Abstract

The article deals with robust bayesian estimates. Different statistical models are proposed which yield estimates of the posterior mean which are able to cope with very large claims, misspecified prior information and misspecified distributional assumptions. The methods developed in the paper are especially useful when dealing with catastrophe risks.

Keywords


1 Introduction

The main advantage of Bayesian statistical methods is to allow the use of a priori information in a coherent way. This is important in an insurance context where in addition to the data there is usually a larger quantity of collateral or of a priori information available. It is essential in a reinsurance context where large claims play an important role and where by definition the amount of data is scarce. In such a context Bayesian techniques are especially powerful (see e.g. R. Schnieper (1993)).

Bayesian statistical models are sometimes criticized for their alleged lack of robustness. In order to keep the models tractable, practitioners often use distributions which have natural conjugate priors and are thus closed under sampling: the posterior density is of the same type as the prior density. The information conveyed by the data only leads to an update of the parameters of the prior density. (Examples are found for instance in J. Aitchison and I.R. Dunsmore (1975).) When it comes to computing the posterior mean, the most commonly used such likelihoods have indeed the property that they do not weigh down 'large' outlying observations thus producing estimates which are not robust. Credibility estimators can be viewed as another attempt to keep the estimation of the posterior mean mathematically tractable. They are least squares estimates of the posterior mean which are linear in the observations. As such they cannot weigh down 'large' outlying observations.

This lack of robustness however is not inherent to Bayesian statistics but is the consequence of the choice of the statistical model i.e. of the likelihood and of the prior distribution. We propose statistical models
which can cope with large outlying observations as well as with misspecification of the prior mean. These models lead to posterior means which have to be computed numerically. This increased mathematical complexity is compensated by the fact the models provide an automatic means of detecting and accommodating aberrant observations as well as misspecified prior information.

Many of the distributions we propose are mixtures of normal distributions. As such they lead to posterior means which are weighted averages of the observations with the weights depending on the observations.

We start by giving a simple presentation of the general bayesian framework and of credibility theory. We then propose two models which provide both an unbiased estimator of an individual pure risk premium and a robust treatment of large claims. Finally we propose a family of distributions, which is a generalization of the Box Tiao distributions and which allows a robust approach to the observations, the prior mean and the distributional assumptions.

2 General Framework

2.1 Bayesian Model

Let $X_1, X_2, \ldots, X_n$ be a random sample from a common distribution with an unknown - possibly vector valued - parameter $\theta$. Given $\theta$, the $X_i$ are independent with common density $f(x \mid \theta)$, which is referred to as the likelihood of $X_i$. In an insurance context the $X_i$’s usually denote claims. As is customary, densities are indexed by their arguments. Thus $f(x)$ and $f(\theta)$ may be different densities. A generalization of the following results to the case where the probability distributions are not differentiable is straightforward and is omitted here. The possible values of $\theta$ are given by a density $f(\theta)$, the prior density. Bayes’ theorem states that given a realization of $X_1, \ldots, X_n$, the posterior density of $\theta$ is

$$f(\theta \mid x) = \frac{f(x \mid \theta)f(\theta)}{f(x)}$$

where

$$f(x \mid \theta) = \prod_{i=1}^{n} f(x_i \mid \theta)$$

and

$$f(x) = \int f(x \mid \theta)f(\theta)d\theta$$

does not depend on $\theta$.

Thus

$$f(\theta \mid x) = kf(x \mid \theta)f(\theta)$$

where $k$ is some constant independent of $\theta$. 
Often one is interested in some function of $\theta$, e.g. in the pure risk premium

$$m(\theta) = E(\bar{X} | \theta) = \int x f(x | \theta) dx.$$ 

A posteriori, i.e. after one has observed a realization of $n$ claims $X = (X_1, \ldots, X_n)$ the 'best' estimate of $m(\theta)$ is

$$m(\bar{x}) = E(m(\theta) | \bar{x}) = \int m(\theta) f(\theta | \bar{x}) dx$$

where 'best' means that $m(\bar{x})$ is the estimator of $m(\theta)$ with the minimum mean square error.

**Example**

Let $\tilde{X}_1, \ldots, \tilde{X}_n$ be normally distributed with unknown mean $\theta$ and known precision $p$ (the precision is the multiplicative inverse of the variance)

$$p(\bar{x} | \theta) = \prod_{i=1}^{n} p(x_i | \theta) = \left( \frac{p}{2\pi} \right)^{\frac{n}{2}} e^{-\frac{p}{2} \sum_{i=1}^{n} (x_i - \theta)^2}.$$ 

The prior distribution of $\tilde{\theta}$, the unknown sample mean, is itself normal with known mean $m$ and known precision $q$

$$p(\theta) = \sqrt{\frac{q}{2\pi}} e^{-\frac{q}{2} (\theta - m)^2}$$

According to Bayes’ theorem, the posterior density is

$$p(\theta | \bar{x}) = k \cdot p(\bar{x} | \theta)p(\theta)$$

$$= k \cdot e^{-\frac{1}{2}(p \sum_{i=1}^{n} (x_i - \theta)^2 + q(\theta - m)^2)}$$

$$= k \cdot e^{-\frac{np\bar{x} + qm}{np + q}}.$$ 

And it is seen that the posterior distribution of $\tilde{\theta}$ is again a normal distribution, i.e. the normal likelihood with a normal prior is closed under sampling.

The posterior mean is

$$E(\tilde{\theta} | X) = \frac{np \cdot \bar{x} + qm}{np + q}$$

The posterior precision is

$$Var^{-1}(\tilde{\theta} | X) = n \cdot p + q$$
i.e. it is the sum of the precisions of the observations and of the prior density.

The mean square error of the posterior mean is

$$E(E(\tilde{\theta} \mid X) - \tilde{\theta})^2 = E\left[ E(\tilde{\theta} - E(\tilde{\theta} \mid X))^2 \mid X \right]$$

$$= E(Var(\tilde{\theta} \mid X))$$

$$= \frac{1}{np + q}$$

Whereas the mean square error of the classical estimator is

$$E(\bar{X} - \hat{\theta})^2 = E(\hat{\theta}Var(\bar{X} \mid \theta)) = \frac{1}{np}$$

Hence

$$E(E(\tilde{\theta} \mid X) - \tilde{\theta})^2 = E(\bar{X} - \hat{\theta})^2 \cdot \frac{n}{n + \frac{q}{p}}.$$  

If $n$ is small compared to the ratio of precisions $q/p$, the Bayesian estimator is much more precise than the classical estimator.

### 2.2 Credibility Theory

Let $\tilde{X}_1, \ldots, \tilde{X}_n$ be a random sample characterized by an unknown risk parameter $\theta$. Given $\theta$, $\tilde{X}_1, \ldots, \tilde{X}_n$ are independent with finite second moments

$$E(\tilde{X}_i \mid \theta) = m(\theta)$$

$$Var(\tilde{X}_i \mid \theta) = v(\theta).$$

The pure risk premium $m(\theta)$ is to be approximated by a premium which is linear in the observations.

$$\overline{m(\theta)} = \alpha_0 + \sum_{j=1}^{n} \alpha_jX_j$$

and which minimizes the mean square error

$$E(\overline{m(\theta)} - m(\theta))^2 = \min_{\alpha_0, \alpha_1, \ldots, \alpha_n}.$$  

It is easily seen that the optimal linear premium is a weighted average of the individual mean and of the a priori mean

$$\overline{m(\theta)} = z\bar{X} + (1 - z)m$$

where

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$
\[ m = E(m(\theta)) \]

The weight \( z \) given to the individual mean \( \bar{X} \) is called the credibility factor. It is equal to

\[
z = \frac{n \cdot \text{Var}(m(\theta))}{n \cdot \text{Var}(m(\theta)) + E(\text{Var}(\bar{X} | \theta))}
\]

In the notation of the preceding example we have

\[ \text{Var}(m(\theta)) = q^{-1} \quad E(\text{Var}(\bar{X} | \theta)) = p^{-1} \]

and it is seen that in the case of a normal likelihood with known variance and of a normal prior the credibility formula is equal to the a posterior mean. This result generalizes to simple exponential families where the mean is the sufficient statistic and where the prior is the natural conjugate prior (see W.S. Jewell, 1974).

The mean square error of the credibility formula is

\[ E((m(\theta) - m(\theta))^2 = \text{Var}(E(\bar{X} | \theta)) \cdot (1 - z). \]

This is seen in the following way

\[
E((m(\theta) - m(\theta))^2 = (1 - z)^2 E(m - m(\theta))^2 + z^2 E(\bar{X} - m(\theta))^2 + 2(1 - z)z E((m - m(\theta))(\bar{X} - m(\theta))].
\]

The last term of the right hand side of the above equation is 0 which is seen by taking the conditional expectation given \( \theta \). Thus

\[ E((m(\theta) - m(\theta))^2 = (1 - z)^2 q^{-1} + z^2 \frac{p^{-1}}{n} = q^{-1} \frac{p^{-1}}{nq^{-1} + p^{-1}} = \text{Var}(E(\bar{X} | \theta)) \cdot (1 - z) \]

which proves the above statement.

The mean square error of the classical estimator is

\[ E((\bar{X} - m(\theta))^2 = \frac{p^{-1}}{n} = \frac{1}{n} E_\theta(\text{Var}(\bar{X} | \theta)) \]

Hence

\[ E((m(\theta) - m(\theta))^2 = z \cdot E((\bar{X} - m(\theta))^2 \]

which is seen in the following way

\[ E((m(\theta) - m(\theta))^2 = q^{-1} \frac{p^{-1}}{nq^{-1} + p^{-1}} = \frac{p^{-1}}{n} \frac{nq^{-1}}{nq^{-1} + p^{-1}} = E((\bar{X} - m(\theta))^2 \cdot z. \]

From the above it is seen that for \( n \) small compared to the ratio of within and between variances \( E(\text{Var}(\bar{X} | \theta)) / \text{Var}(m(\theta)) \), the Bayesian estimator is much more precise than the classical estimator.

The general Bayesian framework and the credibility theory presented here can be generalized to different cohorts of risks, to observations with heterogeneous weights, to general linear models, etc.
3 Treatment of Large Claims

We propose different Bayesian statistical models which allow for an automatic and coherent treatment of outliers - or in an insurance context of large claims. The estimators derived from these models are bias-free. This is a very important feature in an insurance context since large claims must be paid and cannot be simply truncated or discarded when it comes to pricing a risk.

3.1 t-distributed likelihood and normal prior

Assumptions

1. Given $\theta$, the observations are independent and t-distributed with location parameter $\theta$ and $c$ degrees of freedom

$$p(x \mid \theta) = \frac{\Gamma\left(\frac{c+1}{2}\right)}{(c\pi)^{\frac{1}{2}}\Gamma\left(\frac{c}{2}\right)} \cdot \left(1 + \frac{(x - \theta)^2}{c}\right)^{-\frac{c+1}{2}}$$

We assume $c > 1$ so that $m(\theta) = E(\bar{X} \mid \theta)$ exists and is equal to $\theta$.

2. $\theta$ is normally distributed with mean $m$ and precision $p$

$$p(\theta) = \left(\frac{p}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{p}{2}(\theta - m)^2}$$

Given $n$ claims $X = (X_1, \ldots, X_n)$ the best estimator of $m(\theta)$ (i.e. the estimator which minimizes the mean square error) is the posterior mean

$$E(m(\theta) \mid \bar{x}) = E(\bar{X} \mid \bar{x}) = \int \theta p(\theta \mid \bar{x}) d\theta$$

where the posterior density $p(\theta \mid \bar{x})$ is given by Bayes’ theorem

$$p(\theta \mid \bar{x}) = \frac{p(\bar{x} \mid \theta)p(\theta)}{p(\bar{x})}$$

In the following example we illustrate the behaviors of $E(m(\theta) \mid \bar{x})$ and compare it to the behavior of the credibility estimator $t(\bar{x})$.

Example

We have only one observation and we assume $p = Var^{-1}(m(\bar{X})) = \frac{1}{3}$ and $c = 2, 3$ corresponding to $Var^{-1}(\bar{X} \mid \theta) = 0$ and $\frac{1}{3}$ respectively. We also assume $m = 0$, which amounts to a choice of the origin. We have

$$t(\bar{X}) = \frac{n \cdot E^{-1}(Var(\bar{X} \mid \theta)) \cdot \bar{X} + Var^{-1}(m(\theta)) \cdot m}{n \cdot E^{-1}(Var(\bar{X} \mid \theta)) + Var^{-1}(m(\theta))}$$
and $E(m(\tilde{\theta}) \mid X)$ is computed according to the formula given above. We obtain

<table>
<thead>
<tr>
<th>$c = 2$</th>
<th>$x$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E(m(\tilde{\theta}) \mid x)$</td>
<td>0.7</td>
<td>1.3</td>
<td>1.8</td>
<td>2.1</td>
<td>2.2</td>
<td>2.0</td>
<td>1.7</td>
<td>1.4</td>
<td>1.2</td>
<td>1.0</td>
<td>0.6</td>
<td>0.5</td>
<td>0.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t(x)$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$c = 3$</th>
<th>$x$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E(m(\tilde{\theta}) \mid x)$</td>
<td>0.7</td>
<td>1.3</td>
<td>1.9</td>
<td>2.4</td>
<td>2.6</td>
<td>2.6</td>
<td>2.3</td>
<td>1.9</td>
<td>1.6</td>
<td>1.4</td>
<td>0.8</td>
<td>0.6</td>
<td>0.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t(x)$</td>
<td>0</td>
<td>0.5</td>
<td>1</td>
<td>1.5</td>
<td>2</td>
<td>2.5</td>
<td>3</td>
<td>3.5</td>
<td>4</td>
<td>4.5</td>
<td>5</td>
<td>7.5</td>
<td>10</td>
<td>15</td>
<td></td>
</tr>
</tbody>
</table>

In both cases $c = 2$ and $c = 3$ the posterior mean discards very large claims. Nevertheless the estimator is biasfree since $E(E(m(\tilde{\theta}) \mid X)) = E(m(\tilde{\theta})) = \theta$. A further very important feature of the posterior mean is that it gives 'small' claims a higher weight than does the credibility estimator. The credibility estimator produces unnatural results both in the case $c = 2$ and $c = 3$. In the first case the claim is ignored altogether and in the second case the claim gets a weight of 0.5 irrespectively of how far apart it is from the prior mean. The second case can be adjusted through e.g. truncation of the claim. In such a case an ad hoc adjustment must be made to the credibility formula in order to keep it biasfree. (See for instance Gisler, 1980). There is not much that can be done to fix the credibility formula in the first case.

The behaviour of the posterior mean with respect to large claims observed in the above example holds generally true within the framework of the above model.

**Theorem 1**

Under assumptions (1) and (2) above, we have

(i) $\lim_{x \to \pm \infty} E \left[ m(\tilde{\theta}) \mid x \right] = E \left[ m(\tilde{\theta}) \right]$

(ii) $\lim_{x_{k+1}, \ldots, x_{k+l} \to \pm \infty} E \left[ m(\tilde{\theta}) \mid x_1, \ldots, x_k, x_{k+1}, \ldots, x_{k+l} \right]$

$$= E \left[ m(\tilde{\theta}) \mid x_1, \ldots, x_k \right]$$

if $k \geq l$

i.e. $l$ 'large' claims are all rejected if there are at least $2l$ claims.

The proof of the theorem is found in O'Hagan (1979).
The reason why large claims are weighed down by the posterior mean in the context of the above model is best understood by analyzing the t-distribution as a mixture of normal distributions with unknown, $\chi^2$-distributed precisions.

It is easily seen that assumption 1 above is equivalent with 1'.

1'(i) Given $\theta$ and $\rho$, the observations are independent and normally distributed with mean $\theta$ and precision $\rho$.

$$p(x \mid \theta, \rho) = \left(\frac{\rho}{2\pi}\right)^\frac{1}{2} e^{-\frac{\rho}{2}(x-\theta)^2}$$

1'(ii) $\theta$ and $\rho$ are independent with

$$p(\rho) = k \cdot \rho^{\frac{k}{2} - 1} e^{-\frac{k}{2}\rho}.$$

The proof of the equivalence is straightforward and is omitted here. The posterior mean can now be rewritten in the following way

$$E\left(m(\bar{\theta}) \mid \bar{x}\right) = \int E\left(m(\bar{\theta}) \mid \bar{\rho}, \bar{x}\right) p(\bar{\rho} \mid \bar{x}) \, d\rho_1...d\rho_n$$

It is easily seen that

$$E\left(m(\bar{\theta}) \mid \bar{\rho}, \bar{x}\right) = \frac{p \cdot m + \rho_1 x_1 + ... + \rho_n x_n}{p + \rho_1 + ... + \rho_n}$$

Hence

$$E\left(m(\bar{\theta}) \mid \bar{x}\right) = \sum_{i=0}^{n} E\left(\frac{\rho_i}{\rho_0 + ... + \rho_n} \mid \bar{x}\right) \cdot x_i$$

with $\rho_0 = p$ and $x_0 = m$, and it is seen that the posterior mean can be represented as a weighted average of the observations and of the prior mean with the weights depending on the observations.

The posterior density of the precisions is obtained in the following way

$$p(\bar{\rho} \mid \bar{x}) = k \cdot p(\bar{x} \mid \bar{\rho}) p(\bar{\rho})$$

where $p(\bar{x} \mid \bar{\rho})$ is given by the following theorem.
Theorem 2

Assumptions

1. The observations $\widetilde{X}_1, \widetilde{X}_2, ..., \widetilde{X}_n$ are normally distributed with the following, unknown parameters: $E(\widetilde{X}_i) = \theta, Var(\widetilde{X}_i) = \rho_i^{-1}$.

2. The common, unknown mean $\widetilde{\theta}$ is a normally distributed random variable with $E(\widetilde{\theta}) = x_0$ and $Var(\widetilde{X}) = \rho_0^{-1}$. $\rho_0$ and $x_0$ are given.

3. The random variables $\widetilde{X}_1, ..., \widetilde{X}_n$ are independent given $\rho_1, ..., \rho_n$ and $\theta$.

4. The random vector $(\rho_1, \widetilde{\rho}_2, ..., \widetilde{\rho}_n)$ and $\widetilde{\theta}$ are independent.

The following relation holds true

$$p(x_1, ..., x_n \mid \rho_1, ..., \rho_n) = k \cdot \left( \prod_{i=0}^{n} \rho_i \right)^{\frac{1}{2}} \cdot \frac{1}{\sqrt{\sum_{i=0}^{n} \rho_i}} \cdot e^{-\frac{1}{2} \sum_{i=0}^{n} \rho_i (x_i - \mu)^2}$$

Proof

$$p(x \mid \rho) = \int p(x \mid \rho, \theta)p(\theta \mid \rho) d\theta = \int p(x \mid \rho, \theta)p(\theta) d\theta$$

where we have used assumption 4.

$$p(x \mid \rho) = k \int_{-\infty}^{\infty} \prod_{i=1}^{n} \rho_i^{\frac{1}{2}} e^{-\frac{1}{2} (x_i - \theta)^2} \rho_0^{\frac{1}{2}} e^{-\frac{1}{2} \theta^2} d\theta$$

$$= k \cdot \left( \prod_{i=0}^{n} \rho_i \right)^{\frac{1}{2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2} \sum_{i=0}^{n} \rho_i (x_i - \theta)^2} d\theta$$

And the expression in square brackets above is

$$\sum_{i=0}^{n} \rho_i (x_i - \theta)^2 = \sum_{i=0}^{n} \rho_i (x_i^2 - 2\theta x_i + \theta^2)$$

$$= \sum_{i=0}^{n} \rho_i x_i^2 - 2\theta \sum_{i=0}^{n} \rho_i x_i + \theta^2 \sum_{i=0}^{n} \rho_i$$

$$= \sum_{i=0}^{n} \rho_i x_i^2 + \sum_{i=0}^{n} \rho_i \left( \theta^2 - 2\theta \frac{\sum_{i=0}^{n} \rho_i x_i}{\sum \rho_i} + \left( \frac{\sum_{i=0}^{n} \rho_i x_i}{\sum \rho_i} \right)^2 \right) - \frac{\left( \sum_{i=0}^{n} \rho_i x_i \right)^2}{\sum \rho_i}$$

$$= \sum_{i=0}^{n} \rho_i x_i^2 - \left( \frac{\sum_{i=0}^{n} \rho_i x_i}{\sum \rho_i} \right)^2 + \sum_{i=0}^{n} \rho_i \left( \theta - \frac{\sum_{i=0}^{n} \rho_i x_i}{\sum \rho_i} \right)^2$$

hence
\[ p(x \mid \rho) = k \cdot \left( \prod_{i=0}^{n} \rho_i \right)^{\frac{1}{2}} \cdot e^{-\frac{1}{2} \left( \sum_{i=0}^{n} \rho_i x_i^2 - \left( \sum_{i=0}^{n} \frac{\rho_i}{\rho_i} \right)^2 \right)} \]

and the value of the above integral is \( k \cdot \left( \sum_{i=0}^{n} \rho_i \right)^{-\frac{1}{2}} \cdot e^{-\frac{1}{2} \left( \sum_{i=0}^{n} \rho_i x_i^2 - \left( \sum_{i=0}^{n} \frac{\rho_i}{\rho_i} \right)^2 \right)} \). Hence

\[ p(x \mid \rho) = k \cdot \left( \prod_{i=0}^{n} \rho_i \right)^{\frac{1}{2}} \cdot e^{-\frac{1}{2} \left( \sum_{i=0}^{n} \rho_i x_i^2 - \left( \sum_{i=0}^{n} \frac{\rho_i}{\rho_i} \right)^2 \right)} \]

The expression in square brackets above is

\[ \sum_{i=0}^{n} \rho_i x_i^2 - \left( \sum_{i=0}^{n} \frac{\rho_i}{\rho_i} \right)^2 = \left( \sum_{i=0}^{n} \rho_i \right)^{-1} \left( \sum_{i,j} \rho_i \rho_j x_i x_j - \sum_{i,j} \rho_i \rho_j x_i x_j \right) \]

\[ = \left( \sum_{i=0}^{n} \rho_i \right)^{-1} \sum_{i,j} \rho_i \rho_j (x_i - x_j)^2 \]

Hence

\[ p(x \mid \rho) = k \cdot \left( \prod_{i=0}^{n} \rho_i \right)^{\frac{1}{2}} \cdot e^{-\frac{1}{2} \left( \sum_{i,j} \rho_i \rho_j (x_i - x_j)^2 \right)} \]

which proves the theorem. \( \blacksquare \)

We can now compute the observation dependent weights and the posterior mean. As an illustration we recompute the above example for \( c=3 \).

<table>
<thead>
<tr>
<th>x</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>w(x)</td>
<td>0.69</td>
<td>0.68</td>
<td>0.67</td>
<td>0.64</td>
<td>0.59</td>
<td>0.52</td>
<td>0.43</td>
<td>0.33</td>
<td>0.24</td>
<td>0.18</td>
<td>0.14</td>
<td>0.06</td>
<td>0.03</td>
<td>0.01</td>
</tr>
<tr>
<td>( E(\tilde{\theta} \mid x) )</td>
<td>0</td>
<td>0.7</td>
<td>1.3</td>
<td>1.9</td>
<td>2.4</td>
<td>2.6</td>
<td>2.6</td>
<td>2.3</td>
<td>1.9</td>
<td>1.6</td>
<td>1.4</td>
<td>0.8</td>
<td>0.6</td>
<td>0.4</td>
</tr>
</tbody>
</table>

where \( w(x) = E(\frac{\tilde{\rho}}{\tilde{\rho}+\tilde{\rho}} \mid x) \).

From the above representation it is seen that within the framework of the model the observations carry information not only on the mean.
but also on the precision. Hence, the weight given to an individual observation depends on this observation. In the above example, the weight varies between 0.69 and 0, whereas in the case of a normal likelihood, or in the case of the credibility formula, the weight given to an individual observation is constant and does not depend on the value of the observation. Other examples of models leading to observation-dependent credibility weights are found in Jewell, Schnieper (1983). In particular it should be noticed that the model can be specified by defining the a priori density of the precision \( p(\rho) \), irrespectively of whether the likelihood

\[
p(x | \theta) = \int_0^p p(x | \theta, \rho)p(\rho | \theta)d\rho
\]

is available in analytical form or not.

The use of a likelihood which is a scale mixture of normal densities, in order to deal with outliers, is an idea which goes back to De Finetti (1961). In the following section we discuss another example of a scale mixture of normal densities, the Laplace density. It is seen that in that case large outlying observations are truncated rather than discarded.

### 3.2 Laplace distributed likelihood and normal prior

**Assumptions**

1. Given \( \theta \), the claims are independent and Laplace distributed with mean \( \theta \) and precision \( q \)

\[
p(x | \theta) = \sqrt{\frac{q}{2}} e^{-\sqrt{2q}|x-\theta|}
\]

2. \( \theta \) is normally distributed with mean \( m \) and precision \( p \)

\[
p(\theta) = \sqrt{\frac{p}{2\pi}} e^{-\frac{1}{2}(\theta-m)^2}
\]

The computation of the posterior mean is straightforward. Using the same parameters as in the preceding example \( (m = 0, p = q = \frac{1}{3}) \) we obtain

<table>
<thead>
<tr>
<th>( x )</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>( E(\tilde{m}(\theta)</td>
<td>x) )</td>
<td>0.0</td>
<td>0.6</td>
<td>1.2</td>
<td>1.7</td>
<td>2.1</td>
<td>2.3</td>
<td>2.4</td>
<td>2.4</td>
<td>2.4</td>
<td>2.4</td>
<td>2.4</td>
</tr>
</tbody>
</table>

and it is seen that large claims are truncated rather than discarded as was the case for a \( t \)-distributed likelihood.

A graph of \( E(m(\theta) | x) \) for the above example and for the preceding example \( -t \) likelihood with 3 degrees of freedom- is shown in appendix 1.

The general behavior of \( E(m(\theta) | x) \) is described by the following
Theorem

Under the assumption (1) and (2) above we have

(i) \( \lim_{x \to \pm \infty} E \left[ \tilde{\theta} \mid x \right] = E \left[ \tilde{\theta} \right] \pm \sqrt{2q} \frac{p}{\pi} \)

(ii) \( E \left[ \tilde{\theta} \mid x \right] \) is an increasing function of \( x \)

Proof

\[
E \left[ m(\tilde{\theta}) \mid x \right] = \frac{\int_{-\infty}^{\infty} \theta e^{-\sqrt{2q|x-\theta|} - \frac{\theta}{2}(\theta-m)^2} d\theta}{\int_{-\infty}^{\infty} e^{-\sqrt{2q|x-\theta|} - \frac{\theta}{2}(\theta-m)^2} d\theta}
\]

without loss of generality we set \( m=0 \) which amounts to a choice of the origin and obtain after some straightforward simplifications

\[
E \left[ m(\tilde{\theta}) \mid x \right] = \frac{\int_{-\infty}^{\infty} \theta e^{\sqrt{2q\theta} - \frac{x}{2}\theta^2} d\theta + e^{2\sqrt{2q}x} \int_{-\infty}^{\infty} \theta e^{-\sqrt{2q\theta} - \frac{x}{2}\theta^2} d\theta}{\int_{-\infty}^{\infty} e^{\sqrt{2q\theta} - \frac{x}{2}\theta^2} d\theta + e^{2\sqrt{2q}x} \int_{-\infty}^{\infty} e^{-\sqrt{2q\theta} - \frac{x}{2}\theta^2} d\theta}
\]

on the other hand

\[
e^{2\sqrt{2q}x} \int_{-\infty}^{\infty} \theta e^{-\sqrt{2q\theta} - \frac{x}{2}\theta^2} d\theta \leq e^{2\sqrt{2q}x} \int_{-\infty}^{\infty} \theta e^{-\frac{x}{2}\theta^2} d\theta = e^{2\sqrt{2q}x} \frac{1}{p} e^{-\frac{x}{2}x^2}
\]

hence the above term goes to 0 for \( x \to \infty \)

and

\[
e^{2\sqrt{2q}x} \int_{-\infty}^{\infty} e^{-\sqrt{2q\theta} - \frac{x}{2}\theta^2} d\theta \leq e^{2\sqrt{2q}x} \int_{-\infty}^{\infty} e^{-\frac{x}{2}\theta^2} d\theta
\]

for \( x \geq 1 \), and it follows that this term also goes to 0 for \( x \to \infty \).

Hence

\[
\lim_{x \to \infty} E \left[ m(\tilde{\theta}) \mid x \right] = \lim_{x \to \infty} \frac{\int_{-\infty}^{\infty} \theta e^{\sqrt{2q\theta} - \frac{x}{2}\theta^2} d\theta}{\int_{-\infty}^{\infty} e^{\sqrt{2q\theta} - \frac{x}{2}\theta^2} d\theta} = \frac{\int_{-\infty}^{\infty} \theta e^{-\frac{p}{2}(2\sqrt{2q}x)} d\theta}{\int_{-\infty}^{\infty} e^{-\frac{p}{2}(2\sqrt{2q}x)} d\theta} = \frac{\sqrt{2q}}{p}
\]

And a similar exercise for \( x \to -\infty \) proves statement (i) of the theorem.

We now turn to the proof of statement (ii).
\[
E \left[ m(\bar{\theta}) \mid x \right] = \frac{\int_{-\infty}^{x} \theta e^{-\sqrt{2q}(x-\theta)} - \frac{p_2}{2} \theta^2 d\theta + \int_{x}^{\infty} \theta e^{-\sqrt{2q}(\theta-x)} - \frac{p_2}{2} \theta^2 d\theta}{\int_{-\infty}^{x} e^{-\sqrt{2q}(x-\theta)} - \frac{p_2}{2} \theta^2 d\theta + \int_{x}^{\infty} e^{-\sqrt{2q}(\theta-x)} - \frac{p_2}{2} \theta^2 d\theta}
\]

\[
\frac{d}{dx} E \left[ m(\bar{\theta}) \mid x \right] = \frac{u'v - uv'}{v^2}
\]

with

\[
u' = xe^{-px^2} - \sqrt{2q} \int_{-\infty}^{x} \theta e^{-\sqrt{2q}(x-\theta)} - \frac{p_2}{2} \theta^2 d\theta
\]

\[
u' = xe^{-px^2} - \sqrt{2q} \int_{-\infty}^{x} \theta e^{-\sqrt{2q}(x-\theta)} - \frac{p_2}{2} \theta^2 d\theta
\]

\[
u' = xe^{-px^2} - \sqrt{2q} \int_{-\infty}^{x} \theta e^{-\sqrt{2q}(x-\theta)} - \frac{p_2}{2} \theta^2 d\theta
\]

Hence

\[
u'v - uv' = \sqrt{2q} ((-A_1 + A_2) (B_1 + B_2) - (A_1 + A_2) (-B_1 + B_2))
\]

\[
u'v - uv' = 2\sqrt{2q} (-A_1 B_2 + A_2 B_1)
\]

\[
A_2 B_1 - A_1 B_2 = \int_{-\infty}^{\infty} \theta e^{-\sqrt{2q}(\theta-x)} - \frac{p_2}{2} \theta^2 d\theta \int_{x}^{\infty} e^{-\sqrt{2q}(\theta-x)} - \frac{p_2}{2} \theta^2 d\theta
\]

We introduce the following random variable
\[ X_- \sim N \left( \mu = -\frac{\sqrt{2q}}{p}, \sigma^2 = p^{-1} \right) \]
\[ X_+ \sim N \left( \mu = \frac{\sqrt{2q}}{p}, \sigma^2 = p^{-1} \right) \]

We obtain
\[ A_2 B_1 - A_1 B_2 = e^{\frac{q}{2}} \left( E[X_-; X_+ > x] \cdot \Pr[X_+ < x] - E[X_+; X_- < x] \cdot \Pr[X_- > x] \right) \]
\[ = e^{\frac{q}{2}} \Pr[X_+ < x] \cdot \Pr[X_- > x] \cdot (E[X_-; X_- > x] - E[X_+; X_+ < x]) \]
\[ > 0 \]
And it is seen that \( \frac{d}{dx} E\left[ m(\theta) \mid x \right] > 0 \)
which proves statement (ii) of the theorem.

**Remark**

From the proof of the theorem it is seen at once that in the case of \( n \) claims, the following relation holds true
\[ \lim_{x_1, \ldots, x_n \to -\infty} E(m(\tilde{\theta}) \mid x_1, \ldots, x_n) = m + n \frac{\sqrt{2q}}{p} \]

The Laplace distribution too can be analysed as a scale mixture of normal distributions with unknown, exponential distributed variance.

To be more precise, assumption 1 above is equivalent with

1'(i) Given \( \theta \) and \( \rho \), the claims are independent and normally distributed with mean \( \theta \) and precision \( \rho \).

\[ p(x \mid \theta, \rho) = \left( \frac{\rho}{2\pi} \right)^{\frac{1}{2}} e^{-\frac{\rho}{2}(x-\theta)^2} \]

1'(ii) \( \theta \) and \( \rho \) are independent with

\[ p(\rho) = q \cdot \rho^{-2} e^{-q\rho^{-1}}. \]

Assumption 1'(ii) means that \( \bar{\rho}^{-1} \), i.e. the variance of the claims, is exponentially distributed with expectation \( q^{-1} \).

The equivalence between (1) and (1') is seen in the following way

\[ p(x \mid \theta) = \frac{q}{\sqrt{2\pi}} \int_0^\infty \rho^{-\frac{3}{2}} e^{-\frac{1}{2}((x-\theta)\rho + 2q\rho^{-1})} d\rho \]

making the change of variable \( \rho = \tau^2 \), we obtain
where we have used the following result

\[
\int_0^\infty t^{-2}e^{-\frac{1}{2}(a^2t^2+b^2t^{-2})}dt = \sqrt{\frac{\pi}{2}} |ab| e^{-|ab|}
\]

Using theorem 2 of section 3.1, we can now compute the posterior mean as a weighted average of the prior mean and of the observation with the weight depending on the observation

\[
E(m(\tilde{\theta}) \mid x) = (1 - w(x)) m + w(x) \cdot x
\]

with

\[
w(x) = E\left(\frac{\tilde{\rho}}{\rho + \tilde{\rho}} \mid x\right)
\]

\(w(x)\) is computed as in section 3.1 except that the prior density of \(\tilde{\rho}\), \(p(\rho)\), is now as in 1'(ii) above. We recompute the example above for \(m = 0\), \(p = q = \frac{1}{2}\) and we obtain

<table>
<thead>
<tr>
<th>x</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>w(x)</td>
<td>0.63</td>
<td>0.62</td>
<td>0.60</td>
<td>0.56</td>
<td>0.51</td>
<td>0.46</td>
<td>0.40</td>
<td>0.35</td>
<td>0.31</td>
<td>0.27</td>
<td>0.24</td>
<td>0.16</td>
<td>0.12</td>
<td>0.08</td>
</tr>
<tr>
<td>E(m(\tilde{\theta}) \mid x)</td>
<td>0.0</td>
<td>0.6</td>
<td>1.2</td>
<td>1.7</td>
<td>2.1</td>
<td>2.3</td>
<td>2.4</td>
<td>2.4</td>
<td>2.4</td>
<td>2.4</td>
<td>2.4</td>
<td>2.4</td>
<td>2.4</td>
<td>2.4</td>
</tr>
</tbody>
</table>

The fact that, in the case of the Laplace distribution, large, outlying observations are truncated rather than discarded as was the case for the t-distribution, makes the Laplace distribution more appealing for insurance applications. It would indeed be difficult to justify a pure risk premium which is a decreasing function of the claim.

Appendix 2 and 3 show the contour plot of the posterior mean as a function of two claims in the case of a t- and of a Laplace likelihood respectively. In the case of a t-likelihood two 'large' claims lead to a higher pure risk premium whereas one 'large' claim only is discarded. Two 'excessively large' claims are discarded. In the case of a Laplace likelihood one or two 'large' claims lead to a high pure risk premium. One or more 'excessively large' claims however get truncated.
3.3 Scale Mixtures of Normal Distributions


**Theorem**

If $X$ has a density function $f$ symmetric about 0 then there exist independent random variables $V, Z$ with $Z$ standard normal such that

$$X = \frac{Z}{V}$$

if and only if the derivatives of $f$ satisfy

$$\left( - \frac{d}{dy} \right)^k f \left( y^{\frac{1}{2}} \right) \geq 0 \text{ for } y > 0$$

The proof is found in the above mentioned article.

In addition to the t- and the Laplace distribution, the logistic distribution with density

$$p(x) = \frac{e^{-x}}{(1 + e^{-x})^2}$$

is also a scale mixture of normal distributions.

4 Robust Priors

Sometimes there is uncertainty about the prior information. The best way to handle such a situation is to assume a prior density with a fat tail. If the information conveyed by the data contradicts the prior information, the latter will be discarded. This statement is illustrated by the following two models.

4.1 t-distributed prior and likelihood

**Assumptions**

1. Given $\theta$, the observations are independent and t-distributed with location parameter $\theta$ and $c$ degrees of freedom

$$p(x \mid \theta) \propto \left( 1 + \frac{(x - c)^2}{c} \right)^{-\frac{c+1}{2}}$$

2. $\tilde{\theta}$ is t-distributed with mean $m$, $d$ degrees of freedom and scale parameter $k$

$$p(\theta) \propto \left( 1 + \frac{(m - \theta)^2}{k^2 \cdot d} \right)^{-\frac{d+1}{2}}$$
The scale parameter $k$ is chosen in such a way to keep $Var(\tilde{\theta})$ constant as the degrees of freedom vary.

Appendix 4 shows the posterior mean $E(m(\tilde{\theta}) \mid x)$ as a function of one observation $x$. The likelihood is $t$-distributed with $c = 3$. The three different curves correspond to $d = 3$, $d = 6$ and $d = \infty$ (normal density) respectively. $k$ is varied in order to keep $Var(\tilde{\theta})$ constant and equal to $Var(X \mid \theta) = 3$. In the case $d = 3$, the posterior mean is the arithmetic average of the prior and of the observation. In the case $d = 6$ a large, outlying observation is weighed down, however less so than in the case $d = \infty$.

4.2 Laplace distributed prior and likelihood

Assumptions

1. Given $\theta$, the observations are independent and Laplace distributed with mean $\theta$ and precision $q$

$$p(x \mid \theta) \propto e^{-\sqrt{2q}|x-\theta|}$$

2. $\tilde{\theta}$ is Laplace distributed with mean $m$ and precision $p$

$$p(\theta) \propto e^{-\sqrt{2p}|\theta-m|}$$

Appendix 5 shows the posterior mean $E(m(\tilde{\theta}) \mid x)$ as a function of one observation $x$. The likelihood is Laplace distributed with $q = \frac{1}{3}$. The two different curves correspond to a Laplace prior with $p = \frac{1}{3}$ and to a normal prior with precision $\frac{1}{3}$ respectively. In the case of the Laplace-Laplace model, the posterior mean is the arithmetic average of the prior and of the observation. The Laplace-Normal case is as in section 3.2.

5 Robust Distributional Assumptions

Box and Tiao, 1973 advocate the use of the following family of likelihoods

$$p(x \mid \theta, \sigma, \beta) = w(\beta) \cdot \sigma^{-1} \cdot e^{-c(\beta) \cdot |x-\theta|^{1+\beta}}$$

where

$$c(\beta) = \left( \frac{\Gamma \left( \frac{\beta}{2} (1 + \beta) \right)}{\Gamma \left( \frac{\beta}{2} (1 + \beta) \right)} \right)^{\frac{1}{1+\beta}}$$

$$w(\beta) = \frac{\Gamma \left( \frac{\beta}{2} (1 + \beta) \right)^{\frac{1}{2}}}{(1 + \beta) \cdot \Gamma \left( \frac{\beta}{2} (1 + \beta) \right)^{\frac{3}{2}}}$$

213
\( p(x \mid \theta, \sigma, \beta) \) is a density on \((-\infty, \infty)\). \( \theta \) and \( \sigma \) are the a priori mean and standard deviation respectively. \( \beta \) defines the shape of the density. When \( \beta \) tends to -1 the distribution tends to the rectangular distribution. When \( \beta = 0 \) the distribution is normal, when \( \beta = 1 \) the distribution is the Laplace distribution. Box and Tiao only consider the range \(-1 < \beta < 1\). The density however is defined for \(-1 < \beta < \infty\). The densities corresponding to the range \(1 < \beta < \infty\) have the particularity that the posterior mean discards large, outlying observations. They behave like t-densities. We therefore consider the whole range \(-1 < \beta < \infty\) and refer to the corresponding densities as to the generalised Box Tiao family. Appendix 6 shows the posterior mean as a function of one observation for different values of \( \beta \). Thereby we have chosen \( \sigma = 1 \) and the prior \( p(\theta) \) is \( N(0,1) \) distributed. It is seen that in the case \( \beta = 0 \), the posterior mean is a linear function of the observation, which is due to the fact that this is the normal normal model and that therefore the credibility formula is exact. The case \( \beta = 1 \) corresponds to the Laplace normal model and the posterior mean is a monotone increasing function of the observation which tends to \( \sqrt{2} \) when the observation becomes large. The case \( \beta = -0.5 \) leads to a posterior mean which increases more quickly than a straight line. The case \( \beta = 0.5 \) leads to a behavior of the posterior mean which is intermediary between the normal normal and the Laplace normal model. The cases \( \beta = 2 \) and \( \beta = 3 \) lead to a similar behavior of the posterior mean as in the t-normal model: large, outlying observations are rejected.

The above illustrates the fact that the generalised Box Tiao family is robust in respect of the distributional assumptions: treating \( \beta \) as an unknown parameter within a Bayesian framework leads to a treatment of outliers which is influenced by the actual data rather than imposed by the choice of models.

6 Bühlmann Straub Example Revisited

In order to illustrate the behavior of the generalised Box Tiao family we reanalyse the data of Bühlmann Straub (1970). Bühlmann and Straub consider an example with 7 different treaties producing each 5 burning costs \( X \) in 5 consecutive years. To each treaty year is associated a premium volume \( P \).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Year:</td>
<td>h=1</td>
<td>h=2</td>
<td>h=3</td>
<td>h=4</td>
<td>h=5</td>
<td>h=6</td>
<td>h=7</td>
</tr>
<tr>
<td>g=5</td>
<td>( P )</td>
<td>( X )</td>
<td>( P )</td>
<td>( X )</td>
<td>( P )</td>
<td>( X )</td>
<td>( P )</td>
</tr>
<tr>
<td>g=4</td>
<td>40.0</td>
<td>14.0</td>
<td>11.3</td>
<td>18.0</td>
<td>20.5</td>
<td>4.9</td>
<td>43.9</td>
</tr>
<tr>
<td>g=3</td>
<td>64.0</td>
<td>14.0</td>
<td>25.0</td>
<td>20.1</td>
<td>22.5</td>
<td>9.9</td>
<td>47.1</td>
</tr>
<tr>
<td>g=2</td>
<td>84.2</td>
<td>13.0</td>
<td>18.5</td>
<td>23.7</td>
<td>25.7</td>
<td>6.7</td>
<td>53.0</td>
</tr>
<tr>
<td>g=1</td>
<td>10.0</td>
<td>11.0</td>
<td>14.3</td>
<td>25.3</td>
<td>29.7</td>
<td>10.3</td>
<td>61.1</td>
</tr>
</tbody>
</table>

The total premium pertaining to a given treaty

214
\[ P_h = \sum_{g=1}^{5} P_{gh} \]

and the mean claim rate of that treaty

\[ \overline{X}_h = \sum_{g=1}^{5} \frac{P_{gh}}{P_h} X_{gh} \]

are given in the bottom line for each treaty. As an estimate for the a priori mean Bühlmann Straub use

\[ m = \sum_{h=1}^{7} \frac{P_h}{P} \overline{X}_h = 9.6\% . \]

where

\[ P = \sum_{h=1}^{7} P_h. \]

As an estimator for the within and between variance respectively they use

\[ v = 209.0 \cdot 10^{-4} \]

and

\[ w = 12.1 \cdot 10^{-4} . \]

For the derivation of the estimates see the original article.

We shall adopt an empirical Bayes approach and use the same values for \( m, v \) and \( w \) as in Bühlmann Straub.

The likelihood of a given treaty year is thus

\[ \ell(x_{ij} \mid \theta_j, \beta_j) = w(\beta_j) \cdot \sqrt{\frac{p_{ij}}{209}} e^{-c(\beta_j)(|x_{ij} - \theta_j|\sqrt{\frac{p_{ij}}{209}})^{1+\beta_j}} \]

with \( c(\beta) \) and \( w(\beta) \) as in section 5.

Note that for \( \beta = 1 \), the \( x_{ij} \) are normally distributed given \( \theta \).

We assume that \( \tilde{\theta}_j \) is normally distributed with mean \( m = 9.6\% \) and precision \( p = \frac{1}{12.1} \).

We assume that the a priori density of \( \tilde{\beta}_j \) is proportional to the standard normal density on \([-0.9, 3]\) and is 0 outside this interval.

\( \tilde{\theta}_j, \tilde{\beta}_j \) are independent random variables.
The pairs \( \left( \tilde{\theta}_1, \tilde{\beta}_1 \right), \ldots, \left( \tilde{\theta}_7, \tilde{\beta}_7 \right) \) are independent random vectors. We can now compute the posterior densities

\[
f (\theta_j, \beta_j \mid \underline{x}) = f \left( \theta_j, \beta_j \mid x_{1j}, \ldots, x_{5j} \right)
= k \prod_{i=1}^{5} \ell \left( x_{ij} \mid \theta_j, \beta_j \right) \cdot f (\theta_j) f (\beta_j)
\]

Appendix 7 gives a contour plot of \( f (\theta_3, \beta_3) \). It is easily seen that for \( \beta_j = 0 \), the above general model reduces to the normal normal model for which the credibility estimator is exact. In general however the 'best' estimator of \( \theta_j \) is the posterior mean \( E \left( \tilde{\theta}_j \mid \underline{x} \right) \) which is obtained from the above formula.

\[
E \left( \tilde{\theta}_j \mid \underline{x} \right) = \frac{\int \int \theta_j \left( \prod_{i=1}^{5} \ell \left( x_{ij} \mid \theta_j, \beta_j \right) \right) f (\theta_j) f (\beta_j) \, d\theta_j \, d\beta_j}{\int \int \prod_{i=1}^{5} \ell \left( x_{ij} \mid \theta_j, \beta_j \right) f (\theta_j) f (\beta_j) \, d\theta_j \, d\beta_j}
\]

\( E \left( \tilde{\beta}_j \mid \underline{x} \right) \) is computed in a similar way. A strong departure from 0 indicates that the underlying model is strongly different from the normal model.

The two parameters \( E \left( \tilde{\theta}_j \mid \underline{x} \right) \) and \( E \left( \tilde{\beta}_j \mid \underline{x} \right) \) of the seven treaties are tabulated below. They show that the model is close to the normal model and hence that the credibility estimators are close to the posterior mean.

<table>
<thead>
<tr>
<th>Treaty Number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Premium</td>
<td>9.6%</td>
<td>9.6%</td>
<td>9.6%</td>
<td>9.6%</td>
<td>9.6%</td>
<td>9.6%</td>
<td>9.6%</td>
</tr>
<tr>
<td>Credibility Premium</td>
<td>5.0%</td>
<td>17.3%</td>
<td>5.6%</td>
<td>7.3%</td>
<td>9.5%</td>
<td>11.9%</td>
<td>9.2%</td>
</tr>
<tr>
<td>Posterior Mean</td>
<td>5.0%</td>
<td>16.2%</td>
<td>5.6%</td>
<td>7.2%</td>
<td>9.6%</td>
<td>12.0%</td>
<td>9.2%</td>
</tr>
<tr>
<td>( E \left( \tilde{\beta}_j \mid \underline{x} \right) )</td>
<td>0.1</td>
<td>0.7</td>
<td>0.1</td>
<td>0.6</td>
<td>0.4</td>
<td>0.1</td>
<td>0.3</td>
</tr>
<tr>
<td>Individual Premium</td>
<td>3.1%</td>
<td>19.5%</td>
<td>5.0%</td>
<td>7.0%</td>
<td>9.5%</td>
<td>12.1%</td>
<td>9.2%</td>
</tr>
</tbody>
</table>
Literature


Box, G.E.P. and Tiao, G.C (1973) Bayesian Inference in Statistical Analysis, Addison-Wesley.


Appendix 7

Generalised Box Tiao Model
Joint Contour Plot of mean and shape parameter
Treaty Number 3