

ROBUST BAYESIAN CREDIBILITY USING SEMIPARAMETRIC MODELS

VIRGINIA R. YOUNG

University of Wisconsin – Madison

ABSTRACT

In performing Bayesian analysis of insurance losses, one usually chooses a parametric conditional loss distribution for each risk and a parametric prior distribution to describe how the conditional distributions vary across the risks. Young (1997) applies techniques from nonparametric density estimation to estimate the prior and uses the estimated model to calculate the predictive mean of future claims given past claims. A shortcoming of this method is that, in estimating the prior, one assumes the average claim amount equals the conditional claim. In this paper, we consider a class of priors obtained by perturbing the one determined nonparametrically, as in Young (1997). We thereby reflect the uncertainty in the prior that arises from the randomness in the claim data. We, then, calculate intervals for the corresponding predictive means. We illustrate our method with data from Dannenburg et al. (1996) and compare the intervals of the predictive means with nonparametric confidence intervals.

KEYWORDS

Kernel density estimation, claim estimation, robust Bayesian estimation, Dempster-Shafer belief functions, upper and lower expectations.

1. INTRODUCTION

Analyzing insurance losses is an important task for actuaries. Accurately predicting insurance losses aids in creating equitable premiums for policyholders – one wishes the premium to be low enough for the insurance product to be competitive while large enough for the insurance company to remain solvent. In performing Bayesian analysis of insurance losses, one usually chooses a parametric conditional loss distribution for each risk and a parametric prior distribution to describe how the conditional distributions vary across the risks. A criticism of Bayesian analysis is that the prior distribution can be difficult to choose and the resulting model might not represent the loss data very well.

Young (1997) applies nonparametric density estimation to estimate the prior from the claim data and uses the estimated model to calculate the predictive mean of future claims. A shortcoming of this nonparametric method is that, in estimating the prior, one assumes the average claim amount of a risk equals the conditional mean of that risk. The method given by Young (1997), thus, does not reflect the uncertainty in the prior that arises from the randomness of the claim data.

As in Young (1997), we use a semiparametric mixture model to represent the insurance losses of a portfolio of risks: We choose a parametric conditional loss distribution for each risk with unknown conditional mean that varies across the risks. We apply techniques from nonparametric density estimation to estimate a prior for the distribution of the conditional means, π_0 . Semiparametric methods in statistics provide a bridge between nonparametric and parametric methods. Nonparametric methods do not yield 'tight' results when one has information about the loss distributions involved; however, parametric methods do not give reliable results when one uses the 'wrong' loss distributions.

One reason that a semiparametric method is appropriate for analyzing insurance losses is that actuaries often have information about the loss distributions for individual policyholders or homogeneous risk classes – either through historical claim data or theoretical considerations. Thus, parametric loss distributions for the risk classes', or the policyholders', claims are reasonable to use in this case. However, actuaries are not always able to describe a prior distribution for the conditional means without using the data, this 'second-level' information is otherwise difficult to specify. Therefore, we propose estimating the prior nonparametrically.

By using the data to estimate the prior, one should reflect the uncertainty in the prior π_0 that arises from the randomness of the claim data. To do this, we consider a class of priors obtained by locally perturbing π_0 based on the variability of the claim data. Local perturbation priors are recommended by Wasserman (1990b) to account for uncertainty in a given prior. We calculate the range of the predictive mean as given by the infimum and supremum of the predictive means over this set of priors. Such a procedure of calculating a quantity, like the predictive mean, over a class of priors is part of robust Bayesian analysis, an important area in Bayesian statistics (Berger, 1994).

In Section 2, we briefly review Young (1997) and set the notation for this paper. In Section 3, we describe how local perturbation priors can be defined via the Dempster-Shafer theory of belief functions, (Dempster, 1967, 1968) and (Shafer, 1979). By relying on the Dempster-Shafer theory, calculating the range of the conditional expectations is simplified, (Wasserman, 1990a, b) and (Walley, 1991). We illustrate the computations with an analytically tractable example. Finally, in Section 4, we apply our method to data from Dannenburg et al. (1996) and compare the intervals of the conditional expectations with nonparametric confidence intervals.

2. NOTATION AND KERNEL DENSITY ESTIMATION OF THE PRIOR

2.1. Notation and Assumptions

Assume that the underlying claim of risk i per unit of exposure is a conditional random variable $Y|\theta_i, i = 1, 2, \dots, I$, with probability density function $f(y|\theta_i)$. For each of the I risks, one observes the average claims per unit of exposure $x_i = (x_{i1}, x_{i2}, \dots, x_{iT_i})$, with an associated exposure vector $w_i = (w_{i1}, w_{i2}, \dots, w_{iT_i})$, $i = 1, 2, \dots, I$. Thus, the observed average claim x_{it} is the arithmetic average of w_{it} claims, each of which is an independent realization of the conditional random variable $Y|\theta_i$. For example, if a risk is in the t^{th} policy period of the i^{th} risk class, and w_{it} may be the number of policies in the i^{th} class during the t^{th} policy period.

Assume that the parameter θ is the conditional mean, $E[Y|\theta] = \theta$. Assume that parameters, other than the conditional mean, are fixed across the risks. The loss distribution of a given risk is, therefore, characterized by its conditional mean, although that mean is generally unknown. Denote the probability density function of θ by $\pi(\theta)$, also called the *structure function* (Bühlmann, 1967, 1970). The structure function characterizes how the conditional mean θ varies from risk to risk. Note that our model is a special case of the one given by Bühlmann and Straub (1970).

The goal of credibility theory is to estimate the conditional mean $E[Y|\theta]$ of a risk, given that the risk's claim experience is x and exposure is w . As in Young (1997), set the credibility formula equal to the predictive mean $E[Y|\bar{x}]$, given the weighted sample average \bar{x} , weighted by the exposure w . Also restrict attention to parametric conditional distributions for which $E[Y|\theta] = \theta$,¹ the sample mean is a sufficient statistic for θ , and the functional form of $f(y|\theta)$ is closed under averaging.² Families of densities that satisfy these properties are (1) the normal, with mean θ and fixed variance σ^2 , (2) the gamma, with mean $\theta = \frac{\alpha}{\beta}$ and fixed shape parameter α , and (3) the inverse gaussian, with mean θ and fixed $\lambda = \frac{\alpha}{\theta^3}$.

2.2. Kernel Density Estimation

Young (1997) uses kernel density estimation (Silverman, 1986) to estimate the probability density $\pi(\theta)$. A *kernel* K acts as a weight function and satisfies the condition $\int_{-\infty}^{\infty} K(t)dt = 1$. If one were to observe directly the

¹ Thus, the predictive mean equals the posterior mean; that is, $E[Y|\bar{x}] = E[\theta|\bar{x}]$.

² That is, if \bar{X} is an average of w claims that follow the distribution given by $f(y|\theta)$, then the density of \bar{X} has the same functional form as $f(y|\theta)$.

conditional means $\theta_1, \theta_2, \dots, \theta_I$, then the kernel density estimate of $\pi(\theta)$ with kernel K would be given by

$$\frac{1}{I} \sum_{i=1}^I \frac{1}{h_i} K\left(\frac{\theta - \theta_i}{h_i}\right), \tag{2.1}$$

in which h_i is a positive parameter called the *window width*, or *bandwidth*. Assume that the kernel is symmetric.

Because one observes only data x_i and w_i and not the true conditional means θ_i , one might use the sample mean \bar{x}_i to estimate θ_i consistently, $i = 1, 2, \dots, I$, (Serfling, 1980). In the expression in (2.1), one might wish to weight the terms in the sum according to the relative number of claims for the i^{th} risk so that the expectation of θ is the sample mean

$$\bar{\bar{x}} = \frac{\sum_{i=1}^I \sum_{t=1}^{T_i} w_{it} x_{it}}{\sum_{i=1}^I \sum_{t=1}^{T_i} w_{it}} = \frac{\sum_{i=1}^I w_i \bar{x}_i}{\sum_{i=1}^I w_i},$$

in which $w_i = \sum_{t=1}^{T_i} w_{it}$. Young, therefore,

proposes the following kernel density estimator for $\pi(\theta)$

$$\pi_0(\theta) = \sum_{i=1}^I \frac{w_i}{w_{tot}} \frac{1}{h_i} K\left(\frac{\theta - \bar{x}_i}{h_i}\right), \tag{2.2}$$

in which $w_{tot} = \sum_{i=1}^I w_i$.

One problem with using the sample mean \bar{x}_i to estimate θ_i is that for small sample sizes, the sample mean might not give a good estimate of θ_i . One can measure the strength of the estimate by the standard error of \bar{x}_i , $se(\bar{x}_i)$:

$$se(\bar{x}_i) = \sqrt{\frac{\sum_{t=1}^{T_i} w_{it} (x_{it} - \bar{x}_i)^2}{(T_i - 1)w_i}}. \tag{2.3}$$

In practice, we propose using the standard error of \bar{x}_i , $i = 1, \dots, I$, to perturb π_0 and thereby create a class of local perturbation priors (Wasserman, 1990b). This class of priors accounts for the randomness inherent in using the sample mean \bar{x}_i to estimate θ_i . We define local perturbation priors in the next section after giving a summary of the Dempster-Shafer theory of belief functions.

3. LOCAL PERTURBATION PRIORS

Local perturbation priors can be defined readily by using the Dempster-Shafer theory of belief functions. One can also use that theory and results of Wasserman (1990a,b) and Walley (1991) to calculate intervals of conditional expectations. Therefore, we first summarize Dempster-Shafer's theory, then show how local perturbation priors are defined using this theory, and finally show how to calculate conditional expectations.

3.1. Dempster-Shafer Theory of Belief Functions

Belief functions are set functions that are numerical representations of uncertainty, as in the uncertainty surrounding a prior distribution in Bayesian credibility theory. The theory of belief functions began when Dempster (1967, 1968) defined upper and lower probabilities and used them to illustrate a new method of statistical inference. Shafer (1979) continued the work of Dempster and developed the theory in more detail. More recently, statisticians have been applying Dempster-Shafer's theory to robust Bayesian analysis; see, for example, Berger (1994), Wasserman (1990b), and the references therein.

Let $(\Theta, B(\Theta))$ be a measurable space, and let $(\Omega, B(\Omega), \mu)$ be a probability space. Θ is the space of interest; in the credibility context in this paper, Θ is the set of positive real numbers, the set of possible values of the conditional mean. The sets Θ and Ω are both complete, separable metric spaces, and $B(\Theta)$ and $B(\Omega)$ are the Borel subsets of Θ and Ω , respectively.

Let Γ be a mapping from Ω to the nonempty, closed subsets of Θ . Define a *belief function*, Bel , and a *plausibility function*, Pl , on $B(\Theta)$ by

$$Bel(A) = \mu(\{\omega \in \Omega : \Gamma(\omega) \subset A\}) \text{ and } Pl(A) = \mu(\{\omega \in \Omega : \Gamma(\omega) \cap A \neq \emptyset\}).$$

Loosely speaking, $Bel(A)$ is the probability that the subset $\Gamma(\omega)$ is contained in A , and $Pl(A)$ is the probability that $\Gamma(\omega)$ intersects A . Bel and Pl are related by the relationship $Bel(A) = 1 - Pl(A^c)$. Note that $Bel(\emptyset) = Pl(\emptyset) = 0$, $Bel(\Theta) = Pl(\Theta) = 1$, and $Bel(A) \leq Pl(A)$ with equality if and only if Bel is a probability measure. The four-tuple $(\Omega, B(\Omega), \mu, \Gamma)$ is called the *source* for Bel .

A probability measure P on $(\Theta, B(\Theta))$ is said to be *compatible* with Bel and Pl if, for every $A \in B(\Theta)$, one has $Bel(A) \leq P(A) \leq Pl(A)$. Let Π_{Bel} be the set of all probability measures compatible with Bel and Pl .¹ It can be shown that Π_{Bel} is nonempty and that for each $A \in B(\Theta)$,

$$Bel(A) = \inf_{P \in \Pi_{Bel}} P(A), \quad (3.1)$$

and

$$Pl(A) = \sup_{P \in \Pi_{Bel}} P(A), \quad (3.2)$$

(Shafer, 1979). Thus, Bel and Pl may be thought of as the lower and upper bounds, respectively, of a class of probability measures. For this reason, Bel and Pl are also called *lower* and *upper probabilities*, respectively. The class Π_{Bel} of probability measures compatible with Bel and Pl is useful in robust Bayesian analysis.

¹ Π_{Bel} is also called the *core* of Bel (Denneberg, 1994).

Walley (1991) provides a behavioral interpretation for upper and lower probabilities. For example, let $A_t = \{\theta : \theta > t\}$ be the subset of risks with conditional mean larger than t . In this case, $\underline{P}(A_t)$ is the supremum buying price for the gamble that pays \$1 if a risk chosen at random has conditional mean θ larger than t . Conversely, $\bar{P}(A_t)$ is the infimum selling price for the gamble that pays \$1 if a randomly chosen risk has θ larger than t . Note that $\bar{P}(A_t)$ is also 1 minus the supremum betting rate at which one will bet that a randomly chosen risk has θ less than or equal to t ; that is, $\bar{P}(A_t) = 1 - \underline{P}(A_t^c)$, as previously observed. To avoid sure loss in this setting, one has that $\underline{P}(A) \leq \bar{P}(A)$, for all $A \in \mathcal{B}(\Theta)$, which follows from (3.1) and (3.2).

Wasserman (1990b) defines a class of *local perturbations* of a given prior π_0 on $(\Theta, \mathcal{B}(\Theta))$ as follows. Let $(\Theta, \mathcal{B}(\Theta), \pi_0, \Gamma)$ be the source, for which $\Gamma : \Theta \rightarrow \mathcal{B}(\Theta)$ is a function such that $\theta \in \Gamma(\theta)$. The set $\Gamma(\theta)$ represents one's uncertainty about the probability of θ . The class of local perturbation priors of π_0 is defined to be the class Π_{Bel} generated by such a multi-valued mapping Γ .

Example 3.1: Suppose $X|\theta \sim U(0, 2\theta)$, and that the base prior for θ is given by

$$\pi_0(\theta) = \frac{1}{1000}, \quad 1000 < \theta < 2000.$$

Define Γ on $[1000, 2000]$ by $\Gamma(\theta) = [\theta - d, \theta + d]$, for some fixed $d < 1000$, $\theta \in [1000, 2000]$. Then, the lower probability, or belief, of an interval $[a, b]$ is

$$\begin{aligned} \underline{P}[a, b] &= \begin{cases} \int_{a+c}^{b-c} \pi_0(\theta) d\theta, & \text{if } a + d < b - d, \\ 0, & \text{else;} \end{cases} \\ &= \begin{cases} \frac{b - a - 2d}{1000}, & \text{if } a + d < b - d, \\ 0, & \text{else;} \end{cases} \end{aligned}$$

Also, the upper probability, or plausibility, of an interval $[a, b]$ is

$$\begin{aligned} \bar{P}[a, b] &= \int_{a-d}^{b+d} \pi_0(\theta) d\theta \\ &= \frac{\min(b + d, 2000) - \max(a - d, 1000)}{1000}. \end{aligned}$$

In the credibility context of this paper, one might set $\Gamma(\theta) = [\theta - c \text{ se}(\theta), \theta + c \text{ se}(\theta)]$, in which c is a positive constant, and se is a function determined by the standard errors $\text{se}(\bar{x}_i)$ from (2.3). For example, one might define se by setting θ_i equal to \bar{x}_i and by interpolating a polynomial through the I pairs $(\theta_i, \text{se}(\theta_i)) = (\bar{x}_i, \text{se}(\bar{x}_i))$.

3.2. Conditional Expectations

To apply robust Bayesian analysis to the credibility context of this paper, one calculates a range of predictive means (or posterior means of θ , in this case) over the collection of local perturbation priors of π_0 , given \bar{x} . Specifically, one determines the *lower conditional expectation*

$$\underline{E}[\theta|\bar{x}] = \inf_{P \in \Pi} \int \theta d(P_{\bar{x}}(\theta)), \tag{3.3}$$

and the *upper conditional expectation*

$$\bar{E}[\theta|\bar{x}] = \sup_{P \in \Pi} \int \theta d(P_{\bar{x}}(\theta)), \tag{3.4}$$

in which $P_{\bar{x}}$ is the posterior measure of P given \bar{x} .¹

To simplify the formulas for these conditional expectations, we first introduce some notation. Denote the likelihood $f(\bar{x}|\theta)$ of θ given \bar{x} by $L(\theta)$. By Walley (1991, Chapter 8), the lower conditional expectation $\underline{E}[\theta|\bar{x}]$ is the unique solution α to the equation

$$\underline{E}[(\theta - \alpha)L(\theta)] = 0, \tag{3.5}$$

in which $\underline{E}[Z(\theta)] = \inf_{P \in \Pi} \int Z(\theta) d(P(\theta))$ is the lower (unconditional) expectation of a function Z of θ . Similarly, the upper conditional expectation $\bar{E}[\theta|\bar{x}]$ is the unique solution β to the equation

$$\bar{E}[(\theta - \beta)L(\theta)] = 0, \tag{3.6}$$

in which $\bar{E}[Z(\theta)] = \sup_{P \in \Pi} \int Z(\theta) d(P(\theta))$ is the upper (unconditional) expectation of Z . Thus, one can express conditional expectations in terms of unconditional ones.

Upper and lower (unconditional) expectations of a function Z of θ can be calculated by (Wasserman, 1990a, b)

$$\underline{E}[Z(\theta)] = \int \inf_{\theta^* \in \Gamma(\theta)} Z(\theta^*) \pi_0(\theta) d\theta, \tag{3.7}$$

and

$$\bar{E}[Z(\theta)] = \int \sup_{\theta^* \in \Gamma(\theta)} Z(\theta^*) \pi_0(\theta) d\theta. \tag{3.8}$$

¹ In general, the lower conditional expectation of a function Z of θ is given by $\underline{E}[Z(\theta)|\bar{x}] = \inf_{P \in \Pi} \int Z(\theta) d(P_{\bar{x}}(\theta))$. Similarly for the upper conditional expectation – simply replace the inf with a sup.

Recall that $\Gamma(\theta)$ is a union of closed intervals; therefore, if Z is continuous, then $Z(\theta^*)$ attains its inf or sup on $\Gamma(\theta)$. If Z is a positive function of θ , then it might be easier in some cases to calculate the expectations using the following formulas

$$\underline{E}[Z(\theta)] = \int_0^\infty \underline{P}\{\theta : Z(\theta) > t\} d\theta, \tag{3.9}$$

and

$$\bar{E}[Z(\theta)] = \int_0^\infty \bar{P}\{\theta : Z(\theta) > t\} d\theta, \tag{3.10}$$

(Young, 1998). In the following example, we use the formulas in (3.9) and (3.10) to calculate expectations; however, we use (3.7) and (3.8) in the example in Section 4.

Example 3.1 (continued): Continue Example 3.1 by calculating the (prior) lower and upper expectations of θ , and then calculating the posterior lower and upper expectations of θ , given a single observation x .

$$\begin{aligned} \underline{E}[\theta] &= \int_0^{2000} \underline{P}\{\theta : \theta > t\} dt \\ &= \int_0^{1000} 1 dt + \int_0^{2000-d} \frac{2000 - (t + d)}{1000} dt + \int_{2000-d}^{2000} 0 dt \\ &= 1500 - d + \frac{d^2}{2000}, \end{aligned}$$

and

$$\begin{aligned} \bar{E}[\theta] &= \int_0^{2000} \bar{E}\{\theta : \theta > t\} dt \\ &= \int_0^{1000+d} 1 dt + \int_{1000+d}^{2000} \frac{2000 - (t - d)}{1000} dt \\ &= 1500 + d - \frac{d^2}{2000}. \end{aligned}$$

It is straightforward to check that one obtains the same expectations by using (3.7) and (3.8).

Next, use (3.5) to calculate the lower expectation of θ given x :

$$0 = \underline{E}[(\theta - \alpha)L(\theta)] = \int_{\max(\frac{x}{3}, 1000)}^{2000} \inf_{\theta^* \in \Gamma(\theta)} \left(\frac{\theta^* - \alpha}{\theta^*} \right) \frac{1}{1000} d\theta.$$

If $x \leq 2000 + 2d$, then

$$0 = \int_{\max(\frac{x}{2}, 1000)}^{1000+d} \left(1 - \frac{\alpha}{1000}\right) \frac{1}{1000} d\theta + \int_{1000+d}^{2000} \left(1 - \frac{\alpha}{\theta - d}\right) \frac{1}{1000} d\theta,$$

from which it follows that

$$\alpha = \underline{E}[\theta|x] = \frac{\min(4000 - x, 2000)}{\frac{\min(2000 + 2d - x, 2d)}{1000} + 2 \ln\left(\frac{2000 - d}{1000}\right)}.$$

Similarly, if $x > 2000 + 2d$, then

$$\underline{E}[\theta|x] = \frac{4000 - x}{2 \ln\left(\frac{4000 - 2d}{x - 2d}\right)}.$$

Finally, use (3.6) to calculate the upper expectation of θ given x :

$$0 = \bar{E}[(\theta - \beta)L(\theta)] = \int_{\max(\frac{x}{2}, 1000)}^{2000} \sup_{\theta^* \in \Gamma(\theta)} \left(\frac{\theta^* - \beta}{\theta^*}\right) \frac{1}{1000} d\theta.$$

If $x \leq 4000 - 2d$, then

$$0 = \int_{\max(\frac{x}{2}, 1000)}^{2000-d} \left(1 - \frac{\beta}{\theta + d}\right) \frac{1}{1000} d\theta + \int_{1000+d}^{2000} \left(1 - \frac{\beta}{2000}\right) \frac{1}{1000} d\theta,$$

from which it follows that

$$\beta = \bar{E}[\theta|x] = \frac{\min(4000 - x, 2000)}{\frac{d}{1000} + 2 \ln\left(\frac{4000}{\max(x + 2d, 2000 + 2d)}\right)}.$$

Similarly, if $x > 4000 - 2d$, then $\bar{E}[\theta|x] = 2000$.

In Figure 3.1, we graph the base posterior expectation,

$$E_0[\theta|x] = \frac{4000 - \max(x, 2000)}{2 \ln\left(\frac{4000}{\max(x, 2000)}\right)},$$

along with the lower and upper expectations of θ given x , for $d = 100$ and $d = 500$. Note that the intervals for $d = 500$ are much wider than those for $d = 100$, as expected.

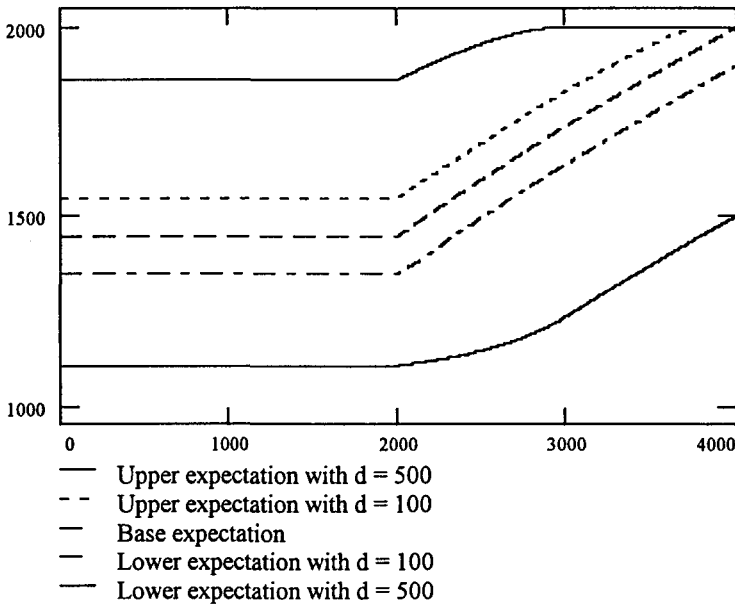


FIGURE 3.1: Lower, Upper, and Base Conditional Expectations

In general, it is difficult to calculate the lower and upper conditional expectations via equations (3.5) and (3.6). To calculate the lower conditional expectation, Walley (1991, Chapter 8) suggests using the following algorithm:

1. Let α_0 be an initial estimate of $\underline{E}[\theta|\bar{x}]$. For example, one might set α_0 equal to the Bühlmann-Straub credibility estimate minus c standard errors, if $\Gamma(\theta) = [\theta - c se(\theta), \theta + c se(\theta)]$, as in the example in Section 4.
2. Define $\alpha_{n+1} = \alpha_n + 2 \frac{\underline{E}[(\theta - \alpha_n)L(\theta)]}{\underline{E}[L(\theta)] + \underline{E}[L(\theta)]}$, $n = 0, 1, 2, \dots$

The sequence $\{\alpha_n\}$ converges to $\underline{E}[\theta|\bar{x}]$. Similarly, one can calculate $\bar{E}[\theta|\bar{x}]$ recursively.

If $\bar{E}[\theta|\bar{x}] - \underline{E}[\theta|\bar{x}]$ is small, then one says that the credibility estimate $E[\theta|\bar{x}] = \int \theta \pi_0(\theta|\bar{x})d\theta$ is a robust estimate of the conditional expectation $E[Y|\theta]$.

4. ILLUSTRATIVE EXAMPLE

In this section, we apply our robust Bayesian method to data from Dannenburg et al. (1996, Section 2.4) and calculate a credibility estimate of future claims Y given \bar{x} and corresponding lower and upper conditional expectations. Consider an insurance portfolio of nine fleets of cars, $i = 1, 2, \dots, 9$, which have been observed for ten years, $t = 1, 2, \dots, 10$. For the i^{th} fleet in year t , x_{it} is the average insurance claim over w_{it} cars. See Table 2.1 in Dannenburg et al. (1996, p. 35) for a list of the data. In Table 4.1, we give the average for each fleet \bar{x}_i , together with the standard error of \bar{x}_i , and exposure w_i (= car years).

TABLE 4.1
Average Claims for Each Fleet over the Ten Years

Fleet	1	2	3	4	5	6	7	8	9
Average	509.3	178.2	300.5	359.9	653.9	176.9	441.1	506.4	795.3
Std Error	16.29	34.74	134.5	64.30	59.93	103.0	32.63	84.27	237.7
Exposure	526	250	60	138	174	40	158	128	36

Assume that the conditional density of the underlying individual claim $Y|\theta$ is given by a normal density with mean θ and known variance σ^2 ; that is,

$$f(y|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\theta)^2}{2\sigma^2}}, \quad y > 0.$$

Given that a risk's claim experience is $\mathbf{x} = (x_1, \dots, x_J)$ with exposure $\mathbf{w} = (w_1, \dots, w_J)$, the likelihood of θ is, up to a constant multiple,

$$L(\theta) = \exp\left(-\frac{w(\theta - \bar{x})^2}{2\sigma^2}\right),$$

in which $w = \sum_{j=1}^J w_j$ and $\bar{x} = \frac{\sum_{j=1}^J w_j x_j}{w}$. Note that the variance σ^2 is the expected process variance, so we estimate it by s^2 given by (Dannenburg et al., 1996)

$$s^2 = \frac{1}{9(10-1)} \sum_{i,t} w_{it} (x_{it} - \bar{x}_i)^2 \approx (833.73)^2.$$

Two commonly used symmetric kernels are (1) the Gaussian kernel, G ,

$$G(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}, \quad -\infty < t < \infty,$$

and (2) the Epanechnikov kernel, *Epa*,

$$Epa(t) = \begin{cases} \frac{3}{4} \frac{\left(1 - \frac{t^2}{5}\right)}{\sqrt{5}}, & -\sqrt{5} < t < \sqrt{5} \\ 0, & \text{else.} \end{cases}$$

In this example, we use the Epanechnikov kernel because its domain is bounded, and one can, therefore, easily restrict the support of $\pi_0(\theta)$ to lie in the nonnegative real numbers.

There are many techniques for choosing the window width h_i ; see, for example, Silverman (1986, Section 3.4). We use a (modified) fixed window width selected by reference to a standard distribution (Silverman, 1986, Section 3.4.2). The window width h that minimizes the mean integrated squared error is given by

$$h = \left\{ \int t^2 K(t) dt \right\}^{-2/5} \left\{ \int K(t)^2 dt \right\}^{1/5} \left\{ \int \pi''(\theta) d\theta \right\}^{-1/5} r^{-1/5}. \tag{4.1}$$

To approximate this optimal window width h , assume that $\pi(\theta)$ is say, normal, with mean 0 and standard deviation τ . In that case, the term $\int \pi''(\theta) d\theta$ equals $\frac{3}{8} \pi^{-1/2} \tau^{-5}$. Estimate the standard deviation τ by an estimate of the standard deviation of the hypothetical means (Dannenburg et al., 1996)

$$\sqrt{\frac{w_{tot} \sum_{i=1}^9 w_i (\bar{x}_i - \bar{\bar{x}})^2 - 8s^2}{w_{tot}^2 - \sum_{i=1}^9 w_i^2}} \approx 161.85,$$

in which $s^2 \approx (833.73)^2$.

The bandwidth h is $(1)^{-2/5} (0.268)^{1/5} (0.212)^{-1/5} (161.85) (9)^{-1/5} \approx 109.4$, as given by (4.1). Truncate this bandwidth h for a given risk if, by otherwise using it, the prior density would have a negative support. Specifically, if $h > \frac{\bar{x}_i}{\sqrt{5}}$, then set the bandwidth h_i equal to $\frac{\bar{x}_i}{\sqrt{5}}$, to guarantee that the support of the estimated density of θ be contained in the nonnegative real numbers. See Figure 4.1 for a graph of the prior density π_0 , given by (2.2). One can also vary the window width with the weight w_i for each risk, $i = 1, 2, \dots, I$, for example, $h_i \propto \frac{1}{\sqrt{w_i}}$.

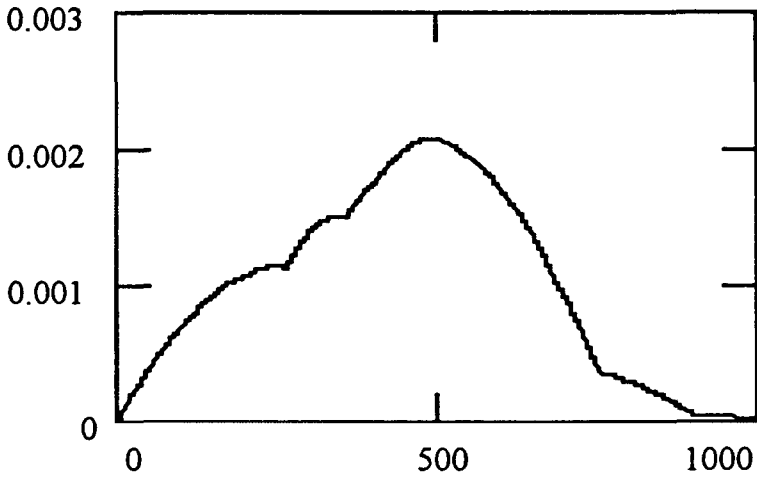


FIGURE 4.1: Base Prior Density π_0

See Table 4.2a for a list of the posterior expectations of θ given the experience of the nine fleets, in which π_0 is the prior density. Also in Table 4.2a are the lower and upper conditional expectations, (3.3) and (3.4), respectively, using $c = 1$ and 2 in the definition of the multivalued mapping Γ , given by $\Gamma(\theta) = [\theta - c se(\theta), \theta + c se(\theta)]$. To create se , we linearly interpolated the nine ordered pairs $(\bar{x}_i, se(\bar{x}_i))$. See the Appendix for more details concerning the calculation of the entries in Table 4.2a. In Table 4.2b are the differences between the lower conditional expectations and the base conditional expectation, as well between the upper and base conditional expectations.

TABLE 4.2a

Posterior Expectations for the Nine Auto Fleets using Local Perturbation Priors

<i>Fleet</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>
Lower, $c=2$	453	76	226	278	500	85	357	433	479
Lower, $c=1$	473	128	270	316	558	170	395	457	537
Expectation	509	187	329	372	631	246	447	504	661
Upper, $c=1$	561	273	418	456	688	371	503	557	785
Upper, $c=2$	580	308	479	519	725	419	540	589	841

TABLE 4.2b
Distance From Base Expectation to Lower and Upper Conditional Expectations
using Local Perturbation Priors

<i>Fleet</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>
Lower, $c=2$	-56	-111	-103	-94	-131	-161	-90	-71	-182
Lower, $c=1$	-36	-59	-59	-56	-73	-76	-52	-47	-124
Upper, $c=1$	52	86	89	84	57	125	26	53	124
Upper, $c=2$	71	121	150	147	94	173	93	85	180
Exposure	526	250	60	138	174	40	158	128	36

Note in Table 4.2b, that although fleet 8 has much lower exposure than fleet 2, fleet 8 has a tighter interval for its conditional expectations than does fleet 2. This occurs because the sample average of fleet 8 is closer to the center of the prior density, as seen in Figure 4.1. This phenomenon is also observed when one compares fleets 5 and 7 and fleets 4 and 8, two fleets with nearly equal exposure, as well as fleets 7 and 8, to some extent. Thus, the intervals for the conditional expectations of a fleet, a measure of the robustness of the base expectation given in Table 4.2a, depends on the fleet's exposure and on the location of its sample mean relative to the other sample means.

For comparison, we also include the Bühlmann-Straub credibility estimates in Table 4.3, along with the estimates ± 1 and ± 2 standard errors, (Dannenburg et al., 1996) or (Frees et al., 1998). Note that the standard error of a fleet's credibility estimate, a measure of the confidence one has in the estimate, varies inversely with the fleet's exposure and is not related to the location of its sample mean. In other words, the width of the Bühlmann-Straub confidence intervals accounts only for the exposure of the fleet, while ignoring the 'likelihood' of observing the given sample mean. On the other hand, the intervals of the conditional expectations for the local perturbation priors depend on both the exposure and the location of the sample mean. They implicitly treat 'unusual' sample means, observations in the tails of the distribution, as being less reliable.

TABLE 4.3
Credibility Estimates for the Nine Auto Fleets using the Bühlmann-Straub Model

<i>Fleet</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>
Lower, -2 se	434	102	158	240	506	69	316	358	425
Lower, -1 se	470	152	250	306	565	174	378	426	533
Cred est	506	203	341	372	625	279	440	494	642
Upper, $+1 \text{ se}$	541	253	433	438	684	384	502	562	750
Upper, $+2 \text{ se}$	577	304	524	503	744	489	564	630	859
Std Error	36	51	91	66	60	105	62	68	109

5. SUMMARY AND CONCLUSIONS

We have presented a method for calculating robust Bayesian confidence intervals for the predictive mean of future insurance claims given the sample mean of past insurance claims. We showed how our method can be framed in the context of the Dempster-Shafer theory of belief functions. Within this framework, one can calculate robust Bayesian intervals numerically by using an iterative algorithm given by Walley (1991). We demonstrated our procedure with insurance claim data from Dannenburg et al. (1996). We showed that robust Bayesian intervals account for both the randomness of the data and for the location of the sample mean relative to the other sample means, while (classical) nonparametric confidence intervals account for the former but not the latter.

Berger (1994, p. 8) eloquently argues for robust Bayesian methods: 'Rightly or wrongly, the majority of the statistical world resists use of Bayesian methods. The most often vocalized reason is fear of using a subjective prior ... Robust Bayesian methods, which can operate with a wide class of prior distributions, ... seems to be an effective way to eliminate this fear.'

ACKNOWLEDGEMENTS

I thank Larry Wasserman for encouraging comments. I also thank the Graduate School and the School of Business of the University of Wisconsin-Madison for financial support of this project.

APPENDIX

Recall that the lower conditional expectation $\underline{E}[\theta|\bar{x}]$ is the unique solution α to the equation (3.5)

$$\underline{E}[(\theta - \alpha)L(\theta)] = 0,$$

(Walley, 1991, Chapter 8). Walley (1991, p. 550) gives an algorithm for calculating $\underline{E}[\theta|\bar{x}]$, and we describe it at the end of section 3. To apply this algorithm in our example in Section 4, we must be able to calculate $\underline{E}[(\theta - \alpha)L(\theta)]$, $\bar{E}[L(\theta)]$, and $\underline{E}[L(\theta)]$, for any positive real number α and for the likelihood function L and perturbation function Γ given in Section 4.

We use equations (3.7) and (3.8) to calculate these expectations. Specifically,

$$\underline{E}[(\theta - \alpha)L(\theta)] = \int \inf_{\theta^* \in \Gamma(\theta)} (\theta^* - \alpha)L(\theta^*)\pi_0(\theta)d\theta.$$

The function $(\theta^* - \alpha)L(\theta^*)$ reaches its infimum at either the endpoints $l(\theta)$ or $u(\theta)$ of $\Gamma(\theta) \cap R^+$ or at the critical points θ_+ or θ_- (depending on whether

the critical points lie in $\Gamma(\theta) \cap R^+$, in which $l(\theta) = \theta - c se(\theta)$ if $\theta > c se(\theta)$, 0 else; $u(\theta) = \theta + c se(\theta)$; and $\theta_{\pm} = \frac{1}{2} \left[(\alpha + \bar{x}) \pm \sqrt{(\alpha - \bar{x})^2 + 4 \frac{\sigma^2}{w}} \right]$. Thus,

$$\inf_{\theta^* \in \Gamma(\theta)} (\theta^* - \alpha)L(\theta^*) = \begin{cases} \min[(l(\theta) - \alpha)L(l(\theta)), (u(\theta) - \alpha)L(u(\theta))], & \text{if } \theta_- < \theta_+ < l(\theta) < u(\theta), \\ \min[(l(\theta) - \alpha)L(l(\theta)), (\theta_+ - \alpha)L(\theta_+), (u(\theta) - \alpha)L(u(\theta))], & \text{if } \theta_- < l(\theta) < \theta_+ < u(\theta), \\ \min[(l(\theta) - \alpha)L(l(\theta)), (u(\theta) - \alpha)L(u(\theta))], & \text{if } \theta_- < l(\theta) < u(\theta) < \theta_+, \\ \min[(l(\theta) - \alpha)L(l(\theta)), (\theta_- - \alpha)L(\theta_-), (\theta_+ - \alpha)L(\theta_+), (u(\theta) - \alpha)L(u(\theta))], & \text{if } l(\theta) < \theta_- < \theta_+ < u(\theta), \\ \min[(l(\theta) - \alpha)L(l(\theta)), (\theta_- - \alpha)L(\theta_-), (u(\theta) - \alpha)L(u(\theta))], & \text{if } l(\theta) < \theta_- < u(\theta) < \theta_+, \\ \min[(l(\theta) - \alpha)L(l(\theta)), (u(\theta) - \alpha)L(u(\theta))], & \text{if } l(\theta) < u(\theta) < \theta_- < \theta_+. \end{cases}$$

Similarly,

$$\bar{E}[L(\theta)] = \int \sup_{\theta^* \in \Gamma(\theta)} L(\theta^*) \pi_0(\theta) d\theta,$$

in which

$$\sup_{\theta^* \in \Gamma(\theta)} L(\theta^*) = \begin{cases} L(l(\theta)), & \text{if } \bar{x} < l(\theta) < u(\theta), \\ L(\bar{x}), & \text{if } l(\theta) < \bar{x} < u(\theta), \\ L(u(\theta)), & \text{if } l(\theta) < u(\theta) < \bar{x}. \end{cases}$$

Also,

$$\bar{E}[L(\theta)] = \int \inf_{\theta^* \in \Gamma(\theta)} L(\theta^*) \pi_0(\theta) d\theta,$$

in which

$$\inf_{\theta^* \in \Gamma(\theta)} L(\theta^*) = \begin{cases} L(u(\theta)), & \text{if } \bar{x} < l(\theta) < u(\theta), \\ \min[L(l(\theta)), L(u(\theta))], & \text{if } l(\theta) < \bar{x} < u(\theta), \\ L(l(\theta)), & \text{if } l(\theta) < u(\theta) < \bar{x}. \end{cases}$$

One calculates the ingredients for the algorithm for $\bar{E}[\theta|\bar{x}]$ similarly. Specifically, replace min with max in calculating $\sup_{\theta^* \in \Gamma(\theta)} (\theta^* - \alpha)L(\theta^*)$.

REFERENCES

- BERGER, J.O. (1994), An overview of robust Bayesian analysis, with discussion, *Test*, 3: 5-124.
- BÜHLMANN, H. (1967), Experience rating and credibility, *ASTIN Bulletin*, 4: 199-207.
- BÜHLMANN, H. (1970), *Mathematical Models in Risk Theory*, Springer-Verlag, New York.
- BÜHLMANN, H. and E. STRAUB (1970), Glaubwürdigkeit für Schadensätze, *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker*, 70: 111-133.
- DANNENBURG, D.R., R. KAAS and M.J. GOOVAERTS (1996), *Practical Actuarial Credibility Models*, Institute of Actuarial Science and Econometrics, University of Amsterdam, Amsterdam, The Netherlands.
- DEMPSTER, A.P. (1967), Upper and lower probabilities induced by a multivalued mapping, *Annals of Mathematical Statistics*, 38: 325-339.
- DEMPSTER, A.P. (1968), A generalization of Bayesian inference, with discussion, *Journal of the Royal Statistical Society, Series B*, 30: 205-247.
- DENNEBERG, D. (1994), *Non-Additive Measure and Integral*, Kluwer, Dordrecht.
- FREES, E.W., V.R. YOUNG and Y. LUO (1998), A longitudinal data analysis interpretation of credibility models, working paper.
- SERFLING, R.J. (1980), *Approximation Theorems of Mathematical Statistics*, Wiley, New York.
- SHAFFER, G. (1979), Allocations of probability, *Annals of Probability*, 7: 827-839.
- SILVERMAN, B.W. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman & Hall, London.
- WALLEY, P. (1991), *Statistical Reasoning with Imprecise Probabilities*, Chapman & Hall, London.
- WASSERMAN, L.A. (1990a), Belief functions and statistical inference, *Canadian Journal of Statistics*, 18: 183-196.
- WASSERMAN, L.A. (1990b), Prior envelopes based on belief functions, *Annals of Statistics*, 18: 454-464.
- YOUNG, V.R. (1997), Credibility using semiparametric models, *ASTIN Bulletin*, 27: 273-285.
- YOUNG, V.R. (1998), Families of update rules of non-additive measures: Applications in pricing risks, to appear, *Insurance: Mathematics & Economics*.

VIRGINIA R. YOUNG
School of Business
975 University Avenue
University of Wisconsin-Madison
Madison, WI USA 53706