SCALE MIXTURES DISTRIBUTIONS IN INSURANCE APPLICATIONS

BY

S.T. BORIS CHOY AND C.M. CHAN

ABSTRACT

In this paper non-normal distributions via scale mixtures are introduced into insurance applications. The symmetric distributions of interest are the Student-t and exponential power (EP) distributions. A Bayesian approach is adopted with the aid of simulation to obtain posterior summaries. We shall show that the computational burden for the Bayesian calculations is alleviated via the scale mixtures representations. Illustrative examples are given.

KEYWORDS

Normal Distribution, Student-t Distribution, Exponential Power Distribution, Scale mixtures of Normal Distributions, Scale Mixtures of Uniform distributions, Markov chain Monte Carlo, Gibbs Sampler, Pure Premium, Credibility, Robustness, Outliers

1. INTRODUCTION

Klugman (1992) addressed the role of Bayesian methods in insurance applications. One of the insurance problems is the credibility analysis which involves the calculation of insurance premiums. The premium charged to an individual insured depends on his/her past claim records and the overall average of all insureds. A Bayesian approach provides an ideal way to strike a balance between the specific and overall claim experiences. Past claims of an individual insured, however, may be widely spread from year to year and the claims from a particular class of insureds can be very different from those of other classes, so much so that the conventional assumption of normality becomes inappropriate. To tackle these problems we consider alternative symmetric distributions for modelling individual and class-specific claims. The proposed models can identify those insureds who have experienced unfavourable claims in previous years, but will not penalize them for the excessive claims. While Klugman considers a number of models, the key here is the modelling of kurtosis, that is the heaviness of tails. Klugman does not consider this aspect of modelling claims.

ASTIN BULLETIN, Vol. 33, No. 1, 2003, pp. 93-104

Bayesian robustness is an important topic in Bayesian analysis but is a largely neglected area in insurance applications. Although normality is assumed in many applications, it is well-known that the normal distribution does not provide a robust analysis, i.e. catering for outliers. A way to achieve robustness is to model the data with heavy-tailed distributions. Landsman and Makov (1998, 1999) consider the exponential dispersion family to provide a more flexible choice of tails for insurance claims while Young (1997, 1998) adopts a nonparametric approach for the prior determination. In this paper, we highlight the use of two classes of scale mixtures distributions to achieve Bayesian robustness in insurance context. For more details about Bayesian robustness, see Berger (1994).

Conventional numerical methods and analytic approximations have become unattractive for computations of sophisticated models within a Bayesian framework. The computer-intensive sampling methods, such as Markov chain Monte Carlo (MCMC) methods (see Smith and Roberts, 1993 and Tierney, 1994), provide an efficient alternative for handling complicated Bayesian calculations. In particular, the Gibbs sampling approach (see Gelfand *et al.*, 1990) enables us to simulate posterior samples from a set of iteratively updated conditional distributions. Choy and Smith (1997) make use of normal scale mixtures representations for the Student-t, symmetric stable and exponential power *EP* densities in Bayesian hierarchical modellings, and Landsman and Makov (1999) consider these distributions for insurance claims . In this paper, we shall use a new scale mixtures representation for the *EP* distribution which simplifies the simulation algorithm. If one or more of the conditional distributions to be sampled is difficult and can not be done directly, then a possible resource is the Metropolis-Hastings algorithm. However, this is potentially slow, requires a certain amount of time consuming tuning and can "get stuck", in that there could be a part of the Markov chain output which is fixed at one point. The algorithm here avoids the need for a Metropolis chain and all conditional distributions to be sampled can be done directly.

The structure of this paper is as follows. Section 2 shows how the t-distribution and exponential-power distribution can be expressed into two different scale mixtures forms. Section 3 illustrates how a full Bayesian model is formulated for a specific insurance problem, details of model implementation using MCMC methods are also presented. Section 4 presents some simulation results. Finally, a concluding remark is given in Section 5.

2. Scale mixtures Representations

2.1. Normal scale mixtures for the Student-*t*

To speed up the efficiency of the MCMC algorithms, the Student-t density with location θ, scale σ^2 and degrees of freedom α is always expressed by the following normal scale mixtures form (Andrews and Mallows, 1974; Choy and Smith, 1997)

$$t_\alpha\left(x \,\middle|\, \theta,\sigma^2\right) = \int_0^\infty N\left(x \,\middle|\, \theta,\frac{\sigma^2}{\lambda}\right) G\left(\lambda \,\middle|\, \frac{\alpha}{2},\frac{\alpha}{2}\right) d\lambda$$

https://doi.org/10.2143/AST.33.1.1041 Published online by Cambridge University Press

where $N(\cdot|\cdot)$ and $G(\cdot|\cdot)$ are normal and gamma densities, respectively, Using this representation, the random variable X, conditional on λ, has a normal $N(\theta, \frac{\sigma^2}{\lambda})$ distribution with λ having a gamma $G\left(\frac{\alpha}{2}, \frac{\alpha}{2}\right)$ mixing distribution, i.e.

$$X \,\Big|\, \theta, \sigma^2, \lambda \sim N\left(\theta, \frac{\sigma^2}{\lambda}\right) \text{ and } \lambda \sim G\left(\frac{\alpha}{2}, \frac{\alpha}{2}\right).$$

2.2. Uniform scale mixtures for the *EP*

Another class of symmetric distributions of interest is the *EP* distribution whose density function is given by

$$EP(x \,|\, \theta, \sigma, \beta) = \frac{c_1}{\sigma} \exp\left\{-\left| c_o^{1/2} \sigma^{-1}(x-\theta)\right|^{2/\beta}\right\}, \quad -\infty < x < \infty,$$

where

$$c_o = \frac{\Gamma(3\beta/2)}{\Gamma(\beta/2)}, \quad c_1 = \frac{c_0^{1/2}}{\beta\Gamma(\beta/2)}$$

and $\beta \in (0, 2]$ is the kurtosis parameter. The *EP* family offers a range of symmetric distributions from the uniform shape ($\beta \to 0$) to the double-exponential shape ($\beta = 2$). Tails can be more platykurtic ($\beta < 1$) or more leptokurtic ($\beta > 1$) than the normal distribution ($\beta = 1$). The *EP* distribution has been studied by Box and Tiao (1992) for statistical modelling and for robustness considerations. Recently, Choy and Smith (1997) adopted a normal scale mixtures form for the *EP* density, but this is restricted to the leptokurtic members and expertise in numerical and simulation techniques is essential for practitioners of this normal scale mixtures representation. See Choy and Smith (1997) for technical details. In this paper, a new and more remarkable scale mixtures form, known as the uniform scale mixtures, is proposed for the *EP* density function which is given by

$$EP(x \,|\, \theta, \sigma, \beta) = \int_0^\infty U\left(x \,\middle|\, \theta - \frac{\sigma}{\sqrt{2c_o}} u^{\beta/2},\, \theta + \frac{\sigma}{\sqrt{2c_o}} u^{\beta/2}\right) G\left(u \,\middle|\, 1 + \frac{\beta}{2}, 2^{-1/\beta}\right) du$$

where $U(\cdot \,|\, a, b)$ is the uniform density function with support (a, b). This mixtures form implies that

$$X \,|\, \theta, \sigma, \beta, u \sim U\left(\theta - \frac{\sigma}{\sqrt{2c_o}} u^{\beta/2}, \theta + \frac{\sigma}{\sqrt{2c_o}} u^{\beta/2}\right)$$

and

$$u \sim G\left(1 + \frac{\beta}{2}, 2^{-1/\beta}\right).$$

This mixtures form was proposed by Walker and Gutierrez-Pena (1999) and is valid for the entire range of β values (from 0 to 2). It also simplifies the

https://doi.org/10.2143/AST.33.1.1041 Published online by Cambridge University Press

computational procedures for Bayesian calculations, which will be illustrated in the subsequent parts of this article. See also Choy (1999) for a comparative study on Bayesian computation of the uniform scale mixtures form with normal scale mixtures form of the *EP* density in a simple Bayeisan hierarchical analysis.

3. Modelling Insurance Premium

An insurance problem of interest comes under the name of credibility. This problem is related to the setting of insurance premiums to insureds according to their past claim records and the overall average of the insureds. A simple model for this problem has been studied by Bühlman and Straub (1972) and Meyers (1984). Other models for insurance claims can be found in Klugman *et al.* (1998). Let Y_{ij} be the claim size of the jth policyholder in category/class i. The claim size is defined as the ratio of the actual loss to the total exposure. A natural model for the claim is a three-stage hierarchical model given by

$$y_{ij} \Big| \theta_i, \sigma^2 \sim N\left(\theta_i, \frac{\sigma^2}{p_{ij}}\right) \qquad i = 1, \ldots, k, \;\; j = 1, \ldots, n_i$$

$$\theta_i \Big| \mu, \tau^2 \sim N\left(\mu, \tau^2\right)$$

$$\mu, \sigma^2, \tau^2 \sim N\left(\mu_o, v^2\right) IG(a_1, b_1)\, IG(a_2, b_2)$$

where p_{ij} are known values proportional to the exposures that produce the observations. $IG(a, b)$ is the inverse gamma distribution with arguments $a > 0$, $b > 0$ and mean $b/(a-1)$. Conditional independence within and between groups are assumed in the first two stages. In the third stage of the hierarchy, we assign a normal prior distribution for μ and inverse gamma prior distributions for σ^2 and τ^2 and assume that these prior distributions are independent and the hyperparameters are known. In many situations, however, the assumption of normality for data and the prior specification may not be appropriate. In fact, data may come from a flat-tailed distribution. In addition, this model with normal component is non-robust in the sense that inference is highly sensitive to outliers appearing in the first two stages. Therefore, we shall modify the distributions in the first two stages.

Here we consider a Student-t sampling distribution with known degrees of freedom α and an *EP* prior distribution with kurtosis parameter β which can either be fixed or assigned a prior density $p(\beta)$. By using the scale mixtures representation of the Student-t and *EP* distributions, the modified model is given by

$$y_{ij} \Big| \theta, \sigma^2, \lambda_{ij} \sim N\left(\theta_i, \frac{\sigma^2}{p_{ij}\lambda_{ij}}\right) \qquad i = 1, \ldots, n, \;\; j = 1, \ldots, k$$

$$\theta_i \Big| \mu_i, \tau^2, u_i \sim U\left(\mu - \frac{\tau}{\sqrt{2c_o}} u_i^{\beta/2}, \; \mu + \frac{\tau}{\sqrt{2c_o}} u_i^{\beta/2}\right)$$

$$\mu, \sigma^2, \tau^2, \lambda_{ij}, u_i \sim N\left(\mu_o, v^2\right) IG(a_1, b_1)\, IG(a_2, b_2)\, G\left(\frac{\alpha}{2}, \frac{\alpha}{2}\right) G\left(1 + \frac{\beta}{2}, 2^{-1/\beta}\right)$$

where λ_{ij}'s and u_i's are the mixing parameters of the Student-t and EP distributions, respectively.

To implement this modified model, both analytic and numerical methods are inefficient, but the simulation-based Gibbs sampling approach (see Roberts and Smith, 1993) provides a good alternative to perform the necessary Bayesian calculations. When the kurtosis parameter β is known, the system of full conditional distributions for the Gibbs sampler is given by

$$[\theta_i \,|\, \text{rest}] = N\left(\theta_i \left| \frac{\sum_{j=1}^{n_i} p_{ij}\lambda_{ij}y_{ij}}{\sum_{j=1}^{n_i} p_{ij}\lambda_{ij}}, \frac{\sigma^2}{\sum_{j=1}^{n_i} p_{ij}\lambda_{ij}} \right.\right) I\left(\frac{\mu - \tau u_i^{\beta/2}}{\sqrt{2c_o}} < \theta_i < \frac{\mu + \tau u_i^{\beta/2}}{\sqrt{2c_o}}\right)$$

$$\left[\sigma^2 \,|\, \text{rest}\right] = IG\left(a_1 + \frac{1}{2}\sum_{i=1}^{k} n_i,\; b_1 + \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{n_i} p_{ij}\lambda_{ij}\left(y_{ij} - \theta_i\right)^2\right)$$

$$[\mu \,|\, \text{rest}] = N\left(\mu_o, v^2\right) I\left(\max\left(\theta_i - \frac{\tau u_i^{\beta/2}}{\sqrt{2c_o}}\right) < \mu < \min\left(\theta_i + \frac{\tau u_i^{\beta/2}}{\sqrt{2c_o}}\right)\right)$$

$$\left[\tau^2 \,|\, \text{rest}\right] = IG\left(a_1 + \frac{k}{2}, b_2\right) I\left(\tau^2 > \max\left(2c_o u_i^{-\beta}(\theta_i - \mu)^2\right)\right)$$

$$\left[\lambda_{ij} \,|\, \text{rest}\right] = G\left(\frac{\alpha + 1}{2}, \frac{1}{2}\left(\alpha + \frac{p_{ij}\left(y_{ij} - \theta_i\right)^2}{\sigma^2}\right)\right)$$

$$[u_i \,|\, \text{rest}] = Exp\left(2^{-\beta}\right) I\left(u_i^{\beta} > 2c_o\tau^{-2}(\theta_i - \mu)^2\right)$$

where $[\cdot|\cdot]$ denotes a conditional density function, 'rest' denotes the observed data and the set of all parameters excluding the one of interest, the index i extends from 1 to k whereas j ranges from 1 to n_i for each i. We note that all these conditional distributions are of standard forms in which case we have no difficulty in generating random variates from them. In addition, if a vague prior is assumed for the population parameter μ, then the corresponding full conditional distribution is reduced to a uniform distribution, i.e.

$$[\mu|\text{rest}] = U\left(\max\left(\theta_i - \frac{\tau u_i^{\beta/2}}{\sqrt{2c_o}}\right), \min\left(\theta_i + \frac{\tau u_i^{\beta/2}}{\sqrt{2c_o}}\right)\right).$$

4. Illustrative Example

4.1. The data

Klugman (1992, Data Set 3, p. 197-207) contains data of total workers' compensation and exposure of 124 different occupation classes over a seven-year period without any inflation adjustment. The loss ratio y_{ij} of class i in year j is defined by

$$y_{ij} = 1000 \times \frac{m_{ij}}{p_{ij}}$$

where m_{ij} (in 10^7) and p_{ij} (in 10^7) are the total compensation to workers and the exposure of class i in year j, respectively. That is, the observation is the pure premium charged per \$1000 of compensation for a particular class in a specific year. For easy computation, both m_{ij} and p_{ij} are divided by 10^7 to a reasonable scale. Note that in the model specification, the variance of the pure premium depends on exposures.

In the data set, three occupation classes contain some missing values and are discarded for analysis. Klugman (1992) analyses the first six-year data of the 121 classes using a normal-likelihood-normal-prior model (normal-normal model) in which Bayesian calculations are done through numerical methods. Thus, this high reliance on numerical methods makes the extension for normal family to other families of symmetric distributions impossible in analysis. In this example, we reconsider all seven-year data of the 121 occupation classes using the Student-t sampling distribution and the EP prior distribution. The Klugman's normal-normal model is in fact a special case of our model. Our proposed simulation-based approach will be more computationally efficient than the numerical methods.

4.2. Parameter estimation

In order to obtain Bayes estimates and other posterior quantities of the parameters of interest, we use Gibbs sampling scheme of the Markov chain Monte Carlo methods. The Gibbs sampler was run for a single series of 12000 iteration in which the first 2000 iterations were discarded as the 'burn-in' period. We then pick up values at every 10th value to mimic a random sample of size 1000 from the joint posterior distributions. Convergence of the simulated Markov chains can be assessed by plotting the ergodic average for each parameter. Here the ergodic average of a parameter is the running averages of the outputs from the Markov chain. In the simulations, a vague prior is assigned to ν, i.e. $\mu \to \infty$, and non-informative priors are assumed for σ^2 and τ^2, i.e. $a_1 = b_1 = a_2 = b_2 = 0$.

Tables 1, 2 and 3 exhibit the posterior means and standard deviations of μ, σ and τ for different degrees of freedom α and kurtosis parameter β. For a fixed β, the standard deviation of the sampling distribution, σ, increases significantly when α increases. This means that the heavier the tails of the sampling distribution, the smaller the estimate of σ will be. In other words, the outliers within each occupation class can be captured by the heavy-tailed distribution so that the error variance is reduced. If we inspect the data set in detail, we may find that some occupation classes have made unusually large claims in one or two years during the seven-year period which can be considered possible outliers. When a normal sampling distribution is assumed, the pure premium charged to these occupation classes will be substantially increased. An insurance company adopting this normal-normal model in its premium calculation may be less competitive than those who adopt the flat-tailed alternatives. In fact, the use of the heavy-tailed distribution will automatically downweigh or even "ignore" the possible outliers within each class and the overall analysis of the class parameter θ_i is protected from the distorting effect of the outliers. Comparatively, the estimates of μ and τ are less volatile when α increases because the second stage parameters are less remarkedly affected by the first stage distribution.

For the choice of prior distribution for the pure premium θ_i, the exponential power distribution provides platykurtic to leptykurtic shapes when β varies from 0 to 2. When α is fixed, the inference of μ is protected even when the tails of the *EP* distribution become heavy. Estimates of τ seem to be rather insensitive to both changes in β and α. Finally, the effect of different choices of α and β on the variance ratio τ^2/σ^2 is clearly illustrated in Table 4.

TABLE 1

POSTERIOR MEAN (WITH STANDARD ERROR IN PARENTHESES) OF μ FOR DIFFERENT COMBINATIONS (α, β).

α \ β	**0.25**	**0.50**	**0.75**	**1.00**	**1.25**	**1.50**	**1.75**	**2.00**
1	16.99	15.90	15.16	14.58	13.99	13.67	13.04	12.89
	(0.92)	(0.87)	(0.95)	(0.98)	(0.96)	(1.02)	(0.99)	(0.91)
3	17.58	16.99	16.15	15.66	15.23	14.76	14.60	14.18
	(0.91)	(0.93)	(0.97)	(1.01)	(0.97)	(0.93)	(1.09)	(0.95)
5	18.16	17.43	16.66	15.90	15.39	15.06	14.96	14.63
	(0.96)	(0.99)	(1.02)	(0.97)	(0.97)	(0.95)	(1.05)	(0.90)
10	18.71	17.86	17.02	16.38	15.85	15.65	15.37	14.96
	(0.91)	(1.02)	(1.02)	(1.00)	(0.96)	(0.97)	(0.90)	(0.98)
∞	19.51	18.75	17.77	17.03	16.79	16.36	15.93	15.64
	(0.96)	(1.05)	(1.12)	(0.96)	(1.08)	(1.11)	(1.04)	(0.99)

TABLE 2

POSTERIOR MEAN (WITH STANDARD ERROR IN PARENTHESES) OF σ FOR DIFFERENT COMBINATIONS (α, β).

α \ β	**0.25**	**0.50**	**0.75**	**1.00**	**1.25**	**1.50**	**1.75**	**2.00**
1	5.35	5.32	5.29	5.27	5.26	5.21	5.22	5.20
	(0.29)	(0.29)	(0.29)	(0.29)	(0.29)	(0.28)	(0.28)	(0.29)
3	9.31	9.31	9.29	9.26	9.23	9.21	9.19	9.18
	(0.36)	(0.37)	(0.36)	(0.37)	(0.37)	(0.36)	(0.35)	(0.37)
5	10.97	10.92	10.91	10.88	10.88	10.88	10.85	10.84
	(0.38)	(0.38)	(0.39)	(0.39)	(0.39)	(0.37)	(0.39)	(0.38)
10	12.67	12.65	12.65	12.66	12.63	12.61	12.61	12.61
	(0.39)	(0.39)	(0.38)	(0.39)	(0.39)	(0.40)	(0.39)	(0.39)
∞	16.09	16.11	16.10	16.10	16.07	16.05	16.04	16.01
	(0.43)	(0.43)	(0.42)	(0.41)	(0.42)	(0.41)	(0.43)	(0.43)

TABLE 3

POSTERIOR MEAN (WITH STANDARD ERROR IN PARENTHESES) OF τ FOR DIFFERENT COMBINATIONS (α, β).

$\alpha \setminus \beta$	0.25	0.50	0.75	1.00	1.25	1.50	1.75	2.00
1	9.78	9.53	9.52	9.61	10.05	10.73	10.99	11.56
	(0.59)	(0.61)	(0.65)	(0.78)	(0.85)	(1.00)	(1.06)	(1.09)
3	9.85	9.67	9.67	9.93	10.21	10.25	10.93	11.14
	(0.57)	(0.64)	(0.71)	(0.80)	(0.86)	(0.91)	(1.06)	(1.14)
5	10.08	10.08	9.88	9.89	9.96	10.17	10.48	10.88
	(0.63)	(0.69)	(0.70)	(0.84)	(0.87)	(0.91)	(1.12)	(1.11)
10	10.45	10.18	9.99	9.91	10.14	10.43	10.58	11.02
	(0.57)	(0.63)	(0.74)	(0.79)	(0.81)	(0.95)	(1.07)	(1.22)
∞	10.84	10.54	10.34	10.31	10.61	11.01	11.24	12.30
	(0.66)	(0.72)	(0.86)	(0.86)	(0.91)	(1.06)	(1.09)	(1.34)

TABLE 4

POSTERIOR MEAN OF τ^2 / σ^2 FOR DIFFERENT COMBINATIONS (α, β).

$\alpha \setminus \beta$	0.25	0.50	0.75	1.00	1.25	1.50	1.75	2.00
1	3.39	3.26	3.28	3.38	3.73	4.32	4.53	5.03
3	1.13	1.09	1.09	1.17	1.24	1.26	1.43	1.50
5	0.86	0.86	0.83	0.84	0.85	0.89	0.95	1.02
10	0.68	0.65	0.63	0.62	0.65	0.69	0.71	0.78
∞	0.46	0.43	0.42	0.41	0.44	0.48	0.50	0.60

4.3. Outlier diagnosis

One important feature of the scale mixtures distributions is that the posterior means of the mixing parameters can be used as global indicators of possible outliers. For the Student-*t* distribution, outlying observations are associated with small λ values, whereas for the *EP* distribution, outlying observations are associated with large u values. To identify possible unusual claims made by a particular occupation class, we inspect the posterior means of the λ_{ij}'s. In order to have a better graphical presentation, we plot the negative values of the logarithm of the posterior means of the λ_{ij}'s in Fig. 1 for a *t-EP* model with $\alpha = 5$ and $\beta = 1.5$. Four obvious outlying observations are identified from occupation classes 44, 57, 75 and 85. This finding is consistent with Klugman (1992, p. 141) who identifies the outliers from the residuals. The choice of α and β values here is for illustration purpose and, of course, other α and β values can be used. Posterior estimates of μ, σ, τ, θ_{44}, θ_{47}, θ_{75} and θ_{85} under a normal-normal and a Cauchy-Laplace models are given in Table 5. First of all, we notice that there is a significant improvement on the estimation of μ and σ while τ seems to be quite

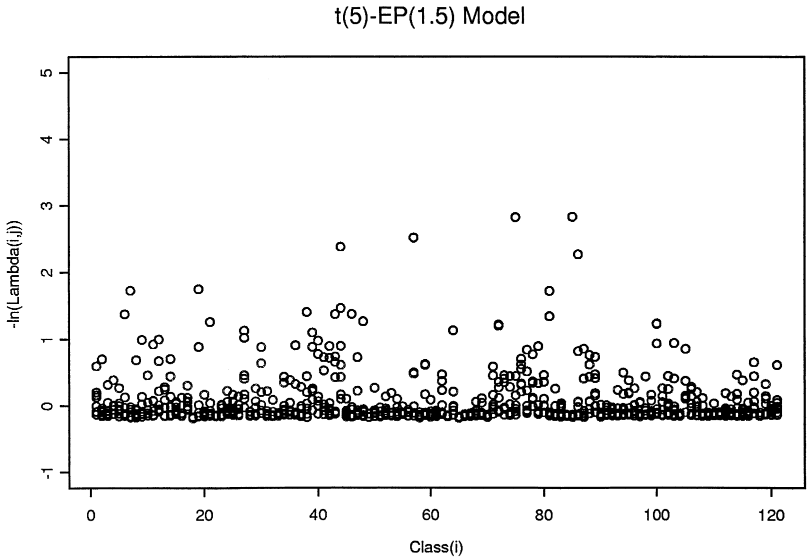

FIGURE 1: Plot of $-\ln(E[\lambda_{ij}|\boldsymbol{y}])$ against occupation class for a t(5)-EP(1.5) model. Large values correspond to the outlying observations.

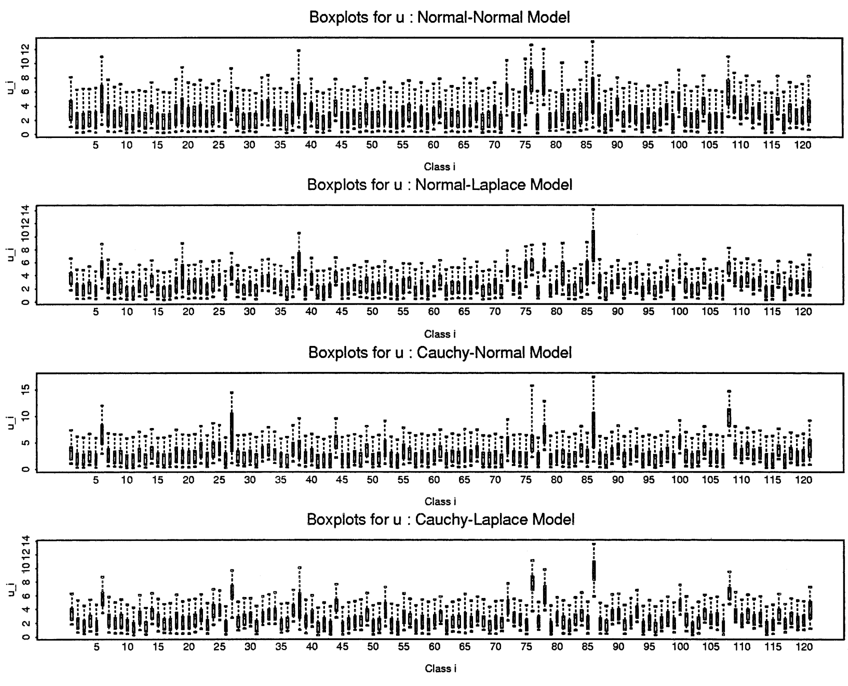

FIGURE 2: Boxplots of the uniform mixing parameter u_i's for the (i) normal-normal, (ii) normal-Laplace, (iii) Cauchy-normal and (iv) Cauchy-Laplace models. Outliers can be clearly identified using the mixing parameter.

insensitive. Occupation classes 75 and 85 give the two most extreme observations and the effects of these observations are almost ignored or being downweighed substantially by the Cauchy component, significantly different estimates for θ_{75} and θ_{85} are thus observed. We also find that occupation class 44 does not provide very unusual observations, but rather the whole class tends to have a large claim size when compared to other occupation classes.

In fact, when we look at the sample means of the pure premium y_{ij} for each occupation class, the sample values for classes 6, 27, 76, 78, 86 and 108 are 40.94, 34.46, 43.32, 44.48, 124.16 and 42.01 respectively, while for classes 44, 57, 75 and 85, sample values are 31.12, 17.12, 38.10 and 41.64. This result contradicts with the findings of Klugman (1992). Fig. 2 displays the boxplots of u_i's for the normal-normal, normal-Laplace, Cauchy-normal and Cauchy-Laplace models and the six outlying classes can be easily identified from the large values of u_i's. After removing observations from these six classes, posterior estimates of μ, σ and τ are presented in Table 6. A more robust analysis on μ and τ is then obtained. Therefore, these six occupation classes are found to belong to the high risk classes and should be analysed separately from the other occupation classes.

TABLE 5

POSTERIOR MEANS (WITH STANDARD ERRORS IN PARENTHESES) OF μ, σ, τ AND KLUGMAN'S OUTLYING θ_i'S FOR NORMAL-NORMAL AND CAUCHY-LAPLACE MODELS

Model	μ	σ	τ	θ_{44}	θ_{57}	θ_{75}	θ_{85}
Normal-Normal	17.03	16.10	10.31	31.24	12.66	33.90	29.40
	(0.96)	(0.41)	(0.86)	(0.90)	(1.12)	(6.50)	(9.19)
Normal-Laplace	15.64	15.99	12.30	31.26	12.83	37.11	34.82
	(0.98)	(0.43)	(1.34)	(0.90)	(1.64)	(7.90)	(14.20)
Cauchy-Normal	14.57	5.26	9.60	31.67	10.65	12.55	9.41
	(0.97)	(0.27)	(0.78)	(2.98)	(1.21)	(4.64)	(5.65)
Cauchy-Laplace	12.89	5.20	11.56	31.44	10.73	12.17	9.49
	(0.91)	(0.29)	(1.09)	(3.12)	(1.16)	(4.09)	(5.34)

TABLE 6

POSTERIOR MEANS (WITH STANDARD ERRORS IN PARENTHESES) OF μ, σ AND τ FOR NORMAL-NORMAL, NORMAL-LAPLACE, CAUCHY-NORMAL AND CAUCHY-LAPLACE MODELS WHEN THE OBSERVATIONS FROM OUTLYING CLASSES, 6, 27, 76, 78, 86 AND 108, ARE DISCARDED.

Model	μ	σ	τ
Normal-Normal	15.58 (0.90)	15.59 (0.44)	8.62 (0.68)
Normal-Laplace	15.11 (0.86)	15.57 (0.41)	9.88 (1.18)
Cauchy-Normal	13.26 (0.76)	5.09 (0.28)	8.02 (0.61)
Cauchy-Laplace	11.99 (0.94)	5.09 (0.29)	9.64 (1.03)

5. Concluding remarks

In this article, we introduce the Student-t and exponential power distributions via scale mixtures representation to insurance applications. As mentioned by Klugman (1992) and Herzog (1996), Bayesian hierarchical models arise naturally in many actuarial and insurance problems such as in credibility analysis. The use of Student-t and EP distributions provides a flexibility for statistical modelling in which modelling with normal distribution can be a special case. The scale mixtures forms of the Student-t and EP distributions allow the Gibbs sampling scheme for conjugate normal-normal models to extend straightforwardly to non-conjugate models without substantial increase in computational effort. In addition, analysis can be protected from the distorting effects of possible outliers at different stages of the models via the use of robustifying distributions. These outliers are easily identified by the mixing parameters of the scale mixtures distributions and can be suitably modelled by the heavy-tailed distributions.

A final remark is that in this article we attempt to illustrate the computing efficiency of using scale mixtures distributions within a Bayesian framework. We do not claim that the analysis for the insurance example studied in Section 4 is the most appropriate because some other factors such as inflation, linearity and autoregressive structure of the year-to-year claims, etc. have not been taken into account. However, our work has provided a starting point for actuarial practitioners to use heavy-tailed distributions via scale mixtures distributions.

Acknowledgements

This work was partially supported by a CRCG research grant award of the University of Hong Kong, HKSAR, China. The authors would like to thank Professor Stephen G. Walker for some valuable comments and Ms Carrie H.K. Yam for carrying out some computer simulation.

References

Andrews, D.F. and Mallows, C.L. (1974) Scale mixtures of normal distributions. *Journal of the Royal Statistical Society, Ser. B*, **36**, 99-102.

Berger, J.O. (1994) An overview of robust Bayesian analysis (with discussion). *TEST* **1**, 5-124.

Box, G.E.P. and Tiao, G.C. (1992) *Bayesian Inference in Statistical Analysis*. Addison Wesley, Massachusettes.

Bühlman, H. and Straub, E. (1972) Credibility for Loss ratios. English Translation in *ARCH*.

Choy, S.T.B. (1999) Contribution to the discussion of 'Robustifying Bayesian Procedures' by Walker, S.G. and Gutierrez-Pena, E. In *Bayesian Statistics 6* (eds. J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith, pp. 685-710, Oxford University Press, New York.

Choy, S.T.B. and Smith, A.F.M. (1997) Hierarchical models with scale mixtures of normal distributions. *TEST* **6**, 205-221.

Choy, S.T.B. and Walker, S.G. (1998) "The Extended Exponential Power Distribution and Bayesian Robustness". *Technical Report*, University of Hong Kong.

GELFAND, A.E., HILLS, S.E., RACINE-POON, A. and SMITH, A.F.M. (1990). Illustration of Bayesian inference in normal models using Gibbs sampling. *Journal of the American Statistical Association* **85**, 972-985.

HERZOG, T.N. (1996) *Introduction to Credibility Theory*. Second edition. ACTEX Publications, USA.

KLUGMAN, S.A. (1992) *Bayesian Statistics in Actuarial Science: With Emphasis on Credibility*. Kluwer Academic Publishers, Massachusetts.

KLUGMAN, S.A., PANJER, H.H. and WILLMOT, G.E. (1998) *Loss Models: From Data to Decisions*. Wiley, New York.

LANDSMAN, Z. and MAKOV, U.E. (1998) Exponential dispersion models and credibility. *Scandinavian Actuarial Journal* **1**, 89-96.

LANDSMAN, Z. and MAKOV, U.E. (1999) Sequential credibility evaluation for symmetric location claim distributions. *Insurance: Mathematics and Economics* **24**, 291-300.

MEYERS, G. (1984) Empirical Bayesian Credibility for Workers' Compensation Classification Ratemaking. *Proceedings of the Casualty Actuarial Society* **71**, 96-121.

SMITH, A.F.M. and ROBERTS, G.O. (1993) Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo method. *Journal of the Royal Statistical Society, Ser. B* **55**, 3-24.

TIERNEY, L. (1994) Markov chains for exploring posterior distributions (with discussion). *The Annals of Statistics* **22**, 1701-1762.

WALKER, S.G. and GUTIERREZ-PENA, E. (1999) Robustifying Bayesian Procedures (with discussion). In *Bayesian Statistics 6* (eds. J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith, pp. 685-710, Oxford University Press, New York.

YOUNG, V.R. (1997) Credibility using semiparametric models. *ASTIN Bulletin* **27**, 273-285.

YOUNG, V.R. (1998) Robust Bayesian credibility using semiparametric models. *ASTIN Bulletin*, **28**, 187-203.

S.T. BORIS CHOY
Department of Statistics and Actuarial Science
University of Hong Kong
Pokfulam Road, Hong Kong.
Email: bchoy@hkustasc.hku.hk

C.M. CHAN
Department of Statistics
The Chinese University of Hong Kong
Shatin, N.T., Hong Kong.
Email: cm_chan@sparc20c.sta.cuhk.edu.hk