Aust. N. Z. J. Stat. **49**(4), 2007, 415–434 doi: 10.1111/j.1467-842X.2007.00492.x

BAYESIAN ESTIMATION OF RUIN PROBABILITIES WITH A HETEROGENEOUS AND HEAVY-TAILED INSURANCE CLAIM-SIZE DISTRIBUTION

M. CONCEPCION AUSIN[1]* AND HEDIBERT F. LOPES[2]

Universidade da Coruña and University of Chicago

Summary

This paper describes a Bayesian approach to make inference for risk reserve processes with an unknown claim-size distribution. A flexible model based on mixtures of Erlang distributions is proposed to approximate the special features frequently observed in insurance claim sizes, such as long tails and heterogeneity. A Bayesian density estimation approach for the claim sizes is implemented using reversible jump Markov chain Monte Carlo methods. An advantage of the considered mixture model is that it belongs to the class of phase-type distributions, and thus explicit evaluations of the ruin probabilities are possible. Furthermore, from a statistical point of view, the parametric structure of the mixtures of the Erlang distribution offers some advantages compared with the whole over-parametrized family of phase-type distributions. Given the observed claim arrivals and claim sizes, we show how to estimate the ruin probabilities, as a function of the initial capital, and predictive intervals that give a measure of the uncertainty in the estimations.

Key words: Bayesian mixtures; heavy tails; multimodality; phase-type distributions; reversible jump MCMC; risk reserve processes.

1. Introduction

The choice of an appropriate model for claim sizes among the plethora discussed in the literature is crucial in the analysis of insurance processes. In the presence of large claims, for example, the ruin probability for a given initial capital can be underestimated if an inadequate model is considered for the claim-size distribution.

One of the most popular families considered to model claim sizes is the class of phase-type distributions (see, for example, Asmussen, 2000 and Rolski *et al*., 2000), which were introduced in a queueing context by Neuts (1981). Assuming phase-type-distributed claim sizes, it is possible to evaluate explicitly ruin probabilities for different risk models, avoiding the use of approximations such as the frequently used Cramér–Lundberg inequality. Furthermore, any positive distribution can be arbitrarily closely approximated by a phase-type distribution, owing to the denseness property. Many different fitting methods have been proposed for phase-type distributions, using moment matching methods (Johnson & Taaffe,

*Author to whom correspondence should be addressed.

[1]Departamento de Matemáticas, Facultade de Informática, Campus de Elviña, Universidade da Coruña, 15071 A Coruña, Spain.
e-mail: mausin@udc.es

[2]Graduate School of Business, University of Chicago, 5807 South Woodlawn Avenue, Chicago, IL 60637, USA.

Acknowledgments. The authors thank the editors and two anonymous referees for their helpful comments. The first author also acknowledges financial support by Xunta de Galicia IPP Program, MEC project SEJ2004-03303 and Graduate School of Business of the University of Chicago.

1991), maximum likelihood estimation (Asmussen, Nerman & Olsson, 1996), and Bayesian techniques (Bladt, Gonzalez & Lauritzen, 2003).

It is well known, however, that the whole class of phase-type distributions is a non-identifiable family because different sets of parameters may lead to the same phase-type distribution; see, for example, Neuts (1981). This over-parametrization can produce problems in fitting algorithms based on observed data, which eventually lead to poor parameter estimates and high computational costs. In order to reduce the number of parameters, two main subsets of phase-type distributions have been proposed in the literature: the Coxian distributions (see, for example, Faddy, 1994), and the mixtures of Erlang distributions (see, for example, Schmickler, 1992). These two distribution families are dense on the positive reals and, consequently, can approximate any positive distribution using a smaller number of parameters.

There has been much work on the calculation of ruin probabilities in insurance risk processes with phase-type claim sizes; see, for example, Asmussen & Rolski (1991), Avram & Usabel (2003), Dickson & Drekic (2004) and Drekic *et al*. (2004). However, their statistical estimation has received comparatively little attention. Inference for risk reserve processes and estimation of ruin probabilities have traditionally been carried out using maximum likelihood estimation (MLE) (see Asmussen, 2000), and non-parametric inference (see Hipp, 1989). A Bayesian approach to risk models has received recent interest in the insurance literature; see Cairns (2000) and Bladt *et al*. (2003). The Bayesian methodology offers a natural way to introduce the inherent parameter and model uncertainty in the estimation of ruin probabilities. Moreover, predictive distributions of the ruin probabilities can be obtained that are more informative than simple point estimations and allow the construction of confidence intervals. It is also possible to estimate the probability of having a positive safety loading and to incorporate this uncertainty directly in the estimation of the required ruin probability. Finally, it is well known that insurance risk models and queueing systems are mathematically closely related. Thus, we can make use of the large number of Bayesian approaches for queueing systems existing in the literature; see, for example, Armero & Bayarri (1994), Rios *et al*. (1998), Wiper (1998), Armero & Conesa (2004), Ausin *et al*. (2003) and Ausin, Wiper & Lillo (2004, 2007).

In this paper, we present a procedure for Bayesian inference and prediction of the classical compound Poisson process with unknown claim-size distribution. We propose a special re-parametrization of mixtures of Erlang distributions as a suitable model with which to approximate the claim-size behaviour, including the possibility of heterogeneity or extreme values. The dense family of mixtures of Erlang distributions includes simpler models such as the exponential, Erlang and hyperexponential distributions as special cases, and belongs to the class of phase-type distributions. Therefore, given the mixture parameters, explicit expressions of the ruin probabilities can be obtained. Bayesian inference for mixtures of Erlang distributions was considered earlier in Ausin *et al*. (2004), in the context of queueing systems and using the standard mixture parametrization. This standard formulation requires the use of a joint proper prior distribution for the mixture parameters, which leads to poor approximations of long-tailed distributions. The new parametrization proposed here will allow us to consider an improper prior and develop a straightforward Markov chain Monte Carlo (MCMC) implementation, obtaining good approximation of both long tails and multimodality with a relatively small number of parameters.

This paper is organized as follows. In Section 2, we briefly introduce the definition and some properties of the class of phase-type distributions and the subset of Coxian distributions, including some discussion on the problems with their parametrization. In Section 3, we carry out Bayesian inference for a re-parametrization of mixtures of Erlang distributions given a non-informative prior and an unknown number of mixture components. In Section 4, we develop Bayesian prediction for the classical compound Poisson process, introducing the uncertainty derived from the estimated claim-size distribution. We analyze the probability of having a positive safety loading and obtain point estimates and credible intervals for the ruin probabilities. We illustrate the methodology in Section 5 with a variety of simulated claim sizes, including long-tails and/or multimodality. We conclude in Section 6 with some discussion and extensions.

2. Phase-type distributions

A positive distribution is of phase-type (PH) with representation $(m, \boldsymbol{\alpha}, \boldsymbol{T})$ (see Neuts, 1981) if it represents the distribution of the time to absorption in a Markov jump process with m transient states and one absorbing state (state $m+1$). Then, the intensity matrix is given by

$$\boldsymbol{Q} = \begin{pmatrix} \boldsymbol{T} & \boldsymbol{t}_0 \\ \boldsymbol{0}^\top & 0 \end{pmatrix},$$

where $\boldsymbol{T}$ is a proper $m \times m$ subintensity matrix and the column vector of exit rates is $\boldsymbol{t}_0 = -\boldsymbol{T}\boldsymbol{e}$, where $\boldsymbol{e}$ is a column vector of ones. The initial probability vector is $(\boldsymbol{\alpha}, \alpha_{m+1})$ with $\boldsymbol{\alpha}\boldsymbol{e} + \alpha_{m+1} = 1$, where $\boldsymbol{\alpha}$ is a $1 \times m$ row vector of probabilities. The distribution function is given by

$$F(x) = 1 - \boldsymbol{\alpha} \exp(\boldsymbol{T}x)\boldsymbol{e}, \qquad \text{for } x > 0,$$

where $\exp(\boldsymbol{T}x)$ is the matrix exponential of $\boldsymbol{T}x$, and the matrix exponential of a matrix A is defined by the power series $\exp(A) = \sum_{k=0}^{\infty} A^k/k!$. The corresponding density function is given by

$$f(x) = \boldsymbol{\alpha} \exp(\boldsymbol{T}x)\boldsymbol{t}_0, \qquad \text{for } x > 0.$$

The simplest phase-type distribution is the exponential density, $f(x) = \mu \exp(-\mu x)$, with representation $(m, \boldsymbol{\alpha}, \boldsymbol{T}) = (1, 1, -\mu)$. A further two classical examples of phase-type distributions are the mixture of exponentials, $f(x) = \sum_{r=1}^{k} \omega_r \mu_r \exp(-\mu_r x)$, with $m = k$, $\boldsymbol{\alpha} = (\omega_1, \ldots, \omega_k)$ and $\boldsymbol{T} = \text{diag}(-\mu_1, \ldots, -\mu_\text{k})$, and the Erlang density

$$f(x) = \frac{(\nu\mu)^\nu}{\Gamma(\nu)} x^{\nu-1} \exp(-\nu\mu x), \qquad \text{for } x > 0, \tag{1}$$

with $m = \nu$, $\boldsymbol{\alpha} = (1, 0, \ldots, 0)$ and

$$\boldsymbol{T} = \begin{pmatrix} -\nu\mu & \nu\mu & & \\ & \ddots & \ddots & \\ & & -\nu\mu & \nu\mu \\ & & & -\nu\mu \end{pmatrix}_{\nu \times \nu}. \tag{2}$$

The whole class of phase-type distributions is a very versatile family and is dense over the set of positive distributions. Thus, any given positive density can be approximated arbitrarily closely by a phase-type distribution (see Neuts, 1981). Unfortunately, however, this distribution family is not identifiable, as the parameters $(m, \boldsymbol{\alpha}, \boldsymbol{T})$ do not determine uniquely one distribution; that is, different phase-type representations can lead to the same distribution. Furthermore, note that, for a given m, the number of free parameters of a phase-type distribution is $m^2 + m$. This over-parametrization will produce instability in the estimation algorithms and difficulties with parameter identifiability. Alternatively, a wide subclass of phase-type distributions, named Coxian distributions and also called acyclic phase-type (APH) or mixed generalized Erlang (MGE) distributions, has frequently been used in the literature (see, for example, Faddy, 1994). The Coxian distribution can be represented by $m, \boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_m)$, and

$$\boldsymbol{T} = \begin{pmatrix} -\lambda_1 & \lambda_1 & & & \\ & \ddots & \ddots & & \\ & & -\lambda_{m-1} & \lambda_{m-1} \\ & & & -\lambda_m \end{pmatrix}, \tag{3}$$

where, without loss of generality, it can be assumed that $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_m$. Note that the number of free parameters in a Coxian distribution is reduced to $2m - 1$, for each value of m. The Coxian family is also dense and is an appropriate model with which to approximate long-tailed distributions (see, for example, Horvath & Telek, 2000). However, the number of parameters required to capture peaked (platykurtic) density functions is still very large. To illustrate this, suppose we have a data set generated from an Erlang distribution, with the density as given in (1) and parameters $\nu = 50$ and $\mu = 0.5$. Then, if we consider a Coxian distribution to model these data, the estimation of the matrix (3) should be close to the matrix given in (2) with an estimated value for m close to 50. Thus, we should estimate around 100 parameters $(2m - 1)$ to obtain a good fitting. Thus, the Coxian model requires a large number of parameters to approximate multimodal distributions, and, clearly, this problem persists for densities that are simultaneously multimodal and platykurtic.

In the next section, we introduce and develop inference for the family of mixtures of Erlang distributions that can capture long tails, multimodality and/or low kurtosis using a smaller number of parameters.

3. Model and inference for the claim-size distribution

In this paper, we assume that the claim-size distribution follows a mixture of Erlang distributions with parameters k, $\boldsymbol{\omega} = (\omega_1, \ldots, w_k)$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)$ and $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_k)$, whose density function is given by

$$f(x \mid \boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\nu}) = \sum_{r=1}^{k} \omega_r \mathrm{Er}(x \mid \nu_r, \mu_r), \qquad \text{for } x > 0, \tag{4}$$

where $\sum_{r=1}^{k} \omega_r = 1$; $\omega_r, \mu_r > 0$; $\nu_r \in \mathbb{N}$; and $\mathrm{Er}(x \mid \nu_r, \mu_r)$ is the Erlang density function given in (1), whose mean and variance are given by $1/\mu_r$ and $1/\nu_r{}^{\mu_r^2}$, respectively, for $r = 1, \ldots, k$. Note that, under this distribution, each generic claim size, x, is with probability ω_r the sum of ν_r exponential phases of rate $\nu_r \mu_r$. Clearly, the distribution model (4) is a mixture of exponentials when $\nu_r = 1$, for $r = 1, \ldots, k$, and a single Erlang distribution when $k = 1$.

Mixtures of exponentials have been successful in modelling loss data for actuarial problems (see, for example, Keatinge, 1999).

A mixture of Erlang distributions is of phase-type as it is a mixture of phase-type distributions (see Neuts, 1981). It can be represented with $m = \sum_{r=1}^{k} \nu_r$; $\boldsymbol{\alpha} = [(\omega_1, 0, \dots 0)_{1\times\nu_1}, \dots, (\omega_k, 0, \dots, 0)_{1\times\nu_k}]$; and $\boldsymbol{T}$ equal to a block diagonal matrix with k blocks such that each one is of size $\nu_r \times \nu_r$, for $r = 1, \dots, k$, and is given by (2). Although it can be shown that the mixtures of Erlang distributions are a subset of the family of Coxian distributions (see Asmussen, 2000), we believe that the flexibility of the Erlang mixture model is, in practice, equivalent to the flexibility of the Coxian family with the advantage of having a more compact parametrization. In practice, we have not found any pattern that can be captured by a Coxian distribution but not by an Erlang mixture using the following Bayesian approach, as will be shown in the examples.

We now describe a Bayesian density estimation method for the claim-size distribution based on the Erlang mixture model (4). Our objective is to make Bayesian inference for the mixture parameters, given a sample of n observed claim sizes, $\boldsymbol{x} = \{x_1, \dots, x_n\}$. First, note that, as the mixture model is identifiable up to permutation of the rates, we can assume that

$$\mu_1 \geq \mu_2 \geq \cdots \geq \mu_k, \tag{5}$$

which is equivalent to assuming an increasing order for the means. Then we can consider the following re-parametrization:

$$\mu_r = \mu_1 \tau_2 \cdots \tau_r, \qquad \text{where } 0 < \tau_j \leq 1, \quad \text{for } r, j = 2, \dots, k.$$

This kind of re-parametrization was considered in Robert & Mengersen (1999) for normal mixtures, in Gruet *et al*. (1999) for exponential mixtures, and in Ausin *et al*. (2007) for Coxian distributions. Some of the known advantages of this type of parametrization are the use of non-informative priors and an improvement of the mixing in the MCMC algorithms.

We now define a prior distribution for the mixture parameters, $(k, \boldsymbol{\omega}, \mu_1, \boldsymbol{\tau}, \boldsymbol{\nu})$. We assume a flat prior for the number of mixture components, k; for example, a discrete uniform distribution defined on the interval $[1, k_{\max}]$. In our examples, we have chosen $k_{\max} = 20$, which, in practice, is large enough to capture the usual patterns observed in claim-size distributions.

Using the new re-parametrization, it is possible to define the following improper prior distribution for the first rate:

$$f(\mu_1) \propto \frac{1}{\mu_1}. \tag{6}$$

The choice of this improper prior for μ_1 will allow for the approximation of long-tailed distributions, as will be shown in the examples. This is because we do not make a strong assumption about the mean of the first component, and thus the remaining mean components of the mixture can be as large or as small as required. Also note that this improper prior distribution for μ_1 will imply an improper joint prior distribution. However, the joint posterior distribution is proper, as shown in the Appendix. Note that when using the standard parametrization, as in Ausin *et al*. (2004), it is not possible to use an improper prior for each mixture component rate, as this leads to an improper posterior distribution (see, for example, Diebolt & Robert, 1994).

Conditional on k, we define proper but diffuse prior distributions for the remaining parameters as follows:

$$\boldsymbol{\omega} \mid k \sim \text{Dirichlet}(\phi_1, \ldots, \phi_k), \tag{7}$$

$$\tau_r \mid k \sim \beta(a, b), \qquad \text{for } r = 1, \ldots, k, \tag{8}$$

$$\nu_r \mid k \sim G(p), \qquad \text{for } r = 1, \ldots, k. \tag{9}$$

In the examples, we have set $\phi_r = 1$ to give a uniform prior over the weights. In addition, we have set $a = b = 1$ to give a uniform prior over each τ_r, for $r = 1, \ldots, k$. Finally, we assume a not very large prior mean $1/p$ for ν_r in order to penalize a large number of phases in the Erlang mixture; for example, in the illustrations we have set $p = 0.05$.

In order to simplify the derivation of the conditional posterior distributions, we also introduce the usual missing data formulation for mixtures setups (see, for example, Diebolt & Robert, 1994), in which a set of latent variables $Z_1, \ldots, Z_n$ is defined such that

$$(X_i \mid Z_i = r) \sim \text{Er}(\nu_r, \mu_r), \qquad \Pr(Z_i = r \mid \boldsymbol{\omega}, k) = w_r,$$

for $r = 1, \ldots, k$. Thus, the observed data, $\boldsymbol{x} = (x_1, \ldots, x_n)$, are completed with a missing data set, $\boldsymbol{z} = (z_1, \ldots, z_n)$, indicating the specific mixture components from which the observations are assumed to arise.

Now, conditional on k, we can construct an MCMC algorithm to sample from the joint posterior distribution of the mixture parameters. The considered parametrization, together with the assumed prior distribution, allows for a straightforward implementation of a Gibbs sampling scheme where all the conditional posterior distributions are explicit. Given k, these conditional distributions are given by

$$\boldsymbol{\omega} \mid \boldsymbol{x}, \boldsymbol{z} \sim \text{Dirichlet}(1 + \text{n}_1, \ldots, 1 + \text{n}_\text{k}),$$

$$\mu_1 \mid \boldsymbol{\nu}, \boldsymbol{x}, \boldsymbol{z} \sim \gamma\Big(\sum_{r=1}^{k} n_r \nu_r, \ \sum_{r=1}^{k} \nu_r s_r \prod_{s=2}^{r} \tau_s\Big),$$

$$\tau_r \mid \tau_{-r}, \boldsymbol{\nu}, \boldsymbol{x}, \boldsymbol{z} \sim \gamma\left(1 + \sum_{j=r}^{k} n_j \nu_j, \ \sum_{j=r}^{k} \mu_1 \nu_j s_j \prod_{s=2, s\neq j}^{j} \tau_s\right) I(0 < \tau_r \leq 1), \tag{10}$$

$$f(\nu_r \mid \boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{x}, z) \propto \frac{\nu_r^{n_r \nu_r}(1-p)^{\nu_r}}{\Gamma(\nu_r)^{n_r}} \exp(-\nu_r(s_r \mu_r - n_r \log \mu_r - \log p_r)), \tag{11}$$

for $r = 2, \ldots, k$, where $\tau_{-r} = (\tau_1, \ldots, \tau_{r-1}, \tau_{r+1} \ldots, \tau_k)$, $I(\cdot)$ represents the indicator function, n_r is the number of observations assigned to the rth component, and s_r and p_r are the sum and the product of these observations, respectively. Note that we assume that s_r and p_r are equal to zero when n_r is zero.

Observe that the distribution of τ_r (10) is a truncated gamma density that can be sampled straightforwardly using a rejection sampling method such as that proposed in Philippe (1997). The conditional posterior distribution of ν_r(11) is a discrete distribution whose support is the whole set of positive integers. This can also be easily generated if we assume a truncation over a finite interval, $1 \le \nu_r \le \nu_{\max}$, for $r = 1, \ldots, k$. We have found that $\nu_{\max} = 100$ is large enough in most practical situations.

In order to sample from the posterior distribution of k, we make use of the reversible jump techniques introduced by Green (1995) and applied for normal mixtures in Richardson & Green (1997). The reversible jump algorithm is a generalization of the Metropolis–Hastings method for variable-dimension parameter spaces. Candidate values for the parameters are proposed, to allow a change in the number of mixture terms from k to $k \pm 1$, and then these candidates are accepted or rejected with the corresponding probability. We consider the so-called split and combine moves, in which two consecutive components in the sense of (5) are combined into one mixture component, and, to allow reversibility, one mixture component is split into two, respectively. If the r_1th and r_2th components are combined into the new rth component, the parameters are modified such that

$$\tilde{w}_r = w_{r_1} + w_{r_2}, \qquad \tilde{\tau}_r = \tau_{r_1}\tau_{r_2}, \qquad \tilde{\nu}_r = \nu_{r_2},$$

which implies that the proposed rate for the new component is $\tilde{\mu}_r = \mu_{r_2}$. Furthermore, it is considered that $\tilde{\mu}_1 = \mu_1 \tau_2$ when $r_1 = 1$. For a split move, the parameters of the new r_1th and r_2th components are given by

$$\begin{aligned} \tilde{w}_{r_1} &= u_1 w_r, & \tilde{w}_{r_2} &= (1-u_1)w_r, \\ \tilde{\tau}_{r_1} &= u_2 + \tau_r(1-u_2), & \tilde{\tau}_{r_2} &= \frac{\tau_r}{u_2 + \tau_r(1-u_2)}, \\ \tilde{\nu}_{r_1} &= u_3, & \tilde{\nu}_{r_2} &= \nu_r, \end{aligned}$$

where u_1 and u_2 are uniform $U(0,1)$, and $(u_3 - 1) \sim \text{Bin}(\nu_{\max}, \frac{\nu_r - 1}{\nu_{\max} - 1})$, a binomial distribution such that $\text{E}(\text{u}_3) = \nu_r$. This split move implies that $\tilde{\mu}_{r_2} = \mu_r$. For the case that $r = 1$, we generate $\tilde{\mu}_{r_1} = \mu_1 / u_2$ and $\tilde{\tau}_2 = u_2$, where $u_2 \sim U(0,1)$. Furthermore, those observations with $z_i = r$ are allocated to one of the r_1th or r_2th components with probability

$$\Pr(\tilde{Z}_i = r_j) \propto \tilde{\omega}_{r_j} \text{Er}(x_i \mid \nu_{r_j}, \mu_{r_j}), \qquad \text{for } j = 1, 2.$$

Note that these moves have been proposed following Gruet *et al*. (1999), and are chosen such that the parameters of the remaining mixture components are not modified. Also note that these moves do not preserve in general the mean and variance of X. The acceptance probability of a split move is $\min\{1, A\}$, where

$$A = \prod_{z_i = r} \frac{\tilde{\omega}_{\tilde{z}_i} \text{Er}(x_i \mid \tilde{\nu}_{\tilde{z}_i}, \tilde{\mu}_{\tilde{z}_i})}{\omega_{r_i} \text{Er}(x_i \mid \nu_r, \mu_r)} \frac{d_{L+1}}{b_L \prod_{z_i = r} \Pr(\tilde{Z}_i = \tilde{z}_i) q(u_3)} \frac{\omega_r (1 - \tau_r)}{u_2 + \tau_r(1-u_2)},$$

when $r > 1$, where d_L and b_L are respectively the probabilities of a combine or a split move, and $q(u_3)$ is the probability of having generated u_3 from the binomial distribution. Note that the last factor refers to the determinant of the Jacobian of the corresponding transformations. The acceptance probability for the reverse combine move can be obtained analogously.

Once we have run the MCMC algorithm, we obtain a sample of size J of the posterior distribution of the model parameters, $\{(k^{(j)}, \boldsymbol{\omega}^{(j)}, \boldsymbol{\mu}^{(j)}, \boldsymbol{\nu}^{(j)})\}_{j=1}^{J}$, and we can approximate the predictive density of the claim-size distribution by the usual Monte Carlo estimation approach:

$$f(x \mid \boldsymbol{x}) \approx \frac{1}{J} \sum_{j=1}^{J} \sum_{r=1}^{k^{(j)}} \omega_r^{(j)} \mathrm{Er}\left(x \mid \nu_r^{(j)}, \mu_r^{(j)}\right). \tag{12}$$

Note that this Bayesian density estimator is in fact a mixture of $\sum_{j=1}^{J} k^{(j)}$ Erlang distributions terms, which provides great flexibility to the resulting estimated density. This predictive density does not provide a single 'best' Erlang mixture model, but a coherent way of combining results over distinct Erlang mixture models. Analogously, we can estimate the predictive cumulated distribution function using the incomplete gamma function formula as follows:

$$F(x \mid \boldsymbol{x}) = 1 - \frac{1}{J} \sum_{j=1}^{J} \sum_{r=1}^{k^{(j)}} \omega_r^{(j)} \exp\left(-\nu_r^{(j)} \mu_r^{(j)} x\right) \sum_{i=0}^{\nu_r^{(j)}-1} \frac{\left(\nu_r^{(j)} \mu_r^{(j)} x\right)^i}{i!}. \tag{13}$$

Finally, we can also perform inference for the mixture parameters. For example, we can estimate the posterior distribution of the number of mixture components by

$$\Pr(k = r \mid \mathbf{x}) = \frac{1}{J} \sum_{j=1}^{J} I\left(k^{(j)=r}\right). \tag{14}$$

4. Bayesian prediction of ruin probabilities

Given an initial capital, u, we are now interested in estimating the ultimate ruin probability of an insurance company, which is defined as

$$\psi(u) = \Pr\left(\inf_{t \geq 0} R_t < 0 \mid R_0 = u\right), \tag{15}$$

where $\{R_t\}_{t \geq 0}$ is the risk reserve process describing the reserves of an insurance company, and is given by

$$R_t = u + ct - \sum_{i=1}^{N_t} U_i, \tag{16}$$

where c is the premium rate, N_t is the number of claims up to time t, and $U_1, U_2, \ldots$ are the claim sizes. We assume a classical compound Poisson model, where N_t is a homogeneous Poisson process with rate λ independent of the claim sizes, which are assumed to be independent and identically distributed (i.i.d.), following a mixture of Erlang distributions as described in the previous section.

In order that the ruin probability $\psi(u)$ differs from one for each value of the initial capital, u, it is required that the safety loading η is positive, where $\eta = c(\lambda\ \mathrm{E(U)})^{-1} - 1$; see, for example, Asmussen (2000). Thus, in risk theory, the safety loading is usually assumed to be positive, which implies that the process tends to plus infinity with probability one, and then the insurance company avoids certain eventual ruin. In practice, however, this condition is not always ensured; that is, it is not always known if the premiums are on average larger that the expected claims. Moreover, there is frequently evidence of a 'heavy traffic' in the data, which means that, on average, the claim sizes are only slightly smaller than the number of claim arrivals per unit time, in which case the safety loading is positive and small. Thus, we will not impose the condition $\eta > 0$ in our risk model, where the claim sizes follow a mixture of Erlang distributions and the safety loading is given by

$$\eta = c\left(\lambda \sum_{r=1}^{k} \frac{\omega_r}{\mu_r}\right)^{-1} - 1. \tag{17}$$

Furthermore, we are interested in estimating the posterior probability that the condition $\eta > 0$ holds, and how to incorporate the resulting uncertainty in the estimated ruin probabilities.

First, we will obtain an explicit expression for the ruin probability when the model parameters are fixed. It is well known that, for the compound Poisson model, the problem of obtaining the ultimate ruin probabilities for each given initial capital, u, is equivalent to calculating the stationary distribution of the waiting time, W, in a M/G/1 queueing system with the same arrival process (with time re-scaled by c) and the same distribution for the service times as for the claim sizes. This relation is given by

$$\psi(u) = P(W > u)$$

when $\eta > 0$ (see, for example, Asmussen, 2000). The stationary distribution of W can be explicitly calculated for the particular case of the M/PH/1 queue, with phase-type-distributed service times, as shown by Neuts (1981). In particular, if λ is the Poisson arrival rate and (m, $\boldsymbol{\alpha}$, $\boldsymbol{T}$) is the phase-type representation of the service time, the stationary waiting time W also follows a phase-type distribution with representation (m, $\boldsymbol{\beta}$, $\boldsymbol{S}$), where

$$\boldsymbol{\beta} = -\lambda\boldsymbol{\alpha}\boldsymbol{T}^{-1}, \qquad \boldsymbol{S} = \boldsymbol{T} + \boldsymbol{t}_0\boldsymbol{\beta}. \tag{18}$$

Thus, the ultimate ruin probability for a compound Poisson model with arrival rate λ, phase-type-distributed claim sizes with representation (m, $\boldsymbol{\alpha}$, $\boldsymbol{T}$), and positive safety loading is given by

$$\psi(u) = \tilde{\boldsymbol{\beta}} \exp(\tilde{\boldsymbol{S}}u)\boldsymbol{e} \tag{19}$$

(Asmussen & Rolski, 1991), where $\boldsymbol{e}$ is a column vector of ones, and $\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{S}}$ are respectively the $1 \times m$ vector and the $m \times m$ matrix given by

$$\tilde{\boldsymbol{\beta}} = -\frac{\lambda}{c}\boldsymbol{\alpha}\boldsymbol{T}^{-1} \qquad \text{and} \qquad \tilde{\boldsymbol{S}} = \boldsymbol{T} + \boldsymbol{t}_0\tilde{\boldsymbol{\beta}}.$$

Clearly, when the safety loading is not positive, the ultimate ruin probability is equal to one. As the Erlang mixture model is a phase-type distribution, we can now compute the ultimate ruin probability $\psi(u)$ for the considered compound Poisson model in which the claim sizes follow a mixture of Erlang distributions. Given a set of fixed parameters (λ, k, $\boldsymbol{\omega}$, $\boldsymbol{\mu}$, $\boldsymbol{\nu}$) such that the safety loading (17) is positive, we find that $m = \sum_{r=1}^{k} \nu_r$; $\tilde{\boldsymbol{\beta}} = (\tilde{\boldsymbol{\beta}}_{\nu_1}, \ldots, \tilde{\boldsymbol{\beta}}_{\nu_1})$, where

$$\tilde{\boldsymbol{\beta}}_{\nu_r} = \frac{\lambda\omega_r}{c\nu_r\mu_r}(1, \ldots, 1)_{1\times\nu_r},$$

and $\tilde{\boldsymbol{S}}$ is a block matrix in which the diagonal blocks are given by

$$\tilde{\boldsymbol{S}}_{\nu_r,\nu_r} = \begin{pmatrix} -\nu_r\mu_r & \nu_r\mu_r & & \\ & \ddots & \ddots & \\ & & -\nu_r\mu_r & \nu_r\mu_r \\ & & & -\nu_r\mu_r \end{pmatrix}_{\nu_r\times\nu_r} + \begin{pmatrix} 0 & \cdots & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 \\ \frac{\lambda\omega_r}{c} & \cdots & \cdots & \frac{\lambda\omega_r}{c} \end{pmatrix}_{\nu_r\times\nu_r},$$

and the off-diagonal block are given by

$$\tilde{\boldsymbol{S}}_{\nu_r,\nu_s} = \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \\ \frac{\lambda\omega_s\nu_r\mu_r}{c\nu_s\mu_s} & \cdots & \frac{\lambda\omega_s\nu_r\mu_r}{c\nu_s\mu_s} \end{pmatrix}_{\nu_r\times\nu_s}, \qquad \text{for } r \neq s.$$

Thus, once we have the expressions for $\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{S}}$, we can calculate the ruin probability $\psi(u)$ given in (19). Note that, for each initial capital u, we have to compute the matrix exponential, $\exp(\tilde{\boldsymbol{S}}u)$. There are a number of algorithms in the literature for computing the exponential of a matrix (see, for example, Moler & Van Loan, 1978). However, observe that we can obtain $\tilde{\boldsymbol{\beta}}\exp(\tilde{\boldsymbol{S}}u)$ directly as the solution of a linear system of differential equations, $\chi'(u) = \chi(u)\tilde{\boldsymbol{S}}$, with initial condition $\chi(0) = \tilde{\boldsymbol{\beta}}$. This can be done using, for example, a classical Runge–Kutta method of low order, (Abramowitz & Stegun, 1964).

We now address the inference problem for which the model parameters are no longer known, but we have a set of observed data from the risk process. Suppose that, in addition to the n previously observed claim sizes, $\boldsymbol{x} = \{x_1, \ldots, x_n\}$, we have observed independently a set of m inter-arrival times between claims, $\boldsymbol{t} = \{t_1, \ldots, t_m\}$, which are assumed to be i.i.d. as an exponential distribution with rate λ. An equivalent experiments have been considered within a queueing framework in a large number of Bayesian articles (see, for example, Armero & Bayarri, 1994). Because independence is assumed between the arrival and sizes of the claims, if we use a prior distribution for λ independent from the prior distribution of the claim-size parameters, $(k, \boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\nu})$, the corresponding posterior distributions will also be independent. Therefore, we assume the natural conjugate prior distribution for the arrival rate, $\lambda \sim \gamma(\xi, \delta)$, where $\xi, \delta > 0$, independent from the prior distribution of the claim-size parameters, $(k, \boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\nu})$, given in Section 2. It is straightforward to show that the posterior distribution of λ is given by

$$\lambda \mid \boldsymbol{t} \sim \gamma\left(\xi + m, \delta + \sum_{i=1}^{m} t_i\right) \tag{20}$$

(see also, for example, Armero & Bayarri, 1994).

Using a sample of size J generated from the posterior distribution of λ given in (20) and an MCMC sample of the same size from the posterior distribution of the claim-size parameters $(k, \boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\nu})$ obtained as described in Section 2, we can estimate various measures of interest by means of the usual Monte Carlo approximation. For example, the posterior probability that the safety loading is positive can be estimated with

$$\Pr(\eta > 0 \mid \boldsymbol{t}, \boldsymbol{x}) \approx \frac{R}{J}, \tag{21}$$

where R is the number of parameter vectors in the posterior sample $\{(\lambda^{(j)}, k^{(j)}, \boldsymbol{\omega}^{(j)}, \boldsymbol{\mu}^{(j)}, \boldsymbol{\nu}^{(j)})\}_{j=1}^{J}$ with positive safety loading calculated from (17). Analogously, we can approximate the posterior mean of the safety loading as an average of the safety loadings over the whole MCMC sample:

$$\mathrm{E}(\eta \mid \boldsymbol{t}, \boldsymbol{x}) \approx \frac{c}{J}\sum_{j=1}^{J}\left(\lambda^{(j)}\sum_{r=1}^{k^{(j)}}\frac{\omega_r^{(j)}}{\mu_r^{(j)}}\right)^{-1} - 1. \tag{22}$$

Note that the posterior mean of the safety loading is known to be finite, as shown in the Appendix. We can also estimate the posterior mean of η assuming stability, $\mathrm{E}(\eta \mid \eta > 0, \boldsymbol{t}, \boldsymbol{x})$, by simply rejecting those draws of the MCMC sample with $\eta \leq 0$. Using the same approach, we can estimate other interesting quantities, such as the posterior mean of the surplus process at a given future time, t:

$$\mathrm{E}(R_t \mid \boldsymbol{t}, \boldsymbol{x}) \approx u + ct - \frac{t}{J}\sum_{j=1}^{J}\left(\lambda^{(j)}\sum_{r=1}^{k^{(j)}}\frac{\omega_r^{(j)}}{\mu_r^{(j)}}\right). \tag{23}$$

The posterior variance of R_t and other statistical measures can be estimated analogously.

Now we can estimate the posterior mean of the ruin probability for each initial capital by the sample mean of the ruin probabilities calculated for each element of the posterior sample,

$$\psi(u \mid \boldsymbol{t}, \boldsymbol{x}) \approx \frac{1}{J}\sum_{j=1}^{J}\psi\left(u \mid \lambda^{(j)}, k^{(j)}, \boldsymbol{\omega}^{(j)}, \boldsymbol{\mu}^{(j)}, \boldsymbol{\nu}^{(j)}\right), \tag{24}$$

where $\psi(u \mid \lambda^{(j)}, \ldots)$ is obtained from (19) when $\eta^{(j)} > 0$, and is equal to one otherwise, where $\eta^{(j)}$ is the safety loading calculated from (17) for each value of the parameters of the MCMC sample. Observe that the estimated ruin probability (24) directly incorporates the uncertainty about the value of η without imposing the equilibrium condition, $\eta > 0$. However, if we wish to impose this condition, we should consider only those elements of the posterior sample that verify $\eta^{(j)} > 0$, and use the following approximation:

$$\psi(u \mid \eta > 0, \boldsymbol{t}, \boldsymbol{x}) \approx \frac{1}{R}\sum_{j:\eta^{(j)}>0}\psi\left(u \mid \lambda^{(j)}, k^{(j)}, \boldsymbol{\omega}^{(j)}, \boldsymbol{\mu}^{(j)}, \boldsymbol{\nu}^{(j)}\right). \tag{25}$$

Finally, note that we can also obtain credible intervals for the estimated ruin probabilities (24) and (25). For example, the posterior median and 95% credible intervals can be obtained by just calculating the median and the 0.025 and 0.975 quantiles of the posterior sample $\{\psi(u \mid \lambda^{(j)}, k^{(j)}, \boldsymbol{\omega}^{(j)}, \boldsymbol{\mu}^{(j)}, \boldsymbol{\nu}^{(j)})\}_{j=1}^{J}$.

5. Examples

In this section, we will illustrate the performance of the proposed methodology with various simulated risk processes covering a variety of features for the claim-size distribution, such as multimodality and/or heavy tails, and considering various values for the safety loading. We will also compare our Bayesian approach with the classical ML estimation based on the Expectation-Maximization (EM) algorithm for phase-type distributions proposed in Asmussen *et al.* (1996).

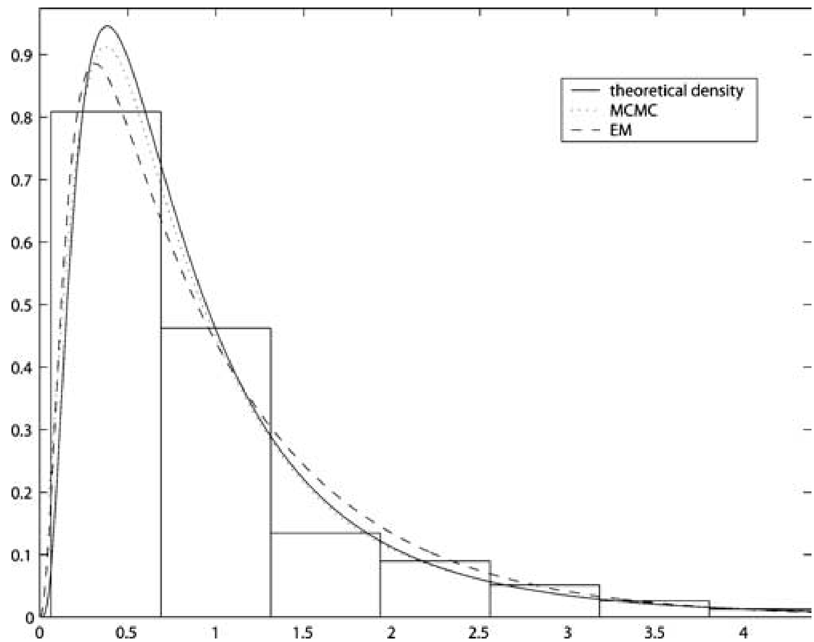

Figure 1. Histogram of the lognormal data set with the theoretical, Bayesian and maximum-likelihood density estimates.

We initially consider a classical simple example based on 250 data values generated from a lognormal distribution, LN $(-0.32, 0.8)$. This example has previously been considered in the literature to model claim-size data (see, for example, Asmussen & Rolski, 1991 and Bladt *et al*., 2003). Figure 1 illustrates a histogram of the data and the predictive density (12), estimated after running the MCMC algorithm for 10 000 burn-in iterations followed by an additional 10 000 iterations. This is compared with the classical density estimate based on the EM algorithm of Asmussen *et al*. (1996). The EM approach requires us to fix in advance the number of phases and then we chose $m = 4$, as suggested in Asmussen & Rolski (1991). With our Bayesian approach, the posterior probability of the mixture size, k, is concentrated between two and four components. Conditional on k, the estimated value for ν_r, $r = 1, \ldots, k$, varies from 1 to 20. Thus the predictive density is an average of densities with different numbers of phases that can be as small as 2 and as high as 140.

We now consider a bimodal and platykurtic data set. We generated 300 data values from a mixture of two Erlang components with parameters $k = 2$, $\boldsymbol{\omega} = (0.5, 0.5)$, $\boldsymbol{\mu} = (3, 0.6)$ and $\boldsymbol{\nu} = (1, 50)$. The sample kurtosis was approximately 1.32. Figure 2 shows the histogram of the data and the predictive density (12) obtained with the proposed MCMC algorithm. Our Bayesian method identified the correct number of components in the mixture, as the posterior mode of k is 2, with probability $P(k = 2 \mid \boldsymbol{x}) \simeq 0.72$, which was obtained using (14). Much smaller posterior probabilities were obtained for $k \neq 2$. Moreover, conditional on $k = 2$, the parameters are estimated correctly; the posterior means for $\boldsymbol{\omega}$, $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ were approximately given by (0.48, 0.52), (3.32, 0.59) and (1, 53.4), respectively. Figure 2 also compares the Bayesian estimation with two ML density estimates obtained from the EM algorithm of Asmussen *et al*. (1996) using a general phase-type and a Coxian distribution with 30 phases.

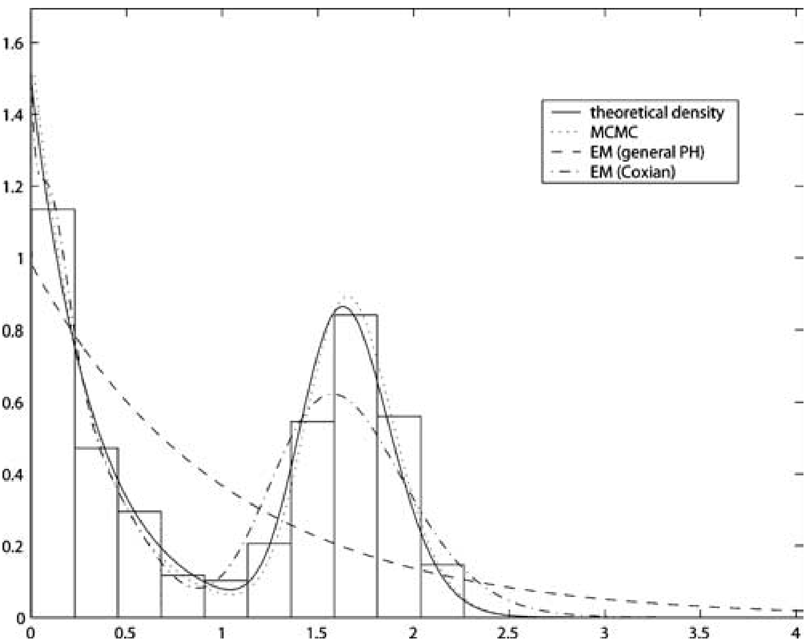

Figure 2. Histogram of the two-component mixture data set with the theoretical and Bayesian predictive densities compared with two maximum-likelihood density estimates based on a general phase-type and a Coxian distribution with 30 phases.

Although we should have fixed the number of phases equal to 51 (note that the true generating Erlang mixture is a phase-type distribution with $\Sigma \nu_r = 51$ phases), the EM algorithm was numerically unfeasible with such a large number of phases. Observe that, using 30 phases, the first ML density estimate, based on a general phase-type distribution, was statistically unstable and could not capture the bimodality of the distribution. The second ML density estimate, based on a Coxian distribution, gave a better fit, as the number of parameters to estimate was smaller. In this case, however, the peaked shape of the second mixture component was not well approximated, and a larger number of phases would be required for a better fit. Furthermore, the MLE required quite a large computational cost of about 40 minutes. In contrast, our Bayesian density estimate, which is quite satisfactory, required a moderate computational cost of approximately 10 minutes. Finally, we also observed that, if the number of mixture components was increased, especially if they had relatively small variances, there were still more difficulties in obtaining good density estimates using the EM approach, whereas the MCMC algorithm was just slightly slower in computational time but continued to be precise.

Finally, we consider two long-tailed distributions. It is well known that, although phase-type distributions are short-tailed, they can be used to approximate long-tailed distributions; see, for example, Feldmann & Whitt (1998) and Horvath & Telek (2000). We generated 300 data from a Weibull distribution, whose cdf is given by $F(x) = \exp(-(dx)^c)$, and we set $c = 0.3$ and d such that the distribution mean was 1. This long-tailed distribution model has been previously considered in Feldmann & Whitt (1998) and Ausin *et al*. (2007). In addition, in order to define a data set with both a bimodal and long-tailed distribution, we

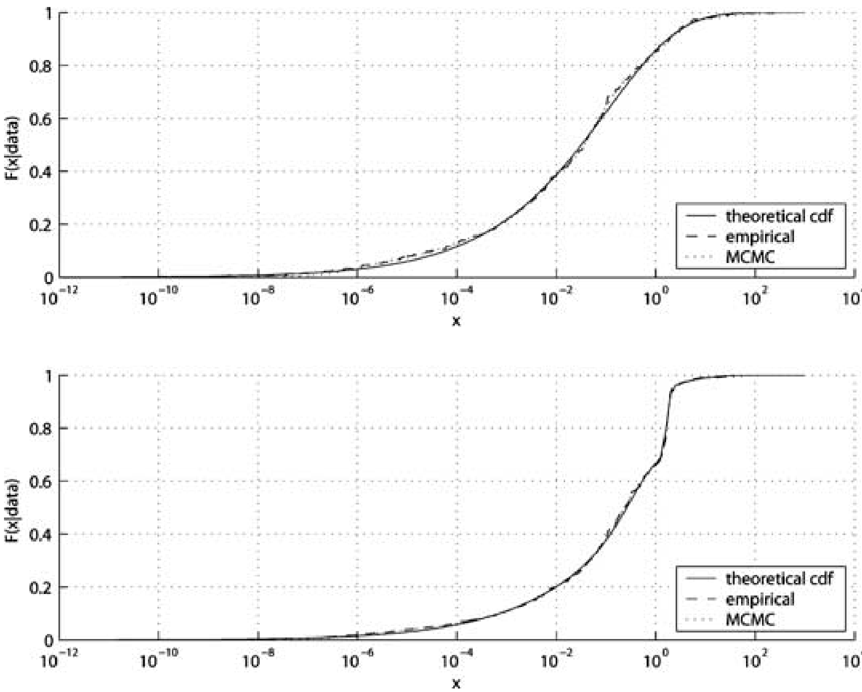

Figure 3. Theoretical, empirical and predictive cumulative distribution functions for the long-tailed Weibull data set (top) and for the mixture of long-tailed and bimodal data sets (bottom).

also considered the 300 observations from the previous example plus these 300 Weibull data. Figure 3 illustrates in log scale the theoretical, empirical and predictive cumulative distribution functions obtained with (13) for both examples. Observe that the fits are quite satisfactory, even for the bimodal case. The number of mixture components required to approximate these data was quite large. In fact, the posterior modes of k were 7 and 8 for the unimodal and bimodal data sets, respectively. Unfortunately, we have not been able to compare the predictive densities with the ML estimates, because the EM algorithm of Asmussen *et al*. (1996) was numerically infeasible for these data sets, whether we considered a general phase-type or a Coxian distribution with more than four phases. However, we compared these predictive distributions with those obtained using a Bayesian approach based on a Coxian distribution proposed in Ausin *et al*. (2007), and we observed that the Coxian model could capture the heavy-tailed behaviour, but that it does not approximate the bimodality very well. Moreover, the number of parameters and the computational cost required was much larger with the Coxian distribution than with the Erlang mixture model.

We can now address the problem of ruin probability estimation. We considered four simulated risk processes using the four claim-size data sets generated previously from a classical unimodal, a bimodal, a long-tailed and a long-tailed bimodal distribution. Note that the first classical example is also a long-tailed distribution, as the lognormal distribution is long-tailed, but the tail of the considered lognormal LN $(-0.32, 0.8)$ does not decay much more slowly than exponentially.

First, in order to examine the effects of 'heavy traffic' in the data, we fixed the safety loading η to be equal to 0.10, which implies that the average income of the insurance company

TABLE 1

Posterior probabilities that the safety loading, η, is positive and their posterior means unconditioned and conditioned on $\eta > 0$ for the four simulated risk processes

	Classical	Bimodal	Long-tailed	Long-tailed & bimodal
P ($\eta > 0$ | data)	0.9014	0.9856	0.0909	0.1469
E [η | data]	0.0918	0.0905	−0.3566	−0.1977
E [η | $\eta > 0$, data]	0.1062	0.0921	0.1636	0.0882

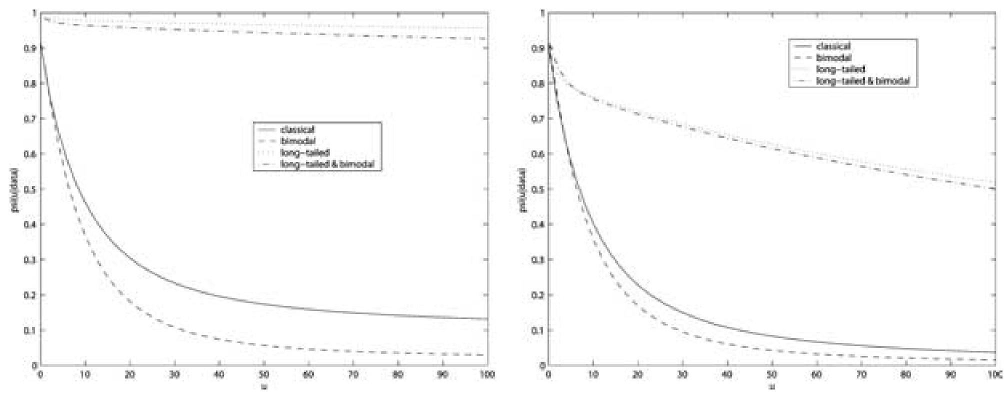

Figure 4. Predictive ruin probabilities as a function of the initial capital, u, for each of the four simulated risk processes unconditioned (left) and conditioned (right) on $\eta > 0$.

is only 10% larger than the average loss per unit of time. Note that the true mean claim size in the four cases is equal to 1, and then the value of the inter-arrival rate is given by $\lambda = 0.9091$ (see 17), which, for simplicity, is assumed to be known. We also assume that the premium rate is $c = 1$.

Table 1 shows the estimated posterior probabilities that the safety loading is positive (see 21), and their posterior means for the four risk processes (see 22). Observe that the posterior probability of having a stable process is very large for the classical and bimodal examples. However, we can observe in the last two cases that the long-tailed claim sizes together with a large inter-arrival rate have produced a small posterior probability of having a stable process. Furthermore, note that the posterior means of the safety loading are close to the true value in the classical and the bimodal example, but in the two long-tailed examples, it is only well estimated when the risk process is assumed to be stable.

Figure 4 illustrates the estimated ruin probabilities as a function of the initial capital, u, for each of the four cases unconditioned and conditioned on $\eta > 0$, obtained from (24) and (25) respectively. As expected, ruin probabilities are smaller when the safety loading is imposed to be positive. These differences are stronger for the long-tailed cases. With or without this restriction, the estimated ruin probabilities are much smaller for the classical and bimodal examples than for the long-tailed distributions. This should be expected because the sample coefficients of variation in the classical and bimodal examples are given by 1.04 and 0.72, respectively, which indicates that the ruin probabilities will be close to those obtained with an exponential claim-size distribution whose coefficient of variation is equal to 1. By contrast, the sample coefficients of variation in the last two long-tailed cases are given by

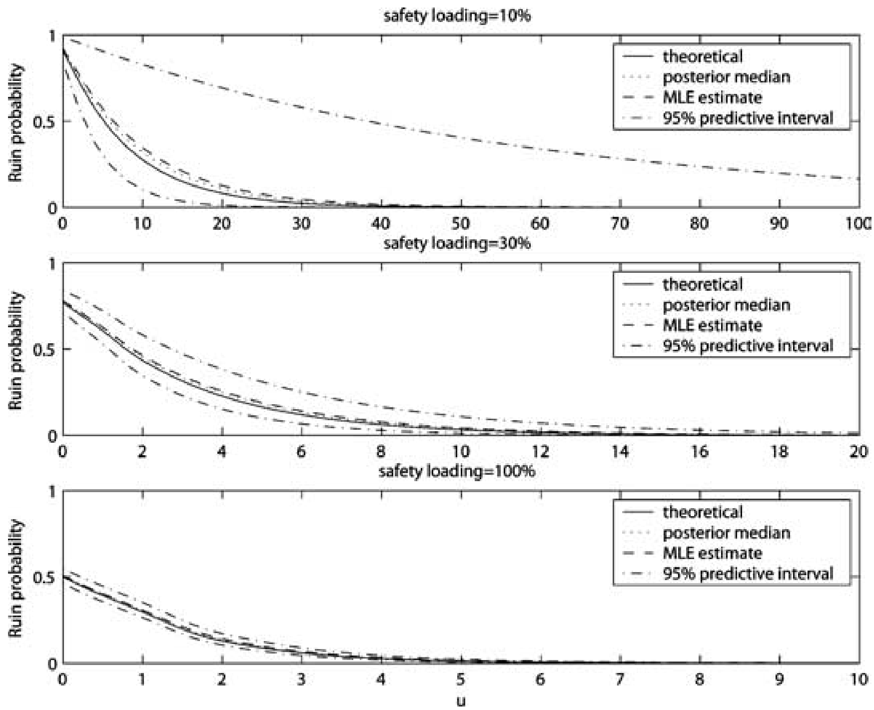

Figure 5. Posterior medians of the ruin probabilities and 95% predictive intervals as a function of the initial capital, u, for the bimodal claim-size data and various values of the safety loading compared with the EM estimates and theoretical values.

8.62 and 7.59, respectively, and then much larger ruin probabilities than for the exponential case should be expected.

Our Bayesian approach also allows for the calculation of predictive intervals for the estimated ruin probabilities, as mentioned in Section 4. Figure 5 illustrates the posterior medians and 95% predictive intervals of the ruin probabilities for the case with a bimodal claim-size distribution and various values of the safety loading. Note that the safety loadings are fixed at 10%, 30% and 100%, which implies that the inter-arrival rates are fixed at 0.909, 0.769, and 0.5, respectively. Observe that, the larger the safety loading, the wider the predictive intervals of the estimated ruin probabilities. Note also that the predictive intervals of the ruin probabilities are left-skewed, especially for larger safety loadings. Thus, the posterior medians of the ruin probabilities will be better estimates than the posterior means. In order to check the accuracy of the estimations, the posterior medians are compared with the theoretical ruin probabilities. Finally, the ruin probability estimates are also compared with those obtained with the EM approach. Note that both the ML estimates and theoretical ruin probabilities are similar to the Bayesian estimates and are always inside the predictive intervals.

Finally, note that, for each initial capital, we have a whole predictive sample of ruin probabilities that has been used to calculate the corresponding 95% predictive interval. This posterior sample can also be used to compute any required quantile of the posterior distribution of each ruin probability. For example, given an initial capital of $u = 6$ units, Figure 6 shows the histograms of the posterior samples of ruin probabilities that have been obtained to calculate

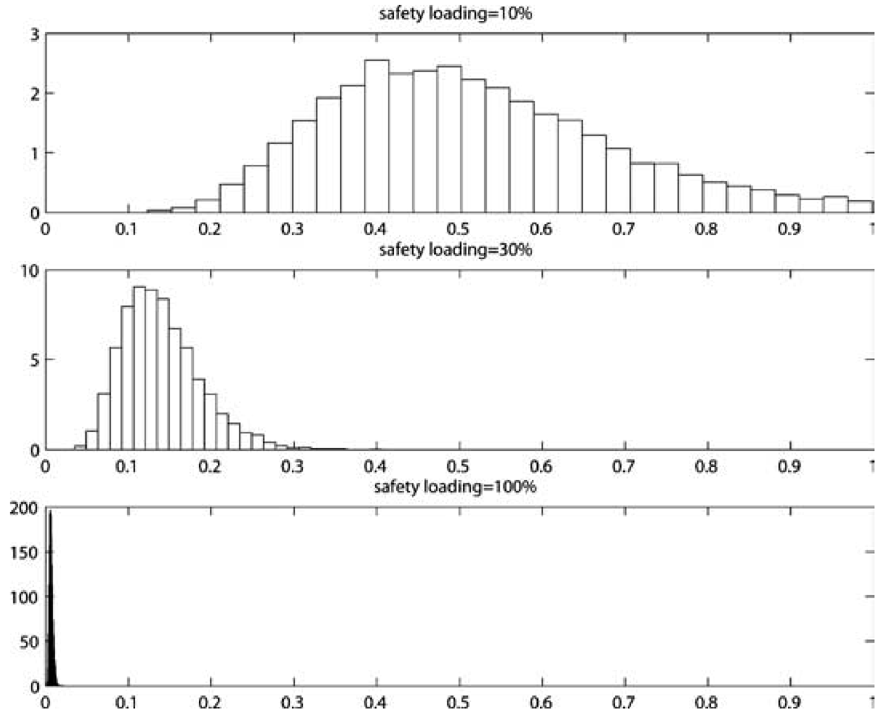

Figure 6. Histogram of the predictive samples of ruin probabilities for the bimodal claim-size data and various values of the safety loading given an initial capital of $u = 6$ units.

the predictive intervals in Figure 5 for $u = 6$. We can again observe the asymmetry of the predictive ruin probability distributions and that their variances decrease when the safety loading increases.

6. Discussion and extensions

In this article, we have developed a Bayesian approach to make inference and prediction for the classical compound Poisson risk reserve process with unknown claim-size distribution. First, we proposed a Bayesian density estimation method based on mixtures of Erlang distributions for the approximation of the general claim-size distribution. We illustrated this using a special re-parametrization of the Erlang mixture model with which it is possible to approximate both long-tailed and/or multimodal claim-size distributions.

Noting that the Erlang mixture model is a phase-type distribution, we also described a Bayesian procedure for the estimation of ruin probabilities and other system quantities. The proposed Bayesian approach provides confidence intervals for the estimated ruin probabilities and directly incorporates the uncertainty in the estimated safety loading without imposing the equilibrium condition.

Our Bayesian approach could be extended to the more general Sparre Andersen risk model by considering Erlang mixtures for both the claim inter-arrival and claim-size distributions. Explicit expressions for the ruin probabilities in this risk model can be derived using the phase-type properties of the Erlang mixture model (see Asmussen & Rolski, 1991 and

Asmussen, 2000). Furthermore, extensions to more complicated situations with non-independent inter-arrival times, such as Markov-modulated Poisson arrivals, should be possible using the Bayesian inference developed in Scott & Smyth (2003). Then ruin probabilities in risk processes with Markov-modulated Poisson arrivals and Erlang mixture claim sizes could be estimated using the results derived in Asmussen & Rolski (1991).

Many other results from the risk theory related to phase-type distributions could be combined with our Bayesian procedure in order to estimate other quantities of interest. For example, given the observed data, we could estimate the deficit at ruin in the aforementioned Sparre Andersen model using the results obtained by Drekic *et al*. (2004) and Dickson & Drekic (2004). Another example is the Bayesian estimation of finite-time ruin probabilities using the explicit expressions of their Laplace transform derived in Avram & Usabel (2003). In this case, the MCMC approach could be combined with numerical inversion methods of Laplace transforms in order to obtain point estimates and credible intervals for the finite-time ruin probabilities. Finally, a further extension is the estimation of ruin probabilities in multivariate compound Poisson risk models that could be developed using the results given in Cai & Li (2005).

Appendix

Here we show that the joint posterior distribution of the model parameters, $\boldsymbol{\theta} = (k, \boldsymbol{\omega}, \mu_1, \boldsymbol{\tau}, \boldsymbol{\nu})$, is proper given the improper joint prior distribution assumed in Section 3 (see (6), (7), (8) and (9)), and the likelihood corresponding to (4). Thus, we have to show that the following integral is finite:

$$\int f(\boldsymbol{\theta}) \prod_{i=1}^{n} \left(\sum_{r=1}^{k} \omega_r \mathrm{Er}(x_i \mid \nu_r, \mu_r) \right) d\boldsymbol{\theta}.$$

Assume first that the number of observations is $n = 1$. This integral is then proportional to

$$\int \frac{f(k) f(\boldsymbol{\omega} \mid k) f(\boldsymbol{\tau} \mid k) f(\boldsymbol{\nu} \mid k)}{\mu_1} \left(\sum_{r=1}^{k} \omega_r \frac{(\nu_r \mu_1 \tau_2 \cdots \tau_r)^{\nu_r}}{\Gamma(\nu_r)} x_1^{\nu_r - 1} \exp(-\nu_r \mu_1 \tau_2 \cdots \tau_r x_1) \right) d\boldsymbol{\theta}$$

$$= \int \sum_{r=1}^{k} \frac{\omega_r}{x_1} \int f(k) f(\boldsymbol{\omega} \mid k) f(\boldsymbol{\tau} \mid k) f(\boldsymbol{\nu} \mid k) \frac{(\nu_r \tau_2 \cdots \tau_r x_1)^{\nu_r}}{\Gamma(\nu_r)} \mu_1^{\nu_r - 1} \exp(-\nu_r \tau_2 \cdots \tau_r x_1 \mu_1) d\boldsymbol{\theta}$$

$$= \int \sum_{r=1}^{k} \frac{\omega_r}{x_1} \int f(k) f(\boldsymbol{\omega} \mid k) f(\boldsymbol{\tau} \mid k) f(\boldsymbol{\nu} \mid k) d\boldsymbol{\theta}_{-\mu_1} = \int \sum_{r=1}^{k} \frac{\omega_r}{x_1} f(\boldsymbol{\omega} \mid k) f(k) d\boldsymbol{\omega}\, dk = \frac{1}{x_1} < \infty,$$

where we have written $\boldsymbol{\theta}_{-\mu_1} = (k, \boldsymbol{\omega}, \boldsymbol{\tau}, \boldsymbol{\nu})$. Observe that it is sufficient to have proved that the integral is finite for $n = 1$, because if $n > 1$ we can define $f(\boldsymbol{\theta} \mid x_1)$ as a new proper *prior* and consider the likelihood based on $\{x_2, \ldots, x_n\}$, which is regular and proper, in which case the posterior is known to be proper.

We now show that the posterior mean of the safety loading given in (22) is also finite. Using the independence assumption between the arrival and size of claims,

$$\mathrm{E}(\eta \mid \mathbf{t}, \mathbf{x}) = \mathrm{E}\left(\left(\lambda \sum_{r=1}^{k} \frac{\omega_r}{\mu_r} \right)^{-1} - 1 \mid \mathbf{t}, \mathbf{x} \right) = \mathrm{E}(\lambda^{-1} \mid \mathbf{t}) \mathrm{E}\left(\left(\sum_{r=1}^{k} \frac{\omega_r}{\mu_r} \right)^{-1} \middle| \mathbf{x} \right) - 1,$$

which is finite, because, using (20),

$$\mathrm{E}(\lambda^{-1} \mid \mathbf{t}) = \frac{\delta + \sum_{i=1}^{m} t_i}{\xi + m - 1} < \infty,$$

for $\gamma > 1$, and, similarly to above, when $n = 1$ we have that $\mathrm{E}((\sum_{r=1}^{k} \frac{\omega_r}{\mu_r})^{-1} \mid x_1)$ is proportional to

$$\int \frac{f(k) f(\boldsymbol{\omega} \mid k) f(\boldsymbol{\tau} \mid k) f(\boldsymbol{\nu} \mid k)}{\mu_1 \sum_{r=1}^{k} \frac{\omega_r}{\mu_1 \tau_2 \cdots \tau_r}} \left(\sum_{r=1}^{k} \omega_r \frac{(\nu_r \mu_1 \tau_2 \cdots \tau_r)^{\nu_r}}{\Gamma(\nu_r)} x_1^{\nu_r - 1} \exp(-\nu_r \mu_1 \tau_2 \cdots \tau_r x_1) \right) d\boldsymbol{\theta}$$

$$= \frac{1}{x_1} \int \frac{f(k) f(\boldsymbol{\omega} \mid k) f(\boldsymbol{\tau} \mid k) f(\boldsymbol{\nu} \mid k)}{\sum_{r=1}^{k} \frac{\omega_r}{\tau_2 \cdots \tau_r}} \left(\sum_{r=1}^{k} \omega_r \frac{(\nu_r \tau_2 \cdots \tau_r x_1)^{\nu_r}}{\Gamma(\nu_r)} \mu_1^{\nu_r} \exp(-\nu_r \mu_1 \tau_2 \cdots \tau_r x_1) \right) d\boldsymbol{\theta}$$

$$= \frac{1}{x_1^2} \int \frac{f(k) f(\boldsymbol{\omega} \mid k) f(\boldsymbol{\tau} \mid k) f(\boldsymbol{\nu} \mid k)}{\sum_{r=1}^{k} \frac{\omega_r}{\tau_2 \cdots \tau_r}} \left(\sum_{r=1}^{k} \frac{\omega_r}{\nu_r \tau_2 \cdots \tau_r} \right) d\boldsymbol{\theta}$$

$$\leq \frac{1}{x_1^2} \int \frac{f(k) f(\boldsymbol{\omega} \mid k) f(\boldsymbol{\tau} \mid k) f(\boldsymbol{\nu} \mid k)}{\sum_{r=1}^{k} \frac{\omega_r}{\tau_2 \cdots \tau_r}} \left(\sum_{r=1}^{k} \frac{\omega_r}{\tau_2 \cdots \tau_r} \right) d\boldsymbol{\theta} = \frac{1}{x_1^2} < \infty.$$

Then, using the same arguments as above, this integral is also finite for $n > 1$.

References

ABRAMOWITZ, M. & STEGUN, I.A. (1964). *Handbook of Mathematical Functions*. New York: Dover.

ARMERO, C. & BAYARRI, M.J. (1994). Bayesian prediction in M/M/1 queues. *Queueing Syst.* **15**, 401–417.

ARMERO, C. & CONESA, D. (2004). Statistical performance of a multiclass bulk production queueing system. *Eur. J. Oper. Res.* **158**, 649–661.

ASMUSSEN, S. (2000). Ruin probabilities. In *Advanced Series on Statistical Science & Applied Probability* (Vol. 2). Singapore: World Scientific.

ASMUSSEN, S., NERMAN, O. & OLSSON, M. (1996). Fitting phase-type distributions via the EM algorithm. *Scand. J. Statist.* **23**, 419–441.

ASMUSSEN, S. & ROLSKI, T. (1991). Computational methods in risk theory: A matrix-algorithmic approach. *Insur. Math. Econ.* **10**, 259–274.

AUSIN, M.C., LILLO, R.E., RUGGERI, F. & WIPER, M.P. (2003). Bayesian modelling of hospital bed occupancy times using a mixed generalized Erlang distribution. In *Bayesian Statistics* 7, eds J.M. BERNARDO, M.J. BAYARRI, J.O. BERGER, A.P. DAWID, D. HECKERMAN, A.F.M. SMITH, M. WEST, pp. 443–452. Oxford: Oxford University Press.

AUSIN, M.C., WIPER, M.P. & LILLO, R.E. (2004). Bayesian estimation for the M/G/1 queue using a phase type approximation. *J. Statist. Plan. Infer.* **118**, 83–101.

AUSIN, M.C., WIPER, M.P. & LILLO, R.E. (2007). Bayesian prediction of the transient behaviour and busy period in short- and long-tailed GI/G/1 queueing systems. *Comput. Stat. Data An.* (in press).

AVRAM, F. & USABEL, M. (2003). Finite time ruin probabilities with one Laplace inversion. *Insur. Math. Econ.* **32**, 371–377.

BLADT, M., GONZALEZ, A. & LAURITZEN, S.L. (2003). The estimation of phase-type related functionals using Markov Chain Monte Carlo methods. *Scand. Actuarial J.* **2003**, 280–300.

CAI, J. & LI, H. (2005). Multivariate risk model of phase type. *Insur. Math. Econ.* **36**, 137–152.

CAIRNS, A.J.G. (2000). A discussion of parameter and model uncertainty in insurance. *Insur. Math. Econ.* **27**, 313–330.

DICKSON, D.C.M. & DREKIC, S. (2004). The joint distribution of the surplus prior to ruin and the deficit at ruin in some Sparre Andersen models. *Insur. Math. Econ.* **34**, 97–107.

DREKIC, S., DICKSON, D.C.M., STANFORD, D.A. & WILLMOT, G.E. (2004). On the distribution of the deficit at ruin when claims are phase-type. *Scand. Actuarial J.* **2**, 105–120.

DIEBOLT, J. & ROBERT, C.P. (1994). Estimation of finite mixture distributions through Bayesian sampling. *J. Roy. Statist. Soc. B* **56**, 363–375.

FADDY, M.J. (1994). Examples of fitting structured phase-type distributions. *Appl. Stoch. Model. Data Anal.* **10**, 247–255.

FELDMANN, A. & WHITT, W. (1998). Fitting mixtures of exponentials to long-tail distributions to analyze network performance models. *Perform. Evaluation* **31**, 245–279.

GREEN, P. (1995). Reversible jump MCMC computation and Bayesian model determination. *Biometrika* **82**, 711–732.

GRUET, M. A., PHILIPPE, A. & ROBERT, C.P. (1999). MCMC control spreadsheets for exponential mixture estimation. *J. Comput. Graph. Statist.* **8**, 298–317.

HIPP, C. (1989). Estimates and boostrap confidence intervals for ruin probabilities. *Astin Bull.* **19**, 57–70.

HORVATH, A. & TELEK, M. (2000). Approximating heavy tailed behavior with phase-type distribution. In *Advances in Algorithmic Methods for Stochastic Models*, eds G. LATOUCHE, P. TAYLOR, pp. 191–214. New Jersey: Notable Publications.

JOHNSON, M.A. & TAAFFE, M.R. (1991). An investigation of phase-distribution moment-matching algorithms for use in queueing models. *Queueing Syst.* **8**, 129–148.

KEATINGE, C.L. (1999). Modeling losses with the mixed exponential distribution. Proceedings of the Casualty Actuarial Society, 1999, Volume LXXXVI, **165**, 654–681.

MOLER, C. & VAN LOAN, C. (1978). Nineteen dubious ways to compute the exponential of a matrix. *SIAM Rev.* **20**, 801–836.

NEUTS, M.F. (1981). *Matrix Geometric Solutions in Stochastic Models*. Baltimore: Johns Hopkins University Press.

PHILIPPE, A. (1997). Simulation of right- and left-truncated gamma distribution by mixtures. *Statist. Comput.* **7**, 173–182.

RICHARDSON, S. & GREEN, P. (1997). On Bayesian analysis of mixtures with an unknown number of components. *J. Roy. Statist. Soc. B* **59**, 731–792.

RIOS, D., WIPER, M.P. & RUGGERI, F. (1998). Bayesian analysis of $M/Er/1$ and $M/H_k/1$ queues. *Queueing Syst.* **30**, 289–308.

ROBERT, C.P. & MENGERSEN, K.L. (1999). Reparameterisation issues in mixture modelling and their bearing on MCMC algorithms. *Comput. Statist. Data Anal.* **29**, 325–343.

ROLSKI, T., SCHMIDLI, H., SCHMIDT, V. & TEUGELS, J. (2000). *Stochastic Processes for Insurance and Finance*. New York: John Wiley & Sons.

SCHMICKLER, L. (1992). MEDA: mixed Erlang distributions as a phase-type representation of empirical distribution functions. *Stoch. Models* **8**, 131–156.

SCOTT, S.L. & SMYTH, P. (2003). The Markov modulated Poisson process and Markov Poisson cascade with applications to web traffic data. In *Bayesian Statistics* 7, eds J.M. BERNARDO, M.J. BAYARRI, J.O. BERGER, A.P. DAWID, D. HECKERMAN, A.F.M. SMITH, M. WEST. pp. 671–680. Oxford: Oxford University Press.

WIPER, M.P. (1998). Bayesian analysis of $Er/M/1$ and $Er/M/c$ queues. *J. Statist. Plan. Infer.* **69**, 65–79.