*Chilean Journal of Statistics*
Vol. 6, No. 2, September 2015, 3–17

Decision Theory
Research Paper

# Credibility premiums for natural exponential family and general 0–1 loss function

Emilio Gómez–Déniz[1] and Enrique Calderín–Ojeda[2,*]

[1]Department of Quantitative Methods in Economics and TiDES Institute.
University of Las Palmas de Gran Canaria, Spain,
[2]Centre for Actuarial Studies, Department of Economics
The University of Melbourne, Australia

(Received: 29 May 2014 · Accepted in final form: 27 June 2015)

**Abstract**

In credibility theory, the premium charged to a policyholder is computed on the basis of his past claims and the accumulated past claims of the corresponding portfolio of policyholders. In order to obtain an appropriate expression for this quantity, different approaches within Bayesian statistical decision theory, such as empirical Bayes, gamma–minimax and posterior regret gamma–minimax estimation have been proposed in the actuarial literature. This paper introduces a new parametric family of Bayesian estimators into the methodology of premium calculation principles. This is achieved by using a general class of 0–1 loss function when the natural exponential family of distribution is used. Consequently, new credibility expressions are obtained.

**Keywords:** Bayesian · Credibility · Natural Exponential Family · Loss Function

**Mathematics Subject Classification:** Primary 62F15 · Secondary 62C10

## 1. Introduction

Credibility theory is an experience rating technique that it is frequently used in assessing automobile insurance, workers' compensation premium, loss reserving and IBNR–Incurred But Not Reported claims. Under this theory, policyholders' premiums are computed by combining the experience of the individual (contract or policyholder) with the experience of a collective (portfolio) by using the expression

$$P = Z(N)\bar{X} + (1 - Z(N))\wp, \tag{1}$$

where $P$ is the credibility adjusted premium, $\wp$ is the overall mean (the expected claim size for the whole portfolio), $\bar{X}$ is the mean for individual risk and $0 \leq Z(N) \leq 1$ is the credibility factor, being $N$ the time of exposition of the risk. This expression can be obtained under different methodologies of Bayesian statistical decision theory such as

*Corresponding author. Email: enrique.calderin@unimelb.edu.au

ISSN: 0718-7912 (print)/ISSN: 0718-7920 (online)

http://www.soche.cl/chjs

empirical Bayes, gamma–minimax and posterior regret gamma–minimax estimation. See Whitney (1918), Eichenauer et al. (1988), Heilmann (1989) and Gómez–Déniz et al. (2006), among others. In this regard, credibility theory is used to determine the expected claims experience of an individual risk when risks are not homogeneous, since the individual risk belongs to a heterogeneous collective. Some old and new references on this topic can be found in Bühlmann (1967), Bühlmann and Gisler (2005), Jewell (1974), Gerber and Arbor (1980), Heilmann (1989), Hooge and Goovaerts (1975), Landsman and Makov (1998, 1999), Eichenauer et al. (1988), Young (1997, 2000), Gómez–Déniz et al. (2006), Gómez–Déniz (2008), Payandeh (2010) and Payandeh et. al (2012), among others.

Basically, credibility theory deals with the problem of estimating the collective (a priori) and Bayes (experience rated) premium under an appropriate loss function (Heilmann, 1989). The weighted squared–error loss function (Furman and Zitikis, 2008) has been traditionally considered to achieve exact credibility premiums under adequate pairs of likelihood and prior distributions (Heilmann, 1989). Furthermore, as a result of its quadratic factor, a convex function is obtained. Alternatively, some authors have pointed out that Bayes estimates should be computed under bounded loss rather than convex loss (Kadane and Chuang, 1978; Smith, 1980). In this sense, convex loss functions have been traditionally preferred due to their simplicity and computational advantages, with a unique optimum value; however, they offer poor approximations to the 0–1 loss and they also show lack of robustness to outliers due to their boundlessness (Hastie et al., 2001).

For the suggestions mentioned above, the general 0–1 loss function appearing in Berger (1985, p.63) is considered in this work. One of its advantages is that, unlike the weighted squared–error loss function, the computation of the Bayes estimator does not require integration. On this subject, its simplicity is based on the fact that only maximization of a given function is needed. This obviously provides more flexibility to choose the prior distribution. Consequently, new credibility expressions will be derived when the natural exponential family of distribution is considered. It is important to point out that many of the most common continuous and discrete probabilistic models, such as Poisson, binomial, negative binomial and gamma, among others, do belong to natural exponential families, which makes their manipulation particularly convenient. The 0–1 loss function has received significant attention in the area of signal processing (Vaseghi, 2000, chapter 4); however, it never has been used in credibility theory to obtain appropriate premium calculation principles. Furthermore, we show that the use of this loss leads to similar but not equal expressions for the credibility factor as those obtained in previous works for a different kind of Bayesian (and collective) premium.

The rest of the paper is organized as follows. Firstly, Section 2 reviews the procedure of premium calculation principle under the Bayesian decision theory. Next, Section 3 describes the generalized loss function used in this manuscript. General expressions for the collective and Bayesian premium are obtained under this loss when the natural exponential family of distributions and its natural conjugate prior are employed. Then, in Section 4 specific examples of new credibility expressions for different pairs of probabilistic models are shown. Later in Section 5 two numerical applications by using two different sets of actuarial data are provided. Finally, some concluding remarks are made in the last Section.

## 2. Preliminaries

A premium calculation principle can be viewed as a procedure of obtaining an estimate of an unknown risk parameter (assumption usually taken in the credibility theory) $\theta$, from observations $\underline{X}$ obtained according to a likelihood $f(\underline{X}|\theta)$, in the presence of a prior distribution (structure function) $\pi(\theta)$ for a given loss function $L(\theta, \widehat{\theta})$. Here $\widehat{\theta}$ is an estimator of $\theta$. The number of claims (or size of a claim) of one contract in one period is specified by a random variable $X \in \mathcal{X}$ following a probability density function $f(x|\theta)$ depending on an unknown risk parameter $\theta \in \Theta$. Let $X_i,\ i = 1, 2, \ldots, N$ be independent and identically distributed random variables with values in $\mathcal{X} \subset \mathbb{R}$ representing the number of claims (in the discrete case) and the claim size of a single risk (in the continuous case) during the last $N$ periods. In risk theory, a premium calculation principle $P$ assigns to each risk parameter $\theta$ a fair premium $P(\theta)$, usually called risk premium, within the set $\mathcal{P} \subset \mathbb{R}$ of admissible premiums.

In the case of the net premium principle the function $P : \Theta \longrightarrow \mathbb{R}$ is given by

$$P(\theta) = E(X|\theta), \quad \theta \in \Theta.$$

This premium calculation principle can be obtained by minimizing with respect to $P$ the expected loss $E_f\left[L(x, P)\right]$ where $L$ is taken as the squared–error loss function $L(a, x) = (x - P)^2$. Other premium calculation principles can also be generated by using a different loss function. For example for the weighted squared–error loss $L(x, P) = \omega(x)(x - P)^2$ and by taking $\omega(x) = \exp(s\,x), s > 0$ and $\omega(x) = x$ we obtain the Esscher and variance premium principles, respectively (Heilmann, 1989; Gómez–Déniz et al., 2006; Furman and Zitikis, 2008).

In practical situations, the premium computed above can only be applied if the distribution of the risk $X$ is known; likewise, the ratemaking procedure incorporates individual claim experience. Henceforward, as it is usually assumed in credibility framework, the distribution of $X$ is specified up to an unknown parameter.

Firstly, if experience is not available, the actuary computes the collective premium, $\wp$, by minimizing the risk function, i.e. minimizing $E_\pi\left[L(P(\theta), \wp)\right]$, where $\pi(\theta)$ is the prior distribution on the unknown parameter $\theta \in \Theta$. On the other hand, if experience is available, the actuary takes a sample $\underline{X} = (X_1, X_2, \ldots, X_N)$ from the random variables $X_i,\ i = 1, 2, \ldots, N$, assuming $X_i$ independent and identically distributed, and then the practitioner uses this information to estimate the unknown risk premium $P(\theta)$, through the Bayes premium $\Pi(\underline{X})$, obtained by minimizing the Bayes risk, i.e. minimizing $E_{\pi_{\underline{X}}}\left[L(P(\theta), \Pi(\underline{X}))\right]$. Here, $\pi_{\underline{X}}$ is the posterior distribution of the risk parameter, $\theta$, given the sample information $\underline{X}$.

Furthermore, as Heilmann (1989) and Gómez–Déniz (2008) have pointed out, it is possible to use different premium calculation principles to obtain the risk, collective and Bayesian premium. Hereafter, following this spirit, we will suppose that the practitioner chooses the squared–error loss function only to obtain the risk premium and a different premium calculation principle to compute the collective and Bayesian premium. In this regard, the squared–error loss function has been traditionally used by decision–theoretic statisticians and economists for many years. In addition to this, it has been preferred in actuarial statistics since the net premium derived from this loss function fulfils some desirable properties such as translation invariance, sub–additivity, iterativity and homogeneity, among others.

Let $\theta$ be a risk parameter characterizing a member of a risk collective, and given $\theta$, let now $f(x|\theta)$, the distribution of the claims (or size of a claim) $X \in \mathcal{X} \subset \mathbb{R}$, be a member of a family of distributions $\{f(x|\theta), \theta \in \Theta \subset \mathbb{R}\}$. Let us also assume that the distribution of $X$ is a member of the Natural Exponential Family (NEF) of distributions. We say that the random variable $X$, given $\Theta = \theta$, is distributed according the NEF of distributions if its probability density function (probability mass function in the discrete case) is given by

$$f(x|\theta) = q(x) \exp\left[-x\theta - \log(\kappa(\theta))\right], \tag{2}$$

where the natural parameter can be written as $\theta = \phi(\varepsilon)$ with parameter $\varepsilon$ being $\theta = \phi(\varepsilon)$ for a function $\phi$ and $\kappa(\theta) = \int_{\mathcal{X}} q(x) \exp(-x\theta)\, dx < \infty$ .

The NEF is a wide class of distributions studied by numerous authors (Diaconis and Ylvisaker, 1979; MacEachern, 1993). In the field of actuarial statistics and particularly in credibility theory, perhaps the most important work dealing with the NEF of distributions is due to Jewell (1974). In his paper, it is proved that when conjugate priors are used and the squared–error loss function is chosen to compute the risk, collective and Bayesian premium, then the latter can be written as in (1). It is important to recall that many of the most common continuous and discrete probabilistic models belong to this family of distributions, which makes their manipulation particularly convenient.

Furthermore, by using the squared–error loss function it is simple to observe that under this family of distributions the risk premium is given by $P(\theta) = E(X|\theta) = -\kappa'(\theta)/\kappa(\theta)$ and under appropriate reparameterization, the risk premium can be expressed as $P(\theta) = \theta$. This is the case of Poisson, negative binomial, binomial, gamma and normal distribution.

## 3. Using the general 0–1 loss function

Broadly speaking, loss functions are used in actuarial statistics, among other settings, to quantify losses associated with deviation from a desired target value. The loss is usually defined as a function of the deviation of an estimator from the true value of the parameter. The general statistical decision problem is described by possible states of nature $\theta \in \Theta$, decisions $\widehat{\theta}$ and a loss function $L(\theta, \widehat{\theta})$.

Let us now consider the loss function given by

$$L(\theta, \widehat{\theta}) = \begin{cases} 0, & \widehat{\theta} = \theta, \\ \eta\, g(\theta), & \widehat{\theta} \neq \theta. \end{cases} \tag{3}$$

Here, $\eta$ is a normalization constant which provides that $\eta g(\theta)\pi(\theta)$ is a proper probability distribution. When the function $g(x)$ is chosen such that $g(x) > 0$, the function above represents a genuine loss function. Here $\theta \in \Theta$ represents the state of nature and $\widehat{\theta}$ the action to be taken. Therefore, the loss function given in (3), as whichever loss function, measures the loss sustained by a decision–maker who takes the action $\widehat{\theta}$ instead of the true value of the parameter $\theta$. In this case the loss is zero if a correct decision is taken and $\eta g(\theta) > 0$ if a wrong decision is made.

The loss function (3) includes as special cases the 0–1 loss function when $\eta g(\theta) = 1$ and the 0–$k$ loss function when $\eta g(\theta) = k > 0$, being $k$ an appropriate constant value. See Berger (1985, p.63) for details on their use. The latter author pointed out that this general loss function is needed to measure in some sense the severity of the mistake by taking an incorrect decision.

Although convex losses have been traditionally preferred because of their simplicity and computational advantages, with a unique optimum value; however, they offer poor approximations to the 0–1 loss and they also show lack of robustness to outliers due to their boundlessness (Hastie et al., 2001). On this subject, some authors have pointed out that Bayes estimates should be made under bounded loss rather that convex loss (Kadane and Chuang, 1978; Smith, 1980; Kamińska and Porosiński, 2009). This is the case when the loss function given in (3) is taken and $g(\theta)$ is chosen as a bounded function or simply a constant value, said $k$. As it can be observed we have that

$$\inf_{\theta,\widehat{\theta}} L(\theta,\widehat{\theta}) = 0,$$

$$\sup_{\theta,\widehat{\theta}} L(\theta,\widehat{\theta}) = \eta g(\theta) > 0.$$

Figure 1 illustrates the loss sustained under squared–error loss function (SEL) and the general 0–1 loss function introduced above. As it can be observed the error is 0, i.e. the same for the two losses, when $\widehat{\theta} = \theta$. When those values do not coincide, then it can occur that the error under SEL is either smaller or larger than the one calculated under the general 0–1 loss function. In any case, the general 0–1 loss function guarantees that an upper bound on the loss can be achieved. Besides, a bounded loss avoids the potential explosion of the squared–error loss function.

On the other hand by choosing the function $g(\theta)$ appropriately, it is obvious that the loss is not symmetric. As Kamińnska and Porosiński (2009) pointed out, there exist situations where over and under estimation can lead to different consequences and therefore, we should consider estimators based on an asymmetric and bounded loss.

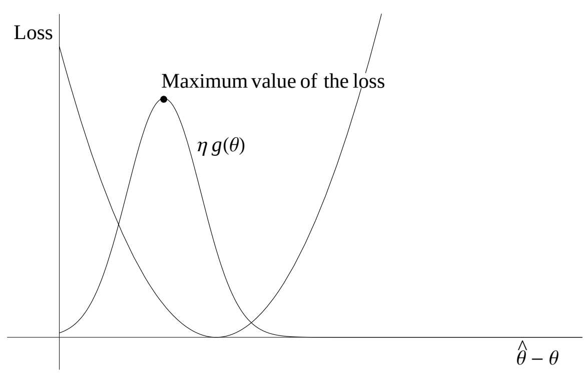


Figure 1. Illustration of the loss sustained under quadratic error loss function and the general 0-1 loss function

As we mentioned in the previous section, if experience is not available, the actuary computes the collective premium, $\wp$, which is given by minimizing the risk function, i.e. minimizing $E_\pi\left[L(P(\theta),\wp)\right]$, where $P(\theta)$ is the unknown risk premium and $\pi(\theta)$ is the prior distribution on the unknown parameter $\theta \in \Theta$. Likewise, the collective premium under the general 0–1 loss function given in (3) is provided by the following proposition.

Proposition 3.1 Let us assume that the risk premium is given by $P(\theta) = \theta > 0$. Then, for the general 0–1 loss function given in (3) the collective premium is given by

$$\wp = \arg\max_{\theta} g(\theta)\pi(\theta). \tag{4}$$

Proof Let $\pi^*(\theta) = \eta g(\theta)\pi(\theta)$, then the result follows by taking into account that under the loss function (3) $\wp$ coincides with the maximum a priori estimate given by

$$\wp = \arg\max_{\theta} \pi^*(\theta) = \arg\max_{\theta} g(\theta)\pi(\theta).$$

■

The Bayes premium, $\Pi(\underline{X})$, is computed in a similar way by replacing $\pi^*(\theta)$ by the posterior $\pi^*(\theta|\underline{X})$. Then $\Pi(\underline{X})$ is obtained as

$$\Pi(\underline{X}) = \arg\max_{\theta} \pi^*(\theta|\underline{X}), \tag{5}$$

which can be written as

$$\begin{aligned}\Pi(\underline{X}) &= \arg\max_{\theta} \pi^*(\theta|\underline{X}) \\ &= \arg\max_{\theta} \{f(\underline{X}|\theta)\pi^*(\theta)\} \\ &= \arg\max_{\theta} \{\log f(\underline{X}|\theta) + \log(\pi^*(\theta))\}. \end{aligned} \tag{6}$$

Expression (6) emphasizes on the fact that the estimator tries to reach a compromise between the a priori knowledge about the risk parameter and the evidence provided by the data via the likelihood function.

One of its advantages is that, unlike the weighted squared–error loss function, the computation of the Bayes estimator does not require integration. On this subject, its simplicity is based on the fact that only maximization of a given function is needed. This clearly provides more flexibility to choose the prior distribution.

As mentioned above, the loss function (3) includes as a particular case the 0–1 loss function (Berger, 1985 and Robert, 2007) when $g(x) = 1$. In this case (5) extracts the value of $\theta$ that causes the maximum of the posterior distribution and it is known in the Bayesian literature as the maximum a posteriori (MAP) estimator (the mode), which coincides with the maximum likelihood estimator of the posterior distribution of the parameter $\theta$. See Micheas and Dey (2004) for a detailed reading of this Bayesian estimator. In this regard, the optimal estimator is the most likely value of $\theta$ given the data and the prior $\pi(\theta)$. If $\theta$ is assumed to be fixed but unknown, then there is no knowledge about $\theta$, which is equivalent to a non–informative improper prior, i.e. $\pi(\theta)$ constant. Then, equation (5) reduces to the familiar maximum likelihood formulation.

Let us now consider the natural conjugate prior for $\theta$, given by

$$\pi(\theta) = \exp\{-\theta x_0 - n_0 \log(\kappa(\theta)) - \log d(n_0, x_0)\}, \tag{7}$$

with $d(n_0, x_0) = \int_\Theta \exp\{-\theta x_0 - n_0 \log(\kappa(\theta))\}\, d\theta$.

It is easy to see that the posterior distribution of the parameter $\theta$ when $\underline{X} = (X_1, X_2, \ldots, X_n)$ is observed is given by

$$\pi(\theta|\underline{X}) \propto \exp\{-\theta x_0^* - n_0^* \log(\kappa(\theta))\},$$

where $n_0^*$ and $x_0^*$ are the updated parameters given by $n_0^* = n_0 + N$ and $x_0^* = x_0 + N\bar{X}$, being $\bar{X} = (1/N)\sum_{i=1}^{N} X_i$ the sample mean.

In the following we will consider $\varepsilon$ as natural parameter. Next result provides the collective premium under the general 0–1 loss function given in (3).

Theorem 3.2 Let us assume that the number of claims or size of a claim follows the distribution given in (2) where the function $\phi$ is monotically decreasing and the prior distribution of the risk parameter $\varepsilon$ is given in (7) with $n_0 > 0$. Let us also suppose that $g'(\varepsilon) = g(\varepsilon)h(\varepsilon)$, with $h(\varepsilon) = a + b\,\phi(\varepsilon)$, $b > 0$ and $g(\varepsilon) > 0$, and that the practitioner chooses the net premium principle to compute the risk premium and the premium based in the general 0–1 loss function to compute the collective and the Bayes premium. Then, the collective premium is given by

$$\wp = \frac{x_0 - a}{n_0 + b}. \tag{8}$$

Proof By using (4), the derivative of the function $\Phi(\varepsilon) = g(\varepsilon)\,\pi(\varepsilon)$ must be computed and set equal to zero. In this case we get

$$\Phi'(\varepsilon) = g(\varepsilon)\pi(\varepsilon)\left[a + b\phi(\varepsilon) - x_0 - n_0\frac{\kappa'(\varepsilon)}{\kappa(\varepsilon)}\right].$$

Now, since $\theta = \phi(\varepsilon) = -\frac{\kappa'(\varepsilon)}{\kappa(\varepsilon)}$, after equating to zero and by isolating $\theta$ expression (8) is obtained.

In order to ensure that the value of $\wp$ is a maximum and having into account that the derivative of $\kappa'(\varepsilon)/\kappa(\varepsilon)$ is the variance of $X|\varepsilon$, we have, after some algebra, that $\Phi''(\varepsilon)|_{\varepsilon=\wp} = g(\wp)\pi(\wp)b\,\phi'(\wp) - n_0\,var(X|\varepsilon = \wp) < 0$.

Then the theorem holds. ■

In practice, the Bayes premium is directly obtained from the collective premium when conjugate distributions are used, by only replacing the parameters upon which the collective premium depends by the updated parameters. Then, after replacing $n_0$ and $x_0$ in (8) by the updated parameters, we obtain the Bayes premium,

$$\Pi(\underline{X}) = \frac{x_0 + N\bar{X} - a}{n_0 + N + b}. \tag{9}$$

Note that (9) can be rewritten as a credibility expression as in (1) with the credibility factor given by $Z(N) = N/(n_0 + N + b)$.

## 4. Specific examples and applications

We are now interested in illustrating the methodology described in the previous section. For that reason, five examples are considered to supply specific credibility premiums under different pairs of likelihood and prior distributions. Certainly, we must focus our work on conjugate families of distributions. In the following, we will assume that the practitioner chooses the quadratic loss to obtain the risk premium and the loss (3), with appropriate function $g(x)$, to obtain the collective and the Bayes premium.

### The Poisson–Gamma case

Let us suppose that the number of claims of an insurer follows a Poisson distribution with parameter $\theta > 0$. In this case $q(x) = 1/x!$, the natural parameter is $\varepsilon = -\log\theta$ and $\kappa(\varepsilon) = \exp(\exp(-\varepsilon))$. Therefore $P(\theta) = -\kappa'(\varepsilon)/\kappa(\varepsilon) = \exp(-\varepsilon) = \theta$.

Let us now suppose that the prior is a gamma distribution $\pi(\theta) \propto \theta^{\sigma-1}e^{-\lambda\theta}$, $\sigma > 0$, $\lambda > 0$ and let us also assume that $g(x) = x^{\gamma}\exp(-c\,x)$, $\gamma > 0$, $c > 0$, $x > 0$, which gives an upper bound of the loss function given by $(\gamma/c)^{\gamma}e^{-\gamma}$. In this case, we have that $\phi(\varepsilon) = \exp(-\varepsilon)$ and $h(\varepsilon) = -\gamma + c\,\phi(\varepsilon)$. Therefore, we have that $a = -\gamma$, $b = c$ and then, the collective premium is obtained by using (8) and having into account that $x_0 = \sigma - 1$

and $n_0 = \lambda$, getting

$$\wp = \frac{\gamma + \sigma - 1}{\lambda + c}, \quad \sigma > \gamma - 1. \tag{10}$$

By replacing in (10) $\sigma$ and $\lambda$ by the updated parameters, we obtain the Bayes premium

$$\Pi(\underline{X}) = \frac{N\bar{X} + \sigma + \gamma - 1}{\lambda + N + c}. \tag{11}$$

Remark 4.1 Observe that when $\sigma = 1$, we are under the Poisson–exponential model and,in this case, $\wp = \gamma/(\lambda + c)$ and $\Pi(\underline{X}) = (N\bar{X} + \gamma)/(\lambda + N + c)$.

### The Binomial–Beta case

Let us now suppose that the number of claims of an insurer follows a binomial distribution with parameters $n > 0$ (assumed known), $0 < \theta < n$, and probability mass function given by

$$f(x|\theta) = \binom{n}{x}\left(\frac{\theta}{n}\right)^x \left(1 - \frac{\theta}{n}\right)^{n-x}, \quad x = 0, 1, \ldots$$

Now, $q(x) = \binom{n}{x}$, the natural parameter is $\varepsilon = \log(n/\theta - 1)$ and $\kappa(\varepsilon) = (1 + \exp(-\varepsilon))^n$. Therefore $P(\theta) = -\kappa'(\varepsilon)/\kappa(\varepsilon) = \theta$. Let us suppose that the prior is a shifted beta distribution with pdf

$$\pi(\theta) = \frac{1}{n^\alpha B(\alpha, \beta)}\theta^{\alpha-1}\left(1 - \frac{\theta}{n}\right)^{\beta-1}, \; 0 < \theta < n, \; \alpha > 0, \; \beta > 0,$$

where $B(\alpha, \beta)$ represents the beta function and let us also assume that $g(x) = x^\gamma(1 - x/n)^c$, $\gamma > 0$, $c > 0$, $0 < x < n$, which gives an upper bound of the loss function given by $(\frac{\gamma n}{c+\gamma})^\gamma(1 - \frac{\gamma}{c+\gamma})^c$. Besides, $\phi(\varepsilon) = n/(1 + \exp(\varepsilon))$. Some tedious computations yields $h(\varepsilon) = -\gamma + \frac{c+\gamma}{n}\phi(\varepsilon)$ and $x_0 = \alpha - 1$, $n_0 = (\beta + \alpha - 2)/n$. Therefore, $a = -\gamma$, $b = \frac{c+\gamma}{n}$ and the collective premium is

$$\wp = \frac{n(\alpha + \gamma - 1)}{\alpha + \beta + c + \gamma - 2}, \quad \alpha > \gamma - 1. \tag{12}$$

Observe again that $h'(x) = -\gamma/x^2 - c/(n(1 - x/n))^2 < 0$, ensuring that the estimator of the risk premium corresponds to a maximum.

By replacing in (12) $\alpha$ and $\beta$ by the updated parameters, we obtain the Bayes premium

$$\Pi(\underline{X}) = \frac{n(\alpha + N\bar{X} + \gamma - 1)}{\alpha + \beta + N\,n + c + \gamma - 2}. \tag{13}$$

The negative binomial–inverted gamma case

Let us assume that the number of claims follows a negative binomial distribution with probability mass function

$$f(x|\theta) = \binom{r+x-1}{x} \left(\frac{r}{r+\theta}\right)^r \left(\frac{\theta}{r+\theta}\right)^x, \quad x = 0, 1, \ldots$$

where $r > 0$ (assumed known) and $0 < \theta < 1$. We have now that $q(x) = \binom{r+x-1}{x}$, $\varepsilon = \log(1 + r/\theta)$ and $\kappa(\varepsilon) = (\exp(\varepsilon)/(\exp(\varepsilon) - 1))^r$, from which $P(\theta) = \theta$.

Let us now suppose that the prior is an inverted beta distribution with probability density function

$$\pi(\theta) = \frac{r^\beta}{B(\alpha, \beta)} \frac{\theta^{\alpha-1}}{(r+\theta)^{\alpha+\beta}}, \quad \theta > 0,\ \alpha > 0,\ \beta > 0$$

and let us also assume that $g(x) = x^\gamma (r+x)^c$, $c > \gamma > 0$, $x > 0$. Here $\phi(\varepsilon) = r/(\exp(\varepsilon)-1)$ and $h(\varepsilon) = -\gamma - \frac{1}{r}(\gamma + c)\phi(\varepsilon)$. From the prior distribution we have that $x_0 = \alpha - 1$ and $n_0 = \frac{\beta+1}{r}$. Finally using (10), we obtain the collective premium

$$\wp = \frac{r(\alpha + \gamma - 1)}{\beta - c - \gamma + 1}, \quad \alpha > \gamma - 1. \tag{14}$$

Additionaly, by replacing in (14) $\alpha$ and $\beta$ by the updated parameters, we have the Bayesian premium,

$$\Pi(\underline{X}) = \frac{r(\alpha + N\bar{X} + \gamma - 1)}{\beta + N\, r + c - \gamma + 1}. \tag{15}$$

The gamma–inverted gamma case

Let us suppose that the claim size follows a gamma distribution with probability density function

$$f(x|\theta) = \left(\frac{\sigma}{\theta}\right)^\sigma \frac{1}{\Gamma(\sigma)} x^{\sigma-1} \exp\left(-\frac{\sigma\, x}{\theta}\right), \quad x > 0,$$

where $\sigma > 0$ (assumed known) and $\theta > 0$. Then, in this case $q(x) = x^{\sigma-1}/\Gamma(\sigma)$, $\varepsilon = \sigma/\theta$ and $\kappa(\varepsilon) = \varepsilon^{-\sigma}$. Therefore, the risk premium is $P(\theta) = \theta$. Let us suppose that the prior is an inverted gamma distribution with probability density function

$$\pi(\theta) \propto \theta^{-(\alpha+1)} \exp[-1/(\beta\, \theta)],$$

and let us also assume that $g(x) = x^{-\gamma} \exp(-c/x)$, $\gamma > 0$, $c > 0$, $x > 0$. Then, the upper bound for the loss function is given by $(\frac{c}{\gamma})^{-\gamma} e^{-\gamma}$. Now, we have that $\phi(\varepsilon) = \sigma/\varepsilon$, from which it is easy to see that $h(\varepsilon) = -c/\sigma + \gamma/\sigma\, \phi(\varepsilon)$. From the prior distribution we have that $x_0 = 1/(\beta\, \sigma)$ and $n_0 = (\alpha + 1)/\sigma$, therefore the collective premium is

$$\wp = \frac{\beta\, c + 1}{\beta(\alpha + \gamma + 1)}. \tag{16}$$

From (16) and updating the parameters we obtain the Bayesian premium

$$\Pi(\underline{X}) = \frac{\beta(\sigma N\bar{X} + c) + 1}{\beta(\alpha + N\,\sigma + \gamma + 1)}. \tag{17}$$

### The normal–normal case

Finally, let us also suppose that the claim size follows a normal distribution $\mathcal{N}(\theta, \sigma^2)$, $\sigma > 0$ (assumed known) and $\theta \in \mathbb{R}$. We have now that $q(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp(-\frac{x^2}{2\sigma^2})$, the natural parameter is $\varepsilon = -\theta/\sigma^2$ and $\kappa(\varepsilon) = \exp(\frac{\varepsilon^2\sigma^2}{2})$. Thus, the risk premium is given by $P(\theta) = \theta$.

Let us now suppose that the prior is a normal distribution $\mathcal{N}(\alpha, \alpha^2)$, $\alpha \in \mathbb{R}$, and let us also assume that $g(x) = x^\gamma \exp(-x)$, $\gamma > 0$. After some simple computation, the collective and the Bayes premiums are provided by

$$\wp = \sqrt{\gamma}\,\alpha,$$
$$\Pi(\underline{X}) = \frac{\sqrt{\gamma}(N\,\alpha^2\bar{X} + \alpha\,\sigma^2)}{N\,\alpha^2 + \sigma^2}, \tag{18}$$

respectively. Observe also that $h'(x) = -\gamma/x^2 < 0$.

It can be noted that the Bayes premiums given in (11), (13), (15), (17) and (18) are credibility formula, i.e. in all cases considered the experience rated premium adopts the attractive form

$$\Pi(\underline{X}) = Z(N)t(\bar{X}) + (1 - Z(N))\wp, \tag{19}$$

where $Z(N) = N/(\lambda + N + c)$, $Z(N) = n\,N/(\alpha + \beta + Nn + c + \gamma - 2)$, $Z(N) = r\,N/(\beta + Nr - c - \gamma + 1)$, $Z(N) = N\beta\sigma/(\beta(\alpha + N\sigma + \gamma + 1))$ and $Z(N) = N\alpha^2/(N\alpha^2 + \sigma^2)$ for (11), (13), (15), (17) and (18), respectively and $t(x) = x$ in all the cases, except for (18), where $t(x) = \sqrt{\gamma}\,x$. Therefore, (11), (13), (15) and (17) are exact credibility expressions.

Moreover, in the cases given above, under appropriate values of $\gamma$ and $c$ the bound function $g(x)$ es equal to 1. Likewise, expressions (11), (13), (15), (17) and (18) reduce to the MAP estimate of the risk premium.

Finally, it is possible to extend the result obtained for the pair Poisson–gamma under the Esscher premium principle, i.e. the Esscher loss, $L(x, a) = e^{s\,x}(x - a)^2$, $s > 0$, is used to obtain the risk premium, and the loss given in (3) to calculate the collective and Bayes premiums, as shown in the following example.

Example 4.1 Again, under the same assumptions as those in the previous cases, let us also suppose that the practitioner chooses the Esscher loss to compute the risk premium and the loss (3) to obtain the a priori and the experience rated premiums. Then the collective and Bayes premiums are given by

$$\wp = \frac{e^s(\gamma + \sigma - 1)}{\lambda + c},$$
$$\Pi(\underline{X}) = \frac{e^s(\gamma + \sigma + N\bar{X} - 1)}{\lambda + N + c}, \tag{20}$$

respectively. This is easy to observe by following the same argument as in the Poisson–gamma case and also by having into account that in this case the risk premium is $P(\theta) = E(Xe^{s\,X})/E(e^{s\,X}) = \theta\,e^s$.

Besides, note also that (20) can be rewritten as in (19) with $Z(N) = N/(\lambda + N + c)$ and $g(\bar{X}) = e^s\,\bar{X}$.

Remark 4.2 Values of $c$ and $\gamma$ can be appropriately chosen by the practitioner to ensure that the loss incurred by him does not take a value larger than a fixed value.

Appropriate values of the parameters $\gamma$ and $c$ in the $g(x)$ function can be selected by setting equal the upper bound of the loss function, say $M$. For example, we could fix one of the value or assume $\gamma = c$ and numerical computation can be used to obtain the value of the parameter.

### Non–conjugate distributions

Although the examples provided above include conjugate family of distributions, the methodology proposed here can be extended in other circumstances in order to give more flexibility to the practitioner when choosing the prior distribution. As an illustration, a specific example is given. Suppose that $X$ follows a Poisson distribution with parameter $\alpha\theta$, where $\alpha > 0$ and $\theta > \theta_m$ and that the prior distribution is the Pareto distribution with probability density function given by $\pi(\theta) \propto \theta^{-(\beta+1)},\ \theta > \theta_m$. Since the Pareto distribution is non–conjugate with the Poisson distribution we cannot apply the result in Theorem 3.2. Nevertheless, by using the same methodology, we could calculate the collective premium by computing $\arg\max_\theta g(\theta)\pi(\theta)$. Assuming without loss of generality that $\alpha = 1$ and that $g(x) = x^\gamma e^{-cx},\ \gamma > 0, c > 0$, it is a simple exercise to see that in this case $\wp = (\gamma - \beta - 1)/c,\ \gamma > \beta + 1$. Now, the posterior distribution is proportional to $\pi(\theta|\underline{X}) \propto \theta^{N\bar{X}-\beta-1}e^{-N\theta}$ and similarly as above, the Bayes premium, which is given by $\Pi(\underline{X}) = Z_N\bar{X} + (1 - Z_N)\wp$ where the credibility factor is $Z_N = N/(c + N)$, is obtained.

## 5. Numerical experiment

Two numerical applications will be considered in this section. One related to discrete claims and other with claim size, i.e. continuous random variable. Let us first consider a Poisson model where the Poisson parameter $\theta$ represents a driver's propensity to make a claim and the prior indicates how that propensity is distributed throughout the population of insured drivers. This pattern has been used successfully to model the number of vehicle motor accidents in actuarial literature.

The data set corresponds to the number of claims generated by Belgium (1975–76) automobile drivers (Denuit, 1997). It is well–known that the Poisson–gamma model produces as marginal distribution the negative binomial distribution in the following way,

$$f(x) = \frac{\lambda^\sigma}{x!\Gamma(\sigma)}\int_0^\infty e^{-\theta(\lambda+1)}\theta^{\sigma+x-1}\,d\theta = \binom{\sigma + x - 1}{x}\left(\frac{\lambda}{\lambda+1}\right)^\sigma\left(\frac{1}{\lambda+1}\right)^x.$$

Maximum likelihood estimation method has been used to obtain the estimates of the parameters of the model. Those numerical values are $\widehat{\lambda} = 16.138\,(1.506)$ and $\widehat{\sigma} = 1.631\,(0.151)$, where the standard errors are given in brackets. The value of the maximum log–likelihood function is $-36104.1$.

In the following, expression (11) is used to compute the premiums obtained under the squared–error loss function ($\gamma = 1,\ c \to 0$), the maximum a posteriori (MAP) premium ($\gamma \to 0,\ c \to 0$) and the premium under the general 0–1 loss introduced in this paper. Different values for $\gamma$ and $c$ in the neighborhood of the previous ones together with different values of $N$ and $k = N\bar{X}$, the number of claims. The obtained results are displayed in Table 1. As it can be observed, the MAP (0–1 loss function) premium is always lower than the premium based on the squared–error loss function. In addition to this, the former loss function is lower than the general 0–1 loss function except for larger values of $k$ and values of $\gamma$ and $c$ close to zero. This is in accord with the fact for those parameter values $g(\theta)$ decays quickly to zero as $k$ increases. Besides, the premiums calculated under the general 0–1 loss function can be either lower or larger than the ones calculated under the squared error loss function depending on the values of the parameters. In this sense, for values close to zero they are lower, but as $\gamma$ and $c$ increase (see the most right–hand column of Table 1) they are larger.

Expression (17) is now used to compute the premiums for the exponential–inverted gamma distributions. This can be obtained by assuming that $\sigma = 1$. Then, by having into account the following dataset (in millions),

2, 2, 2, 2, 2, 2, 2, 2, 3, 2, 2, 2, 3, 3,
3, 3, 4, 4, 4, 5, 5, 5, 5, 6, 6, 6, 6, 8,
8, 9, 15, 17, 22, 23, 24, 24, 25, 27, 32, 43;

which correspond to 40 wind–related catastrophes in 1977 appearing, for instance, in Hogg and Klugman (1984, p.64) and where only values greater than or equal to \$2,000,000 have been included. It is a simple exercise to see that the marginal distribution is now given by

$$f(x) = \frac{\alpha}{\beta^\alpha}\left(\frac{1}{x+1/\beta}\right)^{\alpha+1}, \quad x > 0,\, \alpha > 0,\, \beta > 0, \tag{21}$$

which is a Pareto type distribution (Lomax distribution).

Note that although this set of data is assumed to arise from a continuous distribution, as Brazauskas and Serfling (2003) and Rizzo (2009) have pointed out, observations can be discretized by grouping. Then, by using the method proposed by these authors, they are de–grouped as follows

$$x_j = \left(1 - \frac{j}{m+1}\right) a + \frac{j}{m+1}\, b, \quad j = 1, 2, \ldots, m,$$

where $(a, b)$ contains $m$ groups of sample observations.

Now using (21), the parameters of this probability density function have been by maximum likelihood estimation method. The resulting estimates are $\widehat{\beta} = 0.018829\,(0.030)$ and $\widehat{\alpha} = 6.72654\,(9.573)$ and the maximum of the log–likelihood function evaluated at those estimates is $-128.598$. Different values of the premiums obtained for the Exponential–Inverted Gamma case are shown n Table 2. Similar comments as in the previous case apply here.

In conclusion, the methodology proposed suggests that these parameter values ($c$ and $\gamma$) can be viewed as a mechanism to assess the influence of the prior on the Bayes premium or the influence of the loss function on the same quantity. This procedure, that has received a lot of attention in the past decades, is called sensitivity or robustness analysis. See Berger (1985) and Micheas and Dey (2004), among others. Therefore, the actuary might choose the value of $\gamma$ and $c$ to calculate a premium according to his preferences. A similar procedure can be developed for other models.

Table 1. Bayes premiums obtained for different loss functions and selected values of $N$, $k$, $c$ and $\gamma$. Poisson–Gamma case.

| $N$ | SEL $\gamma = 1$ $c \to 0$ | MAP $\gamma \to 0$ $c \to 0$ | General 0-1 loss $\gamma = 0.2$ $c = 0.1$ | $\gamma = 0.1$ $c = 0.2$ | $\gamma = 2$ $c = 1$ |
|---|---|---|---|---|---|
| | | | $k = 0$ | | |
| 1 | 0.095166 | 0.036817 | 0.048206 | 0.042160 | 0.145051 |
| 2 | 0.089919 | 0.034788 | 0.045563 | 0.039861 | 0.137472 |
| 3 | 0.085221 | 0.032970 | 0.043194 | 0.037800 | 0.130646 |
| 4 | 0.080989 | 0.031333 | 0.041060 | 0.035941 | 0.124465 |
| 5 | 0.077158 | 0.029850 | 0.039127 | 0.034257 | 0.118843 |
| | | | $k = 2$ | | |
| 1 | 0.211863 | 0.153515 | 0.164226 | 0.157512 | 0.255315 |
| 2 | 0.200183 | 0.145051 | 0.155222 | 0.148922 | 0.241974 |
| 3 | 0.189723 | 0.137472 | 0.147154 | 0.141222 | 0.229959 |
| 4 | 0.180302 | 0.130646 | 0.139883 | 0.134278 | 0.219080 |
| 5 | 0.171773 | 0.124465 | 0.133296 | 0.127985 | 0.209184 |
| | | | $k = 4$ | | |
| 1 | 0.328560 | 0.270212 | 0.280246 | 0.272863 | 0.365578 |
| 2 | 0.310446 | 0.255315 | 0.264881 | 0.257983 | 0.346476 |
| 3 | 0.294225 | 0.241974 | 0.251112 | 0.244643 | 0.329271 |
| 4 | 0.279615 | 0.229959 | 0.238705 | 0.232614 | 0.313695 |
| 5 | 0.266387 | 0.219080 | 0.227465 | 0.221713 | 0.299525 |
| | | | $k = 10$ | | |
| 1 | 0.678651 | 0.620303 | 0.628307 | 0.618915 | 0.696368 |
| 2 | 0.641236 | 0.586105 | 0.593857 | 0.585166 | 0.659982 |
| 3 | 0.607731 | 0.555480 | 0.562989 | 0.554906 | 0.627210 |
| 4 | 0.577553 | 0.527897 | 0.535171 | 0.527623 | 0.597538 |
| 5 | 0.550231 | 0.502924 | 0.509973 | 0.502896 | 0.570547 |

Table 2. Bayes premiums obtained for different loss functions and selected values of $N$, $k$, $c$ and $\gamma$. Exponential–Inverted Gamma case.

| $N$ | SEL $\gamma = 1$ $c \to 0$ | MAP $\gamma \to 0$ $c \to 0$ | General 0-1 loss $\gamma = 0.2$ $c = 0.1$ | $\gamma = 0.1$ $c = 0.2$ | $\gamma = 2$ $c = 1$ |
|---|---|---|---|---|---|
| | | | $k = 0$ | | |
| 1 | 5.46027 | 6.08598 | 5.96082 | 6.03969 | 5.04445 |
| 2 | 4.95123 | 5.46027 | 5.36033 | 5.42506 | 4.61428 |
| 3 | 4.52900 | 4.95123 | 4.86975 | 4.92397 | 4.25171 |
| 4 | 4.17313 | 4.52900 | 4.46144 | 4.50762 | 3.94196 |
| 5 | 3.86911 | 4.17313 | 4.11630 | 4.15619 | 3.67429 |
| | | | $k = 2$ | | |
| 1 | 5.66589 | 6.31516 | 6.18487 | 6.26627 | 5.23091 |
| 2 | 5.13768 | 5.66589 | 5.56181 | 5.62859 | 4.78483 |
| 3 | 4.69956 | 5.13768 | 5.05279 | 5.10870 | 4.40886 |
| 4 | 4.33028 | 4.69956 | 4.62913 | 4.67673 | 4.08767 |
| 5 | 4.01482 | 4.33028 | 4.27102 | 4.31212 | 3.8101 |
| | | | $k = 20$ | | |
| 1 | 7.51650 | 8.37784 | 8.20133 | 8.30558 | 6.90899 |
| 2 | 6.81576 | 7.51650 | 7.37513 | 7.46036 | 6.31981 |
| 3 | 6.23454 | 6.81576 | 6.70016 | 6.77128 | 5.82323 |
| 4 | 5.74465 | 6.23454 | 6.13837 | 6.19873 | 5.39900 |
| 5 | 5.32614 | 5.74465 | 5.66351 | 5.71546 | 5.03238 |
| | | | $k = 40$ | | |
| 1 | 9.57273 | 10.6697 | 10.4418 | 10.5715 | 8.77352 |
| 2 | 8.68029 | 9.57273 | 9.38993 | 9.49566 | 8.02534 |
| 3 | 7.94007 | 8.68029 | 8.53056 | 8.61859 | 7.39475 |
| 4 | 7.31617 | 7.94007 | 7.81530 | 7.88984 | 6.85603 |
| 5 | 6.78318 | 7.31617 | 7.21071 | 7.27472 | 6.39047 |

## 6. Concluding remarks

In this paper a new premium calculation principle has been proposed for the Bayesian estimation by considering non–standard and bounded loss functions. As a result of this, new Bayesian premiums that can be expressed as a credibility formula, have been obtained. These expressions have been found to be useful in actuarial practice when experience rating, via Bayesian analysis, is used to compute premiums written as a compromise between the past claims of the policyholder within the portfolio, and the past claims of this portfolio as a whole. The methodology proposed is simple and the credibility formulae are straightforwardly achieved. Additionally, they include as particular case premiums previously obtained under the squared–error loss function and also those ones obtained by using the maximum a posteriori (MAP) estimate.

Finally, after observing the numerical results calculated in the previous section, it can be pointed out that the practitioner has the option to choose a Bayes premium either lower or larger than the one obtained under the square error–loss function.

## Acknowledgements

Authors partially funded by grant ECO2013-47092 (Ministerio de Economía y Competitividad, Spain). The authors are indebted to an anonymous referee for helpful comments and suggestions which, without doubt, helped to improve an earlier version of the paper.

## References

Berger, J., 1985. Statistical Decision Theory and Bayesian Analysis. Second Edition. Springer. New York.

Brazauskas, V., Serfling, R., 2003. Favorable estimator for fitting Pareto models: A study using goodness-of-fit measures with actual data. Astin Bulletin, 33(2), 365–381.

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

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

Denuit, M., 1997. A new distribution of Poisson–type for the number of claims. Astin Bulletin, 27(2), 229–242.

Diaconis, P., Ylvisaker, D., 1979. Conjugate priors for exponential families. Annals of Statistics, 7, 269–281.

Eichenauer, J., Lehn, J., Rettig, S., 1988. A gamma–minimax result in credibility theory. Insurance: Mathematics and Economics, 7(1), 49–57.

Furman, E., Zitikis, R., 2008. Weighted premium calculation principles. Insurance: Mathematics and Economics, 42, 459–465.

Gerber, H.U., Arbor, A., 1980. Credibility for Esscher premiums. Mitteilungen der Vereinigung schweiz. Versicher ungsmathematiker, Heft 3, 307–312.

Gómez–Déniz, E., Pérez, J.M., Vázquez, F.J., 2006. On the use of posterior regret $\Gamma-$minimax actions to obtain credibility premiums. Insurance: Mathematics and Economics, 39, 115–121.

Gómez–Déniz, E., 2008. A generalization of the credibility theory obtained by using the weighted balanced loss function. Insurance: Mathematics and Economics, 42, 850–854.

Hastie, T., Tibshirani, R., Friedman, J., 2001. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, New York.

Heilmann, W., 1989. Decision theoretic foundations of credibility theory. Insurance: Mathematics and Economics, 8(1), 77–95.

Hogg, R. V., Klugman. S. A., 1984. Loss Distributions. John Wiley & Sons, Inc.

Hooge, L., Goovaerts, M.M., 1975. Bayesian inference in credibility theory. Astin Bulletin, 8(2), 164–174.

Jewell, W., 1974. Credible means are exact Bayesian for exponential families. Astin Bulletin, 8(1), 77–90.

Kadane, J., Chuang, D.T., 1978. Stable decision problems. Annals of Statistics, 6, 1095–1110.

Kamińska, A. and Porosiński, Z., 2009. On robust Bayesian estimation under some asymmetric and bounded loss function. Statistics, 43(3), 253–265.

Landsman, Z.M., Makov, U.E., 1998. Exponential dispersion models and credibility. Scandinavian Actuarial Journal, 1, 89–96.

Landsman, Z.M., Makov. U.E., 1999. Credibility evaluation for the exponential dispersion family. Insurance: Mathematics and Economics, 24, 23–29.

MacEachern, S.N., 1993. A characterization of some conjugate prior distributions for exponential families. Scandinavian Journal of Statistics, 20(1), 77–82.

Micheas, A.C., Dey, D.K., 2004. On measuring loss robustness using maximum a posteriori estimate. Communications in Statistics–Theory and Methods, 33(5), 1069–1085.

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

Payandeh, A.T., Hatamib, H., Najafabadi, M.O., 2012. A maximum–entropy approach to the linear credibility formula. Insurance: Mathematics and Economics, 51, 216–221.

Rizzo, M.L., 2009. New goodness-of-fit tests for Pareto distributions. Astin Bulletin, 39(2), 691–715.

Robert, C.P., 2007. The Bayesian Choice. From Decision–Theoretic Foundations to Computational Implementation. Second Edition. Springer, New York.

Smith, J.Q., 1980. Bayes estimates under bounded loss. Biometrika, 67(3), 629–638.

Vaseghi, S.V., 2000. Advanced Digital Signal Processing and Noise Reduction, Second Edition. John Wiley & Sons, New York.

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

Young, V.R., 1997. Credibility using semiparametric models. Astin Bulletin, 27, 273–285.

Young, V.R., 2000. Credibility using semiparametric models and a loss function with a constancy penalty. Insurance: Mathematics and Economics, 26, 151–156.