proposition 1. *Given \mathcal{Q} in (6), which is determined by the credal set \mathcal{C} in (5), the range of corresponding credibility estimators forms a closed and bounded interval in \mathbb{R} .*

Proof. The set \mathcal{Q} is a closed hyper-rectangle in \mathbb{R}^3 , so obviously connected and compact. Since the real-valued function $q \mapsto \hat{\delta}_{\text{opt}}^q(X^n)$ is continuous, the corresponding image of \mathcal{Q} is connected and compact, hence a closed and bounded interval. □

Henceforth, we will refer to this interval as the *imprecise credibility estimator* and denote it by $\mathbb{I}(X^n; \mathcal{Q})$ or $\mathbb{I}(X^n; \mathcal{C})$, or just \mathbb{I}_n for short. Specifically,

$$\mathbb{I}_n = \left[\min_{q \in \mathcal{Q}} \hat{\delta}_{\text{opt}}^q(X^n), \max_{q \in \mathcal{Q}} \hat{\delta}_{\text{opt}}^q(X^n) \right] \tag{7}$$

The imprecise credibility estimator has several appealing features. First, recall that the Bühlmann credibility estimator’s dependence on the statistical model \mathcal{P} was only indirect, through the prior’s implicit dependence on the model. That indirect dependence is eliminated by giving the estimator a parametrisation in terms of the generic quantity q , whose values are interpretable independent of a model. This implies that the imprecise credibility estimator is model-free and, consequently, is not susceptible to model misspecification bias. Second, mapping the actuary’s specified range of q values back to the credal set (5), it is clear that \mathcal{C} covers a very wide range of prior distributions, not just those of a particular parametric form, so (7) is not sensitive to the choice of prior form either. Therefore, we conclude that the imprecise credibility is doubly robust in the sense that its performance is not sensitive to the actuary’s choice of model or prior distribution – because the actuary is not even required to make such specifications.

We should emphasise again that “imprecise” in this context is not synonymous with “inaccurate.” The imprecision in our proposed imprecise credibility estimator is entirely determined by the amount of information available to the actuary. Indeed, in the extreme case of complete certainty, the actuary can choose \mathcal{Q} to be a singleton and Bühlmann’s original credibility estimator emerges. For the more realistic case where the actuary has some degree of uncertainty about the prior specification, the imprecise credibility estimator in (7) seems to be the natural generalisation of Bühlmann’s developments. In other words, our imprecise credibility estimator combines the benefits of Bühlmann’s with an honest assessment of the actuary’s uncertainty about the prior specification.

It is also worth emphasising that the imprecise credibility estimator, while being an interval, is *not* an interval estimator in the traditional sense. That is, the range of values in (7) is completely determined by the prior imprecision. It has nothing at all to do with the sampling distribution properties of an estimator, so one cannot expect it to have any frequentist coverage probability guarantees like a confidence interval would. Moreover, it has nothing at all to do with a posterior distribution, so one cannot expect it to have a certain amount of posterior probability assigned to it like a Bayesian posterior credible interval would. Instead, the imprecise credibility estimator is just the set of all credibility estimators corresponding to the range of prior specifications the actuary is willing/able to make.

Next, computation of the imprecise credibility estimator \mathbb{I}_n in (7) is straightforward, no more difficult than that of Bühlmann’s credibility estimator.

Proposition 2. *Given \mathcal{Q} , which is determined by the tuple $(\underline{m}_1, \bar{m}_1, \underline{m}_2, \bar{m}_2, \underline{v}, \bar{v})$, the endpoints of the imprecise credibility estimator satisfy*

$$\min_{q \in \mathcal{Q}} \hat{\delta}_{\text{opt}}^q(X^n) = \min\{c_1, c_2, c_3, c_4\} \quad \text{and} \quad \max_{q \in \mathcal{Q}} \hat{\delta}_{\text{opt}}^q(X^n) = \max\{c_1, c_2, c_3, c_4\}$$

where

$$\begin{aligned} c_1 &= \frac{n}{n + \bar{v}/\underline{m}_2} \bar{X}_n + \frac{\bar{v}/\underline{m}_2}{n + \bar{v}/\underline{m}_2} \underline{m}_1, \\ c_2 &= \frac{n}{n + \bar{v}/\underline{m}_2} \bar{X}_n + \frac{\bar{v}/\underline{m}_2}{n + \bar{v}/\underline{m}_2} \bar{m}_1, \\ c_3 &= \frac{n}{n + \underline{v}/\bar{m}_2} \bar{X}_n + \frac{\underline{v}/\bar{m}_2}{n + \underline{v}/\bar{m}_2} \underline{m}_1, \\ c_4 &= \frac{n}{n + \underline{v}/\bar{m}_2} \bar{X}_n + \frac{\underline{v}/\bar{m}_2}{n + \underline{v}/\bar{m}_2} \bar{m}_1 \end{aligned}$$

Table 1. Summary statistics for scaled Norwegian fire claims data, 1990–1991

| Year | n | Min | 1st Qu | Median | Mean | 3rd Qu | Max |
|------|-----|-------|--------|--------|-------|--------|---------|
| 1990 | 628 | 1.000 | 1.577 | 2.300 | 3.947 | 3.611 | 157.074 |
| 1991 | 624 | 1.000 | 1.506 | 2.197 | 3.641 | 3.818 | 99.384 |

Proof. Write

$$\delta_{\text{opt}}^q(X^n) = \frac{n\bar{X}_n + (v/m_2)m_1}{n + v/m_2}$$

Then the theorem follows from the fact that $\delta_{\text{opt}}^q(X^n)$ is monotone in m_1 and v/m_2 . □

There is no simple and general rule that can be given for choosing the set \mathcal{Q} , or the corresponding credal set (5), in practice. The reason is that \mathcal{Q} is intended to represent *the actuary’s degree of uncertainty* in a particular application, so clearly we cannot give any specific advice about how he should choose \mathcal{Q} . One can make use of the actual data to help guide this choice, and we have a few remarks about this in section 5.1. Here, however, all we can say is that, clearly, the size of \mathcal{Q} is directly related to the length of \mathbb{I}_n , so it is to the actuary’s advantage to choose \mathcal{Q} as small as he can possibly justify.

Fortunately, the choice of \mathcal{Q} can only affect the results in small or moderate samples. The following result demonstrates that, when n is large, the imprecise credibility estimator collapses, at a root- n rate, to the single point μ^* .

Proposition 3. *Suppose the loss variables X_1, X_2, \dots are iid from distribution P^* , with mean μ^* . Let $\mathbb{I}_n = \mathbb{I}(X^n; \mathcal{Q})$ be the imprecise credibility estimator in (7). Then*

$$\max\{|\max \mathbb{I}_n - \mu^*|, |\min \mathbb{I}_n - \mu^*|\} = O(n^{-1/2}) \quad \text{in } P^*\text{-probability as } n \rightarrow \infty$$

Proof. For $c_k, k = 1, 2, 3, 4$, as defined in Proposition 2, it is easy to check that,

$$c_k - \mu^* = \bar{X}_n - \mu^* + O(n^{-1}), \quad \text{in } P^*\text{-probability, } k = 1, 2, 3, 4$$

Chebyshev’s inequality implies that $\bar{X}_n - \mu^*$ is $O(n^{-1/2})$ in P^* -probability, so the same holds for each $c_k - \mu^*$. Since the max and min operators are continuous, we get

$$\min\{c_1, c_2, c_3, c_4\} - \mu^* = O(n^{-1/2}) \quad \text{and} \quad \max\{c_1, c_2, c_3, c_4\} - \mu^* = O(n^{-1/2})$$

in P^* -probability by the continuous mapping theorem. Then the claim follows since the min and max above correspond to the endpoints of \mathbb{I}_n . □

Therefore, the imprecise credibility estimator has all the benefits of Bühlmann’s original proposal, namely, insensitivity to model and prior specifications and fast asymptotic consistency, while incorporating an honest assessment of the actuary’s prior uncertainty.

4. Example

Consider the Norwegian fire claims data that has recently been analysed by several authors, including Brazauskas & Kleefeld (2016), Mdziniso & Cooray (2018), Hong & Martin (2018, 2020), and Syring *et al.* (2019). The 1990 and 1991 data sets contains $n = 628$ and $n = 624$ entries, respectively, and are available from <http://lstat.kuleuven.be/Wiley/> or the R package `CASdatasets`. We first rescale the data sets by dividing each entry by 500. Table 1 provides the summary statistics for the two scaled data sets.

The situation we have in mind here is one where the 1990 data provides some “prior information” that we use to help construct a credal set, which will then be converted into an imprecise credibility estimate based on the 1991 data. This requires specification of bounds on the three prior hyperparameters (m_1, m_2, ν) , which we carry out as follows. We should emphasise that the assessment of imprecision is a subjective one (see section 5.1) that must be considered on a case-by-case basis. What we present below is an illustration of the kinds of considerations one should make, rather than a recommendation of what credal set to use.

- Since the interpretation of m_1 is the most straightforward of the three, in order to keep our comparisons relatively simple and brief, we suggest here to set $[\underline{m}_1, \overline{m}_1]$ to be the standard 95% confidence interval based on a normal approximation for the distribution of the sample mean which, in this case, is $[\underline{m}_1, \overline{m}_1] = [3.28, 4.61]$. Since the sample size is relatively large, a central limit theorem approximation ought to be reasonable, but we emphasise again that we are not *recommending* practitioners make this choice.
- Since m_2 is more difficult to interpret than m_1 , we apply more care in setting the bound. There is no reason to believe that the variance of the process mean would be especially large, so we proceed with setting a guess \hat{m}_2 equal to the third quartile of the 1990 data, which is 3.61. Then we set the bounds as $\underline{m}_2 = \phi^{-1}\hat{m}_2$ and $\overline{m}_2 = \phi\hat{m}_2$, respectively, where $\phi > 1$ is an *imprecision factor* described in more detail below.
- Like with m_2 , we proceed by taking a guess $\hat{\nu}$ equal to the sample variance from the 1990 data, which is 72.5. Then the bounds for ν are set as $\underline{\nu} = \phi^{-1}\hat{\nu}$ and $\overline{\nu} = \phi\hat{\nu}$, respectively, where $\phi > 1$ is an imprecision factor, not necessarily the same as that for m_2 introduced above.

In our comparisons below, we consider three different *imprecision levels* corresponding to values of the factor ϕ .

Level 1. $\phi = 2$

Level 2. $\phi = 3$

Level 3. $\phi = 4$.

Since we have two such imprecision factors in our credal set construction – one for each of the m_2 and ν bounds – each with three levels, there are nine total imprecision level combinations. In the figures, we denote the different combinations as $A.B$, which is interpreted as m_2 and ν having imprecision levels A and B , respectively. Our goal is to investigate the effects of both imprecision levels and the sample size on the imprecise credibility estimator.

The plots in Figure 1 show the imprecise credibility estimator applied to data that have the same sample mean (3.64) as the 1990 Norwegian fire data but with varying sample sizes: $n \in \{100, 200, 400\}$. The vertical bar represents the interval \mathbb{I}_n itself. As expected, smaller sample sizes and higher imprecision levels correspond to larger intervals. While the intervals in each individual panel do not vary greatly with the imprecision level, especially when the sample size is relatively large, they do vary and the extent of this variability is controlled by the practitioner’s choice of an imprecision factor ϕ . Depending on the application, even our most conservative imprecision level, namely, 3.3, might be too liberal, in which case even wider bounds on the hyperparameters (m_1, m_2, ν) might be warranted, leading to even wider imprecise credibility estimators.

For comparison, consider one of the robust Bayes solutions presented in Gómez-Déniz (2009), which assumes an exponential family model with a conjugate prior. Of course, since these data appear to have been sampled from a heavy-tailed distribution, an exponential family model is questionable. Our reason for introducing the exponential family model is to make a different point, one that is independent of whether that model is appropriate or not, but it is worth emphasising that there are no “standard models” that would be appropriate for these data, hence the proposed method has an advantage in the sense that it does not require the practitioner to specify a model at all. Returning to our main point with the exponential family model, the robust Bayes

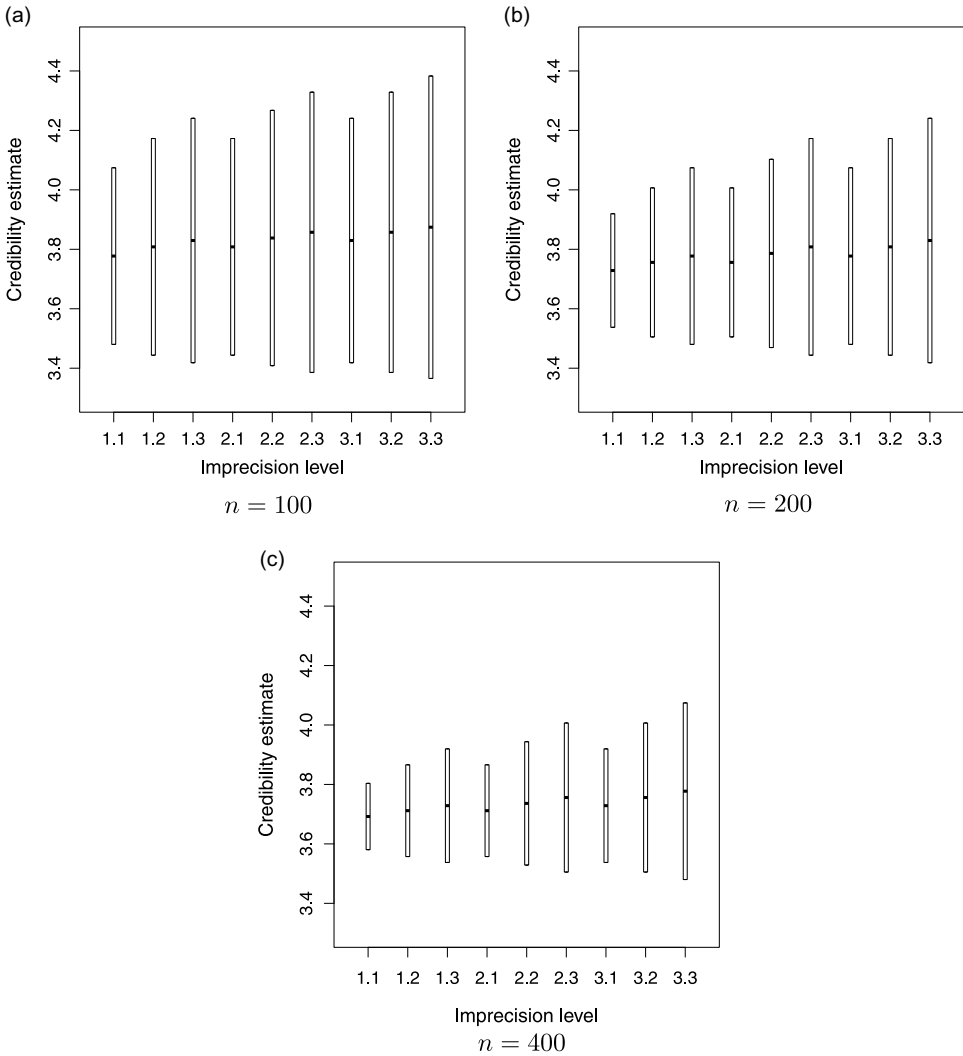


Figure 1. Plots of the imprecise credibility estimator \mathbb{I}_n based on the sample mean from the 1991 Norwegian fire data, with varying sample size n and varying imprecision levels in the construction of the credal set corresponding to the plausible values of (m_1, m_2, v) .

solutions derived in Gómez-Déniz (2009), Theorem 1, all return single values, all very close to the sample mean. Therefore, there are no signs of prior imprecision in the answer returned by the robust Bayes procedure, i.e., the answer is largely insensitive to the choice of credal set, especially when n is large.

5. Remarks

5.1 Default choice of credal set?

As mentioned above, the credal set in (5) or, equivalently, the set \mathcal{Q} in (6) represents what the actuary is willing to assume based on the available prior information in a given application. Therefore, we cannot give any firm advice on how to make that choice. All we can offer are a few remarks about how the data might be used to help guide this choice. We urge the reader to keep in mind that we are not recommending the actuary choose \mathcal{Q} in this way.

Following the now well-developed non- or semi-parametric empirical credibility theory (Bühlmann & Gisler, 2005, section 4.9; Klugman *et al.*, 2008, section 20.4), estimates of the three quantities – m_1 , m_2 , and v – in the credibility formula are available. Since those are relatively simple functions of the current and perhaps historical data, a naive strategy would use the sampling distributions of these functions, perhaps under some simplifying assumptions, to construct approximate “confidence intervals” for m_1 , m_2 , and v . These confidence intervals may not be especially reliable, since they were derived based on some simplifying assumptions. In view of this, the actuary might consider stretching these intervals out to some extent before using them to define the range of the three quantities in \mathcal{Q} .

Similarly, suppose we have a parametric model P_θ and a class of priors Π_λ indexed by the hyperparameter λ . In principle, it would be possible to evaluate the marginal likelihood for λ , given data X^n , produce a maximum marginal likelihood estimator $\hat{\lambda}$. Then $(m_1, m_2, v)(\Pi_\lambda)$ are functions of λ , so some guidance on the choice of \mathcal{Q} can be provided by finding a set of plausible values for λ . Like above, one could use the asymptotic normality of maximum likelihood estimators to construct an approximate “confidence region” for λ which, in turn, could be used to guide the choice of \mathcal{Q} .

5.2 Extension to non-constant risk exposure cases

In the above, our discussion is in the framework of the Bühlmann credibility estimation where the risk exposure is assumed to be one. In some real-world applications, this assumption may not be appropriate. For example, if a policyholder purchases an insurance policy in the middle of a year, then his/her benefits will be effective from the purchase date throughout to the end of the year and the corresponding risk exposure will not be one; if some policyholders drop out of a group insurance in 1 year and some other join in the following year, then risk exposure for this group insurance will vary from year to year. To accommodate these cases, Bühlmann & Straub (1970) propose a generalisation of the Bühlmann credibility estimator. This generalisation assumes that losses X_1, \dots, X_n are iid, given θ , with individual premium $\mu(\theta) = E_\theta(X)$ and the process variance $k_i^{-1}V_\theta(X)$, where k_i is a constant proportional to the size of the risk, i.e., it stands for the risk exposure for X_i . Let k_1, \dots, k_n be the risk exposure for the X_1, \dots, X_n , respectively, and $k = k_1 + \dots + k_n$ be the total risk exposure. As in the Bühlmann credibility estimation, we still put $m_1(\Pi) = E_\Pi\{\mu(\theta)\}$, $m_2(\Pi) = V_\Pi\{\mu(\theta)\}$, and $v(\Pi) = E_\Pi\{\sigma^2(\theta)\}$ but take $\bar{X}_n = k^{-1} \sum_{i=1}^n k_i X_i$. Then the Bühlmann–Straub credibility estimator is given by

$$\hat{\mu}_{BS}^c = \frac{v(\Pi)m_1(\Pi)}{km_2(\Pi) + v(\Pi)} + \frac{km_2(\Pi)}{km_2(\Pi) + v(\Pi)}\bar{X}_n \tag{8}$$

If X_i is interpreted as the average loss for a group of k_i members in the i^{th} year, then the credibility premium to be charged for each member in the group for the $(n + 1)^{\text{st}}$ year is $\hat{\mu}_{BS}^c$, while the total premium to charge this group of k_{n+1} members should be $k_{n+1}\hat{\mu}_{BS}^c$.

The imprecise credibility estimator can be constructed for the Bühlmann–Straub credibility estimator too. The procedure is completely similar to the one for the the Bühlmann credibility estimator. After choosing a credal set \mathcal{Q} based on his prior knowledge, the actuary obtains the corresponding imprecise credibility estimator $\mathbb{I}_{BS}(X^n; \mathcal{C})$ via (8) as before. It is also easy to see that Propositions 1–3 all extend in a straightforward way to the imprecise Bühlmann–Straub credibility estimator.

5.3 Doubly robust imprecise Gibbs posteriors

Hong & Martin (2020) demonstrate that the classical credibility estimator can be interpreted as the posterior mean of an appropriate *Gibbs posterior*. The Gibbs posterior distribution, which has its origins in statistical physics, is the output of a generalisation of the Bayesian framework,

one where the quantity of interest is *defined* by the solution to an optimisation problem, namely, as the minimiser of an expected loss; here, “loss” is in the sense of decision theory, not a loss to the insurance company. In our present context, the quantity of interest is the mean of the X distribution, which is most directly characterised as the solution to the optimisation problem

$$\mu^* = \arg \min_{\mu} R(\mu)$$

where $R(\mu) = E\ell_{\mu}(X)$ and $\ell_{\mu}(x) = (x - \mu)^2$. This is just the classical result which states that the mean of a distribution is the minimiser of the expected squared error loss. For problems like this, where the quantity of interest is defined via a loss function rather than a likelihood, Bissiri *et al.* (2016) argued that the proper generalisation of the Bayesian approach results in a posterior distribution for μ with a density function

$$\pi_n(\mu) \propto e^{-\omega n R_n(\mu)} \pi(\mu), \quad \mu \in \mathbb{R}$$

where $R_n(\mu) = n^{-1} \sum_{i=1}^n \ell_{\mu}(X_i)$ is the empirical version of the risk, $\omega > 0$ is a so-called *learning rate* parameter (e.g. Grünwald, 2012), and π is a prior density for μ . In the present setting, with squared error loss, the first term on the right-hand side amounts to a Gaussian likelihood; therefore, if π is a suitable Gaussian prior, then the corresponding Gibbs posterior mean has form similar to that of the classical credibility estimator, justifying the above claim. For other loss functions, the Gibbs posterior form could be quite different; see Syring & Martin (2017, 2019, 2020), Syring *et al.* (2019), and Bhattacharya & Martin (2020) for examples and general theoretical properties.

One advantage of the Gibbs posterior is its robustness to model misspecification – this happens automatically because the Gibbs posterior does not require the user to specify a model. If, on top of this model-free Gibbs formulation, one considered a credal set of candidate prior distributions for μ , then we get an imprecise Gibbs posterior that is also robust to prior specification, hence *doubly robust*. In fact, if the credal set of priors for μ consists of Gaussian distributions, then each would return a Gibbs posterior mean with form like the credibility estimator, so the imprecise Gibbs posterior mean would return an interval like that in (7). Of course, in the present context, it is better to directly write down the imprecise credibility estimator, as we have done here, but this imprecise Gibbs posterior perspective is more general and deserving of further investigation.

6. Conclusion

Prior uncertainty and model misspecification are two major concerns for actuaries when they want to use the Bühlmann credibility estimator. In this paper, we proposed a method for credibility estimation by representing the prior uncertainty in terms of a credal set. The proposed method leads to an interval of credibility estimators, which we refer to as the imprecise credibility estimator, that preserves the key features of the classical credibility estimator, while being honest about the inherent prior imprecision. This makes our method doubly robust in the sense that it is robust to both model misspecification and prior uncertainty. Our method also extends to the Bühlmann–Straub credibility model.

Again, we want to emphasise that imprecision is not synonymous with “inaccurate” in the present context. More often than not, the available prior information is not sufficiently detailed to identify exactly one prior distribution to be used in a Bayesian analysis, so this imprecision is inherent and should be accounted for. By not working with the entire class of prior distributions compatible with the available information, whether it be for the sake of simplicity or some other reason, the actuary is making potentially unjustifiable assumptions that can bias his estimates and predictions. By working with the proposed imprecise credibility estimator, the actuary can avoid these potential biases.

A big picture view of what is proposed here suggests some potentially interesting future investigations. The original Bayes framework starts with a fully specified statistical model and prior distribution, yielding optimal decision procedures, relative to the assumed model, etc. Bühlmann, building on the early work of Whitney (1918), recognised the challenges in specifying a full statistical model, and that such specifications are not necessary for the estimation problem at hand, so he relaxed the fully Bayesian formulation. Later, other researchers recognised the difficulty in specification of a prior distribution and developed a corresponding robust Bayesian framework, leading to Γ -minimax optimal decision procedures. Here, we have combined those latter two practical modifications of the original Bayesian framework to provide a new tool that is both statistically efficient and accommodates available prior information, without requiring the actuary to specify a statistical model or assume any particular prior form. To generalise this “best of both worlds” beyond this relatively simple credibility estimation context, the formulation in terms of an *imprecise Gibbs posterior* as described in section 5.3 above seems especially promising and deserving of further investigation.

Acknowledgements. We thank the editor and the anonymous reviewers for many useful comments and suggestions. This work is partially supported by the National Science Foundation, DMS–1811802.

References

- Augustin, T., Coolen, F., de Cooman, G. & Troffaes, M. (Eds.) (2014). *Introduction to Imprecise Probabilities*. John Wiley & Sons, Chichester.
- Berger, J.O. (2006). *Statistical Decision Theory and Bayesian Analysis*, 2nd edition. Springer, New York.
- Bhattachary, I. & Martin, R. (2020). Gibbs posterior inference on multivariate quantiles, arXiv:2002.01052.
- Bissiri, P., Holmes, C. & Walker, S. (2016). A general framework for updating belief distribution. *Journal of Royal Statistical Society, Series B—Statistical Methodology*, 78(5), 1103–1130.
- Brazauskas, Y. & Kleefeld, A. (2016). Modeling severity and measuring tail risk of Norwegian fire claims. *North American Actuarial Journal*, 20(1), 1–16.
- Bühlmann, H. (1967). Experience rating and credibility. *ASTIN Bulletin*, 4, 199–207.
- Bühlmann, H. & Gisler, A. (2005). *A Course in Credibility Theory and its Applications*. Springer, New York.
- Bühlmann, H. & Straub, E. (1970). Glaubwürdigkeit für Schadensätze. *Mitteilungen der Vereinigung Schweizerischer Versicherungs-Mathematiker*, 70, 111–133.
- Ghosh, J.K., Delampady, M. & Samanta, T. (2006). *An Introduction to Bayesian Analysis*. Springer, New York.
- Gómez-Déniz, E. (2009). Some Bayesian credibility premiums obtained by using posterior regret Γ -minimax methodology. *Bayesian Analysis*, 4(2), 223–242.
- Gómez-Déniz, E., Pérez-Sánchez, J.M. & Vázquez-Polo, J.F. (2006). On the use of posterior regret Γ -minimax actions to obtain credibility premiums. *Insurance Mathematics and Economics*, 39, 115–121.
- Grünwald, P.D. (2012). The safe Bayesian: Learning the learning rate via the mixability gap. In *Proceedings of the 23rd International Conference on Algorithmic Learning Theory* (pp. 169–183).
- Hahn, P.R., Martin, R. & Walker, S.G. (2018). On recursive Bayesian predictive distributions. *Journal of the American Statistical Association*, 113(523), 1085–1093.
- Hong, L. & Martin, R. (2017a). A flexible Bayesian nonparametric model for predicting future insurance claims. *North American Actuarial Journal*, 21, 228–241.
- Hong, L. & Martin, R. (2017b). A review of Bayesian asymptotics in general insurance applications. *European Actuarial Journal*, 7, 231–255.
- Hong, L. & Martin, R. (2018). Dirichlet process mixture models for insurance loss data. *Scandinavian Actuarial Journal*, 6, 545–554.
- Hong, L. & Martin, R. (2019). Real-time Bayesian nonparametric prediction of solvency risk. *Annals of Actuarial Science*, 13, 67–79.
- Hong, L. & Martin, R. (2020). Model misspecification, Bayesian versus credibility estimation, and Gibbs posteriors. *Scandinavian Actuarial Journal*, 7, 634–649.
- Huang, Y. & Meng, S. W. (2020). A Bayesian nonparametric model and its application in insurance loss prediction. *Insurance: Mathematics and Economics*, 93, 84–94.
- Kallenberg, O. (2002). *Foundations of Modern Probability*, 2nd edition. Springer, New York.
- Klugman, S.A., Panjer, H.H. & Willmot, G.E. (2008). *Loss Models: from Data to Decisions*, 3rd edition. John Wiley & Sons, Hoboken, New Jersey.

- Li, H. and Lu, Y. (2018). A Bayesian non-parametric model for small population mortality. *Scandinavian Actuarial Journal*, 7, 605–628.
- Mdziniso, N.C. & Cooray, K. (2018). Odd Pareto families of distributions for modeling loss payment data. *Scandinavian Actuarial Journal*, 1, 42–63.
- Richardson, R. & Hartman, B. (2018). Bayesian nonparametric regression models for modeling and predicting healthcare claims. *Insurance: Mathematics and Economics*, 83, 1–8.
- Shafer, G. (1976). *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, NJ.
- Shiryayev, A.N. (1996). *Probability*, 2nd edition. Springer, New York.
- Syring, N., Hong, L. & Martin, R. (2019). Gibbs posterior inference on value-at-risk. *Scandinavian Actuarial Journal*, 2019(7), 548–557.
- Syring, N. & Martin, R. (2017). Gibbs posterior inference on the minimum clinically important difference. *Journal of Statistical Planning and Inference*, 187, 67–77.
- Syring, N. & Martin, R. (2019). Calibrating general posterior credible regions. *Biometrika*, 106(2), 479–486.
- Syring, N. & Martin, R. (2020). Robust and rate-optimal Gibbs posterior inference on the boundary of a noisy image. *The Annals of Statistics*, 48(3), 1498–1513.
- Troffaes, M. & de Cooman, G. (2014). *Lower Previsions*. John Wiley & Sons, Chichester.
- van der Vaart, A.W. (1998). *Asymptotic Statistics*. Cambridge University Press, New York.
- Vidakovic, B. (2000). Γ -minimax: A paradigm for conservative robust Bayesians. In D. Ríos Insua & F. Ruggeri (Eds.), *Robust Bayesian Analysis* (pp. 241–259). Springer, New York.
- Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. Chapman & Hall, London.
- Wasserman, L.A. (1990). Prior envelopes based on belief functions. *Annals of Statistics*, 18, 454–464.
- Whitney, A. (1918). The theory of experience rating. *Proceedings of Casualty Actuarial Society*, 4, 274–292.
- Young, V.R. (1998). Robust Bayesian credibility using semiparametric models. *ASTIN Bulletin*, 28(2), 187–203.