Bayesian analysis of aggregate loss models.

M. C. Ausín*, J. M. Vilar, R. Cao and C. González-Fragueiro.

Department of Mathematics, Universidade da Coruña

Abstract

This paper describes a Bayesian approach to make inference for aggregate loss models in the insurance framework. A semiparametric model based on Coxian distributions is proposed for the approximation of both the inter-arrival time between claims and the claim size distributions. A Bayesian density estimation approach for the Coxian distribution is implemented using reversible jump Markov Chain Monte Carlo (MCMC) methods. The family of Coxian distributions is a very flexible mixture model that can capture the special features frequently observed in insurance claims such as long tails and multimodality. Furthermore, given the proposed Coxian approximation, it is possible to obtain closed expressions of the Laplace transforms of the total claim count and the total claim amount random variables. This properties allow us to obtain Bayesian estimations of the distributions of the number of claims and the total claim amount in a future time period, their main characteristics and credible intervals. The possibility of applying deductibles and maximum limits is also analyzed. The methodology is illustrated with a real data set provided by the insurance department of an international commercial company.

Keywords: Aggregate losses, Bayesian inference, mixtures, censored claims, MCMC methods, predictive distributions, Laplace transforms.

[1]Corresponding author. M. C. Ausín, Departamento de Matemáticas, Facultad de Informática, Campus de Elviña, Universidade da Coruña, 15071 A Coruña, Spain. Tel.: +34 981 167 000 (ext. 1318); fax: + 34 981 167 160. E-mail address: mausin@udc.es

1 Introduction

In this paper, we are mainly interested in the estimation of the total claim amount up to time t,

$$S(t) = \sum_{j=1}^{N(t)} Y_j,$$

where $N(t)$ is the number of claims up to time t and $Y_1, Y_2, ...$ are the claim sizes, with the usual convention that $S(t) = 0$ if $N(t) = 0$. It is assumed that $N(t)$ is a renewal process such that the inter-arrival times between successive claims are independent and identically distributed (i.i.d.) and the claim sizes, $Y_1, Y_2, ...$, are also i.i.d. random variables which are independent of the claim arrival process.

The selection of appropriate models for the claim arrival process and the claim size distribution is essential in the estimation of the distributions of the total claim count, $N(t)$, and the total claim amount, $S(t)$. In classical risk theory, it is very common to assume a homogeneous Poisson process for the claim arrival process since this assumption simplifies the derivation of the total claim amount distribution. Also, a gamma distribution model is frequently assumed to describe the usual right skewed shape of the claim size distributions. However, the exponential or gamma distributions are not always realistic models in practice as they cannot capture multimodality, heavy-tails or extreme events which are usually exhibited in insurance data, see e.g. Cizek et al. (2005). Alternatively, in this paper, we propose a renewal model where both inter-arrival claims and claim sizes follow Coxian distributions. The class of Coxian distributions is dense in the set of positive distributions and then, any positive density can be arbitrarily closely approximated by a Coxian distribution, see e.g. Asmussen (2000). Moreover, the Coxian model is a phase-type distribution, see e.g. Neuts (1981), which essentially means that the distribution can be decomposed in a number of exponential stages and then, closed expressions concerning quantities of interest, such as the total claim amount, can be obtained.

In practice, the distributions and parameter values describing the behaviour of claims are unknown and an insurance company only have past information about the frequency and amount of losses. Assume for example that the company have collected data during a past time period, $([0, T]$, observing a sequence of claims at times $0 = t_0 \leq t_1 \leq t_2 \leq \ldots \leq t_n \leq T$. Letting $\tau_j = t_j - t_{j-1}$, for $j = 1, ..., n$, we obtain a data sample of n inter-ocurrence times, $D_\tau = \{\tau_1, ..., \tau_n\}$, which have been assumed to be i.i.d. On the other hand, suppose that the company have observed a sample of n claim sizes which have also been assumed to be i.i.d. and independent of the inter-ocurrence time data. Note that insurance claim size data present frequently left-censoring as some claim sizes are only known to be smaller than certain values, e.g. the deductibles, and their precise values are not registered by the insurance company as it is not in charge of their payment.

Then, assume that we have a data sample of left-censored claim sizes, $D_Y = \{(X_1, \delta_1), \ldots, (X_n, \delta_n)\}$, with $X_j = \max(Y_j, C_j)$, where Y_j is the value of the j-th claim size, C_j is the censoring variable and δ_j is the censoring indicator variable such that $\delta_j = I(Y_j \geq C_j)$. Thus, given the observed claim inter-arrivals and claim sizes, $\{D_\tau, D_Y\}$, which have been collected during a past time period, the insurance company is mainly interested in the total number of claims and the total claim amount in future time periods.

In classical actuarial methods, model and parameter uncertainty is frequently ignored when making predictions for future time periods. For example, the total claim amount distribution is usually estimated based on fitted distributions for the inter-claim times and claim sizes distributions without considering the parameter uncertainty. Alternatively, several statistical approaches can be adopted to measure the uncertainty in the estimation of unobserved future variables. In particular, the Bayesian methodology provides a natural way to calculate predictive distributions which are much more informative than simple density point estimates, see e.g. Klugman (1992) and Dickson et al. (1998). Given past claim data, Bayesian future predictions are based on the posterior predictive densities of the total claim count and the total claim amount. Predictive distributions incorporate both the uncertainty due to the stochastic nature of the model and the parameter uncertainty, see e.g. Cairns (2000). Finally, using Bayesian prediction, we can also obtain credible intervals for the main characteristics of the total claim count and the total claim amount, such as the mean, median, standard deviation, quantiles, etc.

In this paper, we adopt a Bayesian approach for the estimation of predictive distributions of the total claim count and total claim amount in a future time period. Firstly, we carry out Bayesian density estimation based on Coxian distributions for the random variables representing the inter-ocurrence times between claims and the claim sizes. A non-informative prior density is defined for the Coxian parameters in order to develop objective Bayesian inference. Our approach also includes the possibility of censored claim sizes. Given the estimated inter-arrival time and claim size distribution, we obtain estimations of the predictive distributions for the total number of claims and aggregate losses in a future time period. Furthermore, we explore the same problem under the presence of deductibles and policy limits.

The rest of this paper is organized as follows. In Section 2, we introduce the Coxian distribution model which is assumed for both the inter-claim times and the claim sizes. In Section 3, we describe a Bayesian density estimation method for the Coxian distribution given past claim data. Section 4 is devoted to the estimation of the distribution of the number of claims and the total claim amount in a future time period. Section 6 describes how to estimate the aggregate claim distribution when deductibles and maximum limits are applied to the individual losses. A real application of the proposed methodology is presented in Section 5. Results are compared with a different statistical approach, developed in González-Fragueiro et al. (2006), based on nonparametric estimation and bootstrap methods. Section 6 concludes with some discussion.

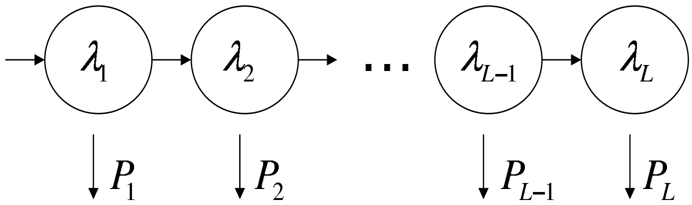

Figure 1: Graphical illustration of the Coxian distribution model.

2 The Coxian distribution model

In this paper, we assume that each inter-ocurrence time, τ, between two consecutive claims follows a Coxian distribution model with parameters $\boldsymbol{\theta}_\tau = \{L, \mathbf{P}, \boldsymbol{\lambda}\}$, where $\mathbf{P} = (P_1, ..., P_L)$ and $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_L)$, that is,

$$\tau = \begin{cases} \eta_1, & \text{with prob} = P_1, \\ \eta_1 + \eta_2, & \text{with prob} = P_2, \\ \vdots & \vdots \\ \eta_1 + ... + \eta_L, & \text{with prob} = P_L, \end{cases} \tag{1}$$

where $\eta_r \sim \exp(\lambda_r)$ and $\sum_{r=1}^{L} P_r = 1$. Then, the corresponding density function of τ can be expressed in terms of a mixture model,

$$f(\tau \mid L, \mathbf{P}, \boldsymbol{\lambda}) = \sum_{r=1}^{L} P_r f_r(\tau \mid \lambda_1, ..., \lambda_r), \qquad \tau > 0, \tag{2}$$

where f_r is the density function of a sum of r exponentials, also called generalized Erlang, whose density is given by,

$$f_r(\tau \mid \lambda_1, ..., \lambda_r) = \sum_{t=1}^{r} \left(\prod_{s \neq t} \frac{\lambda_s}{\lambda_s - \lambda_t} \right) \lambda_t \exp\{-\lambda_t \tau\}, \tag{3}$$

when all rates are distinct, see Johnson and Kotz (1970). Note that, without loss of generality, it can be assumed that $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_L$. Figure 1 shows a graphical illustration of this distribution model.

Throughout, we will also assume that the claim size random variable, Y, follows a Coxian distribution model with parameters $\boldsymbol{\theta}_Y = \{M, \mathbf{Q}, \boldsymbol{\mu}\}$, where $\mathbf{Q} = (Q_1, ..., Q_L)$ and $\boldsymbol{\mu} = (\mu_1, ..., \mu_L)$, as described above. Then, the density function is given by,

$$f(y \mid M, \mathbf{Q}, \boldsymbol{\mu}) = \sum_{r=1}^{M} Q_r \sum_{t=1}^{r} C_{tr}\ \mu_t \exp\{-\mu_t y\}, \qquad y > 0, \tag{4}$$

where,

$$C_{tr} = \prod_{s=1, s\neq t}^{t} \frac{\mu_s}{\mu_s - \mu_t}. \tag{5}$$

The Coxian distribution model is very flexible and appropriate to capture the special features frequently observed in insurance claim sizes such as long tails, multimodality and extreme events. In fact, due to the denseness property of the Coxian distributions, it is possible to approximate any continuous density function over the positive real line by increasing the number of mixture components, L. Note that the Coxian model is a mixture of generalized Erlang distributions and then, it contains the exponential, Erlang and exponential mixture distributions, as special cases.

In the next section, we describe how to develop Bayesian density estimation for this mixture model assuming that all parameters, including the number of mixture components are unknown. We make use of the reversible jump MCMC methods introduced in Richardson and Green (1997) for normal mixtures and considered for many mixtures models in the literature, see e.g. Robert and Mengersen (1999), Gruet et al. (1999), Wiper et al. (2001) and Ausín et al. (2004).

3 Estimation of the claim interarrival and claim size distributions

Given a sample of n inter-occurrence times between claims, $D_\tau = \{\tau_1, ..., \tau_n\}$, following a Coxian distribution with parameters, $\boldsymbol{\theta}_\tau = (L, \mathbf{P}, \boldsymbol{\lambda})$, we wish to develop Bayesian inference and estimate the density of τ given the observed data, D_τ. Thus, we might define a prior distribution for the model parameters, $\pi(\boldsymbol{\theta}_\tau)$, and obtain the posterior distribution,

$$\pi(\boldsymbol{\theta}_\tau \mid D_\tau) \propto l(\boldsymbol{\theta}_\tau \mid D_\tau)\, \pi(\boldsymbol{\theta}_\tau), \tag{6}$$

where $l(\boldsymbol{\theta}_\tau \mid D_\tau)$ is the likelihood function. Given the posterior distribution, a Bayesian density estimation of the inter-arrival time is given by the posterior mean of the Coxian density function, called the predictive density of τ,

$$f(\tau \mid D_\tau) = \int_{\Theta_\tau} f(\tau \mid \boldsymbol{\theta}_\tau)\, \pi(\boldsymbol{\theta}_\tau \mid D_\tau)\, d\boldsymbol{\theta}_\tau, \tag{7}$$

where $f(\tau \mid \boldsymbol{\theta}_\tau)$ is the Coxian density model given in (2). Unfortunately, analytical calculus of the posterior distribution (6) is not straightforward for the Coxian parameters, $\boldsymbol{\theta}_\tau$. However, given a prior distribution, Bayesian inference may be performed using MCMC methods. These involve the construction of a Markov chain $\{\boldsymbol{\theta}_\tau^{(k)} : k = 1, 2, ...\}$, where $\boldsymbol{\theta}_\tau^{(k)} = \left(L^{(k)}, \mathbf{P}^{(k)}, \boldsymbol{\lambda}^{(k)}\right)$, with the posterior distribution $\pi(\boldsymbol{\theta}_\tau \mid D_\tau)$ as its stationary distribution, see e.g. Gilks et al. (1996). Using a sample $\{\boldsymbol{\theta}_\tau^{(1)}, ..., \boldsymbol{\theta}_\tau^{(B_1)}\}$ of the posterior distribution $\pi(\boldsymbol{\theta}_\tau \mid D_\tau)$, the predictive density (7) can be approximated by,

$$f(\tau \mid D_\tau) \simeq \frac{1}{B_1} \sum_{k=1}^{B_1} f\left(\tau \mid \boldsymbol{\theta}_\tau^{(k)}\right), \tag{8}$$

where $f(\tau \mid \boldsymbol{\theta}_\tau^{(k)})$ is the density function (2) given the k-th set of parameters, $\boldsymbol{\theta}_\tau^{(k)}$, of the MCMC sample.

Now, we define a suitable prior distribution for $\boldsymbol{\theta}_\tau$ and describe an MCMC algorithm that can be used to sample from the posterior distribution. Firstly, we reparametrize the rates, $\boldsymbol{\lambda}$, as follows,

$$\lambda_r = \lambda_1 \upsilon_2 ... \upsilon_r, \qquad \text{where } 0 < \upsilon_s \leq 1, \text{ for } r, s = 2, ..., L.$$

This reparameterization facilitates the derivation of a noninformative prior distribution and a straightforward implementation of the MCMC algorithm. This kind of reparameterization has also been considered in Robert and Mengersen (1999) for normal mixtures, and in Gruet et al. (1999) for exponential mixtures.

We now define the following non-informative prior distribution for the Coxian model parameters,

$$\begin{aligned} L &\sim \textit{Uniform}\,(0, 20) \\ \mathbf{P} &\sim \textit{Dirichlet}\,(1, ..., 1)\,, \\ \pi(\lambda_1) &\propto \frac{1}{\lambda_1} \\ \upsilon_r &\sim \textit{Uniform}(0,1), \quad \text{for } r = 1, ..., L. \end{aligned}$$

This prior choice allows for the approximation of long-tailed distribution because no strong assumptions are imposed on the size of the first mixture rate, λ_1, and then, the mean of the remaining mixture components can take values as large or as small as required. Note that the joint prior distribution is improper, although it can be shown that it leads to a proper posterior distribution extending the arguments given in Gruet et al. (1999).

Next, we construct an MCMC algorithm in order to obtain a sample from the posterior distribution of $\boldsymbol{\theta}_\tau = (L, \mathbf{P}, \lambda_1, \boldsymbol{\upsilon})$. This can be carried out by cycling repeatedly through draws of each parameter conditional on the remaining parameters. Thus, we need to be able to sample from the conditional posterior distribution of each parameter. This is facilitated in mixture models with a data augmentation procedure, see for example Richardson and Green (1997), where each inter-occurrence time, τ_j, is assumed to arise from a specific but unknown mixture component, z_j, which is introduced as a missing observation, for $j = 1, ..., n$. Given the missing data, $\mathbf{z} = (z_1, ..., z_n)$, the MCMC algorithm has the following scheme:

`MCMC algorithm`

1. `Set initial values` $\boldsymbol{\theta}_{\tau}^{(0)} = \left(L^{(0)}, \mathbf{P}^{(0)}, \lambda_1^{(0)}, \boldsymbol{\upsilon}^{(0)}\right)$.
2. `Update` $\mathbf{z}$ `by sampling from` $\mathbf{z}^{(k+1)} \sim \mathbf{z} \mid D_{\tau}, L^{(k)}, \mathbf{P}^{(k)}, \lambda_1^{(k)}, \boldsymbol{\upsilon}^{(k)}$.
3. `Update` $\mathbf{P}$ `by sampling from` $\mathbf{P}^{(k+1)} \sim \mathbf{P} \mid D_{\tau}, \mathbf{z}^{(k+1)}, L^{(k)}$.
4. `Update` λ_1 `by sampling from` $\lambda_1^{(k+1)} \sim \lambda_1 \mid D_{\tau}, \mathbf{z}^{(k+1)}, L^{(k)}, \boldsymbol{\upsilon}^{(k)}$.
5. `For` $r = 1, ..., L^{(k)}$:

   `Update` υ_r `by sampling from` $\upsilon_r^{(k+1)} \sim \upsilon_r \mid D_{\tau}, \mathbf{z}^{(k+1)}, L^{(k)}, \lambda_1^{(k+1)}, \upsilon_1^{(k+1)}, ..., \upsilon_{r-1}^{(k+1)}, \upsilon_{r+1}^{(k)}, ..., \upsilon_{L^{(k)}}^{(k)}$.
6. `Update` L `by sampling from` $L^{(k+1)} \sim L \mid D_{\tau}, \mathbf{z}^{(k+1)}, \mathbf{P}^{(k+1)}, \lambda_1^{(k+1)}, \boldsymbol{\upsilon}^{(k+1)}$.
7. $k = k + 1$. `Go to 2.`

In step 2, we sample from the conditional posterior distribution of $\mathbf{z}$ which is given by, for $j = 1, ..., n$,

$$\Pr\left(z_j = r \mid \tau_j, L, \mathbf{P}, \lambda_1, \boldsymbol{\upsilon}\right) \propto P_r f_r\left(\tau_j \mid \lambda_1, \upsilon_2, ..., \upsilon_r\right), \qquad \text{for } r = 1, ..., L,$$

where $f_r\left(\tau \mid \lambda_1, \upsilon_2, ..., \upsilon_r\right)$ denotes the reparametrized Coxian density $f_r(\tau \mid \lambda_1, \lambda_2, .., \lambda_r)$ given in (3).

In step 3, we sample from the conditional posterior distribution for the mixture weights which can be shown to be given by,

$$\mathbf{P} \mid \tau, \mathbf{z}, L \sim \textit{Dirichlet}(1 + n_1, ..., 1 + n_L),$$

where n_r is the number of observations assigned to the r-th mixture component, for $r = 1, ..., L$.

In step 4, we sample from the conditional posterior distributions of λ_1 whose density functions can be evaluated up to the integration constants as,

$$\pi\left(\lambda_1 \mid D_{\tau}, \mathbf{z}, L, \boldsymbol{\upsilon}\right) \propto \prod_{j=1}^{n} f_{z_j}\left(\tau \mid \lambda_1, \upsilon_2, ..., \upsilon_{z_j}\right) \pi\left(\lambda_1\right). \tag{9}$$

Although we can not sample directly from this posterior distribution, we can make use of the Metropolis Hastings method, see Hastings (1970), using a gamma candidate distribution. We generate a candidate $\tilde{\lambda}_1 \sim G\left(2, 2/\lambda^{(k)}\right)$ which is accepted with probability,

$$\min\left\{1, \frac{\pi\left(\tilde{\lambda}_1 | \cdots\right) G\left(\lambda_1^{(k)} \mid \tilde{\lambda}_1\right)}{\pi\left(\lambda_1^{(k)} | \cdots\right) G\left(\tilde{\lambda}_1 \mid \lambda_1^{(k)}\right)}\right\}$$

where $\pi(\tilde{\lambda}_1 | \cdots)$ is given in (9) and $G(\tilde{\lambda}_1 \mid \lambda_1^{(k)})$ is the gamma density used to generate $\tilde{\lambda}_1$.

In step 5, we sample from the conditional posterior distributions,

$$\pi\left(\upsilon_r \mid D_\tau, \mathbf{z}, L, \lambda_1, \boldsymbol{\upsilon}_{-r}\right) \propto \prod_{j=1, z_j \geq r}^{n} f_{z_j}\left(\tau_j \mid \lambda_1, \upsilon_2, ..., \upsilon_{z_j}\right) \pi\left(\upsilon_r\right), \qquad \text{for } r = 2, ..., L,$$

where $\boldsymbol{\upsilon}_{-r} = \left(\upsilon_1, ..., \upsilon_{r-1}, \upsilon_{r+1}, ..., \upsilon_L\right)$. As before, we can make use of a Metropolis Hastings algorithm using a beta candidate distribution to sample from this distribution.

In step 6, we can not either sample directly from the conditional posterior distribution of the mixture size, L. However, we can generate values from this distribution by using the reversible jump methods introduced for normal mixture models by Richardson and Green (1997). This procedure is a generalization of the Metropolis Hastings algorithms for variable dimension parametric spaces, where candidate values are proposed to change the number of mixture components from L to $L \pm 1$. We consider the so called split and combine moves where one mixture component, r , is split into two adjacent components, (r_1, r_2). We consider analogous movements to the proposed in Gruet et al. (1999) for exponential mixtures.

The MCMC algorithm generates values from a Markov chain whose stationary distribution is the joint posterior distribution of interest. Thus, in order to reach the equilibrium, we generate B_0 *burnin* iterations, which will be discarded, followed by another B_1 iterations "in equilibrium" that will be used for the inference. In the real application we have set $B_0 = B_1 = 10\ 000$. Given the MCMC sample of size B_1, $\{\boldsymbol{\theta}_\tau^{(1)}, ..., \boldsymbol{\theta}_\tau^{(B_1)}\}$, we can now approximate the predictive density, $f\left(\tau \mid D_\tau\right)$, of the inter-arrival time between claims, τ, using the approximation (8) based on the sample posterior densities $\{f(\tau \mid \boldsymbol{\theta}_\tau^{(1)}), ..., f(\tau \mid \boldsymbol{\theta}_\tau^{(B_1)})\}$. Also the posterior median and 95% credible intervals for the density can be obtained by just calculating the median and the 0.025 and 0.975 quantiles of this posterior sample, respectively. Analogously, we can approximate the posterior mean, median and confidence intervals for the cumulated distribution function, $F\left(\tau \mid D_\tau\right)$.

Now, we consider Bayesian inference for the claim size variable, Y, which has been assumed to follow a Coxian distribution with parameters, $\boldsymbol{\theta}_Y = (M, \mathbf{Q}, \boldsymbol{\mu})$, given the data sample of n possibly left-censored claim sizes, $D_Y = \{(X_1, \delta_1), \ldots, (X_n, \delta_n)\}$, as described in the introduction.

Censoring can be easily incorporated in an MCMC algorithm by using a data augmentation method as follows. A new set of missing latent variables, $\mathbf{y} = (Y_1, ..., Y_n)$, is introduced such that $Y_j = X_j$ if $\delta_j = 1$, and Y_j follows a Coxian distribution with parameters $\boldsymbol{\theta}_Y$ truncated to $Y_j < C_j$ if $\delta_j = 0$. These missing data set are considered as a new set of parameters that are updated in each iteration of the previous MCMC algorithm by including a new step before step 2. In this new step, the missing values Y_j with $\delta_j = 0$ are simulated from a Coxian random variable conditioned to be less than C_j, for $j = 1, ..., n$. Given the completed data in each iteration, the remaining steps of the MCMC algorithm do not change as the conditional posterior distributions of the remaining parameters are the same as before.

Given the MCMC sample of size B_1 from the joint posterior distribution of $\boldsymbol{\theta}_Y$, we can estimate the predictive density associated to the claim size using,

$$f(y \mid D_Y) \simeq \frac{1}{B_1} \sum_{k=1}^{B_1} f\left(y \mid \boldsymbol{\theta}_Y^{(k)}\right), \tag{10}$$

where $f(y \mid \boldsymbol{\theta}_\tau^{(k)})$ is given in (4). Analogously, the cumulative distribution $F(y \mid D_Y)$ can be estimated.

4 Estimation of the claim count and claim amount distribution

In this section, we are interested in the estimation of the total claim count and the total claim amount in a future time period given past $data = \{D_\tau, D_Y\}$ on claim inter-arrivals and claim sizes. For simplicity of notation, we reset to zero the initial time of the future period such that we are concerned with the distributions of $N(t)$ and $S(t)$. A Bayesian estimation of these distributions could be obtained by calculating the posterior means of their cumulative distributions, also called their predictive cumulative distribution functions,

$$\Pr(N(t) \leq m \mid data) = \int_\Theta \Pr(N(t) \leq m \mid \boldsymbol{\theta})\, \pi(\boldsymbol{\theta} \mid data)\, d\boldsymbol{\theta}, \tag{11}$$

and,

$$\Pr(S(t) \leq x \mid data) = \int_\Theta \Pr(S(t) \leq x \mid \boldsymbol{\theta})\, \pi(\boldsymbol{\theta} \mid data)\, d\boldsymbol{\theta}, \tag{12}$$

where $\boldsymbol{\theta} = \{\boldsymbol{\theta}_\tau, \boldsymbol{\theta}_Y\}$ are the model parameters of the Coxian distributions of τ and Y, and $\pi(\boldsymbol{\theta} \mid data)$ is the joint posterior distribution of these parameters given the observed data. Clearly, we do not have a explicit expression of these predictive distributions, but we can make use of the MCMC sample simulated from the posterior distribution, $\pi(\boldsymbol{\theta} \mid data)$, in the previous section, and approximate (11) and (12) by,

$$\Pr(N(t) \leq m \mid data) \simeq \frac{1}{B_1} \sum_{k=1}^{B_1} \Pr\left(N(t) \leq m \mid \boldsymbol{\theta}^{(k)}\right), \tag{13}$$

and,

$$\Pr(S(t) \leq x \mid data) \simeq \frac{1}{B_1} \sum_{k=1}^{B_1} \Pr\left(S(t) \leq x \mid \boldsymbol{\theta}^{(k)}\right), \tag{14}$$

respectively. In order to obtain these approximations, we also need explicit expression of the distributions of $N(t)$ and $S(t)$ when the model parameters are known. That is, we need to know the value of the probabilities $\Pr(N(t) \leq m \mid \boldsymbol{\theta})$ and $\Pr(S(t) \leq u \mid \boldsymbol{\theta})$, given a set of fixed Coxian parameters of the distributions of the inter-arrival claim time, τ, and the claim size, Y. Although these probabilities are not known directly, we can obtain closed expressions of their Laplace transforms, which can be numerically inverted using for example the Euler algorithm, see Abate and Whitt (1992), which is quite fast an accurate.

Let us first consider how to obtain the Laplace transform of the probability $\Pr\left(N\left(t\right) \leq m \mid \boldsymbol{\theta}\right)$. Note that we assume now that the inter-arrival parameters $\boldsymbol{\theta}_\tau = (L, \mathbf{P}, \boldsymbol{\lambda})$ are fixed. It is well known that in a renewal process, as the considered claim arrival process, we have that, see e.g. Rolski et al. (1999),

$$\Pr\left(N\left(t\right) \geq m \mid \boldsymbol{\theta}_\tau\right) = \Pr\left(\sum_{j=1}^{m} \tau_j \leq t \mid \boldsymbol{\theta}_\tau\right), \tag{15}$$

The Laplace transform of the inter-arrival time, τ, which follows a Coxian distribution with parameters $\boldsymbol{\theta}_\tau = (L, \mathbf{P}, \boldsymbol{\lambda})$, is given by,

$$f_\tau^*\left(s \mid \boldsymbol{\theta}_\tau\right) = E\left[e^{-s\tau} \mid \boldsymbol{\theta}_\tau\right] = \sum_{r=1}^{L} P_r \prod_{i=1}^{r} \left(\frac{\lambda_i}{\lambda_i + s}\right). \tag{16}$$

Then, the Laplace transform of the variable $\sum_{j=1}^{m} \tau_j$, which is a sum of m Coxian variables, is the product of the m Laplace transform of each Coxian distribution, that is,

$$f_{\Sigma\tau}^*\left(s \mid \boldsymbol{\theta}_\tau\right) = E\left[e^{-s\sum_{j=1}^{m}\tau_j} \mid \boldsymbol{\theta}_\tau\right] = \left[\sum_{r=1}^{L} P_r \prod_{i=1}^{r} \left(\frac{\lambda_i}{\lambda_i + s}\right)\right]^m.$$

And the Laplace transform of the cumulated distribution function of $\sum_{j=1}^{m}\tau_j$ is given by,

$$F_{\Sigma\tau}^*\left(s \mid \boldsymbol{\theta}_\tau\right) = \int_0^\infty e^{-st} \Pr\left(\sum_{j=1}^{m} \tau_j \leq t \mid \boldsymbol{\theta}_\tau\right) dt = \frac{1}{s}\left[\sum_{r=1}^{L} P_r \prod_{i=1}^{r} \left(\frac{\lambda_i}{\lambda_i + s}\right)\right]^m.$$

Finally, using the relation (15), we obtain that the Laplace transform of the probability of interest is given by,

$$\int_0^\infty e^{-st} \Pr\left(N\left(t\right) \leq m \mid \boldsymbol{\theta}_\tau\right) dt = \frac{1}{s}\left(1 - \left[\sum_{r=1}^{L} P_r \prod_{i=1}^{r} \left(\frac{\lambda_i}{\lambda_i + s}\right)\right]^{m+1}\right).$$

Then, given t, this Laplace transform can be numerically inverted for $m = 0, 1, 2, ...$ to obtain the probability $\Pr(N\left(t\right) \leq m \mid \boldsymbol{\theta}^{(k)})$ for each value of the Coxian parameters $\boldsymbol{\theta}^{(k)}$ in the MCMC sample such that we can evaluate the approximated predictive probabilities given in (13).

Using the obtained probabilities, $\Pr(N\left(t\right) = m \mid \boldsymbol{\theta}^{(k)})$, it is also possible to approximate the predictive mean of $N(t)$ by,

$$E[N(t) \mid data] \simeq \frac{1}{B_1}\sum_{k=1}^{B_1} E\left[N(t) \mid \boldsymbol{\theta}^{(k)}\right] = \frac{1}{B_1}\sum_{k=1}^{B_1}\sum_{m=0}^{\infty} m \Pr\left(N(t) = m \mid \boldsymbol{\theta}^{(k)}\right). \tag{17}$$

Furthermore, we can obtain a 95% predictive interval for the estimated mean by just calculating the 0.025 and 0.975 quantiles of the posterior sample of means, $\{E[N(t) \mid \boldsymbol{\theta}^{(1)}], ..., E[N(t) \mid \boldsymbol{\theta}^{(B_1)}]\}$. Using an analogous approach, we can estimate other characteristic measures of $N(t)$ such as the variance, median, quantiles, etc., together with their predictive intervals. Note that, in practice, we must truncate the infinite sum in (17) up to a finite value m_0 such that $P(N(t) \geq m_0 \mid \boldsymbol{\theta}^{(k)})$ is very small.

Now, we consider how to obtain the Laplace transform of the probability $\Pr(S(t) \leq x \mid \boldsymbol{\theta})$. Note that we assume now that the inter-arrival parameters, $\boldsymbol{\theta}_\tau = (L, \mathbf{P}, \boldsymbol{\lambda})$, and the claim size parameters, $\boldsymbol{\theta}_Y = (M, \mathbf{Q}, \boldsymbol{\mu})$, are fixed. It can be shown that the Laplace transform of an aggregate loss random variable, such as $S(t)$, is given by, see e.g. Rolski et al. (1999),

$$f^*_{S(t)}(s \mid \boldsymbol{\theta}) = g^*_{N(t)}\left[f^*_Y(s \mid \boldsymbol{\theta}_Y) \mid \boldsymbol{\theta}_\tau\right], \tag{18}$$

where $g^*_{N(t)}[s]$ is the probability generating function of the claim count random variable $N(t)$ and $f^*_Y(s \mid \boldsymbol{\theta}_Y)$ is the Laplace transform of the claim size, Y, which follows a Coxian distribution with parameters $\boldsymbol{\theta}_Y = (M, \mathbf{Q}, \boldsymbol{\mu})$, and is given by,

$$f^*_Y(s \mid \boldsymbol{\theta}_Y) = E\left[e^{-sY} \mid \boldsymbol{\theta}_Y\right] = \sum_{r=1}^{M} Q_r \prod_{t=1}^{r}\left(\frac{\mu_t}{\mu_t + s}\right). \tag{19}$$

Then, from (18) we obtain that,

$$\int_0^\infty e^{-sx} \Pr(S(t) \leq x \mid \boldsymbol{\theta})\, dx = \frac{1}{s}\sum_{m=0}^{\infty}\left[\sum_{r=1}^{M} Q_r \prod_{t=1}^{r}\left(\frac{\mu_t}{\mu_t + s}\right)\right]^m \Pr(N(t) = m). \tag{20}$$

Thus, given t and the probability distribution of $N(t)$ obtained previously, this Laplace transform can be numerically inverted in order to obtain the probability $\Pr(S(t) \leq x \mid \boldsymbol{\theta}^{(k)})$ for each MCMC iteration and use the approximation given in (14). Note that, in practice, we must truncate the infinite sum in (20) up to the previously chosen finite value m_0 such that $P\left(N(t) \geq m_0 \mid \boldsymbol{\theta}^{(k)}\right)$ is very small.

We can also obtain estimations of the mean and variance of $S(t)$ using the following known relationships, see e.g. Rolski et al. (1999),

$$\begin{aligned} E\left[S\left(t\right)\mid \boldsymbol{\theta}\right] &= E\left[N\left(t\right)\mid \boldsymbol{\theta}_{\tau}\right]\times E\left[Y\mid \boldsymbol{\theta}_{Y}\right], \\ V\left[S\left(t\right)\mid \boldsymbol{\theta}\right] &= E\left[N\left(t\right)\mid \boldsymbol{\theta}_{\tau}\right]\times V\left[Y\mid \boldsymbol{\theta}_{Y}\right]+E\left[Y\mid \boldsymbol{\theta}_{Y}\right]^{2}V\left[N\left(t\right)\mid \boldsymbol{\theta}_{\tau}\right], \end{aligned} \tag{21}$$

which can be obtained explicitly considering that the claim size random variable, Y, follows a Coxian distribution a Coxian distribution with parameters $\boldsymbol{\theta}_Y = (M, \mathbf{Q}, \boldsymbol{\mu})$, and then,

$$\begin{aligned} E\left[Y\mid \boldsymbol{\theta}_{Y}\right] &= \sum_{r=1}^{M} Q_r \sum_{s=1}^{r} \frac{1}{\mu_s}, \\ E\left[Y^2\mid \boldsymbol{\theta}_{Y}\right] &= \sum_{r=1}^{M} Q_r \left[\sum_{s=1}^{r} \frac{2}{\mu_s^2} + 2\sum_{s\neq t}^{r} \frac{1}{\mu_s\mu_t}\right]. \end{aligned}$$

Then, the predictive mean of $S\left(t\right)$ can be approximated by,

$$E\left[S\left(t\right)\mid data\right] \simeq \frac{1}{B_1}\sum_{k=1}^{B_1} E\left[S\left(t\right)\mid \boldsymbol{\theta}^{(k)}\right] = \frac{1}{B_1}\sum_{k=1}^{B_1} E\left[N\left(t\right)\mid \boldsymbol{\theta}_{\tau}^{(k)}\right]\times E\left[S\left(t\right)\mid \boldsymbol{\theta}_{Y}^{(k)}\right].$$

As before, we can also obtain predictive intervals for the mean of $S\left(t\right)$ using the percentiles of the predictive sample of means, $\{E[S\left(t\right)\mid \boldsymbol{\theta}^{(1)}], ..., E[S\left(t\right)\mid \boldsymbol{\theta}^{(B_1)}]\}$, and analogously, we can estimate the other characteristic measures of $S\left(t\right)$ such as the variance, median, quantiles, etc., together with their predictive intervals.

5 Estimation under deductibles and maximum limits

In this section, we consider the estimation of the total claim amount distribution when claims are are subject to deductibles and limits. This is a more realistic situation in practice since most insurance contracts contain this kind of clauses. In these cases, the insurer will not pay those losses which are smaller than a previously fixed amount, which is called the deductible, and this amount will be deducted from all payments. Further, a maximum amount, called the limit, is also predetermined in the policy such that the insurer will not pay more than this limit amount minus the deductible. Then, for each claim size, Y, we have the following layer representing the loss from an excess-of-loss cover,

$$\tilde{Y} = \begin{cases} 0, & \text{if } 0 < Y < a, \\ Y - a, & \text{if } a \leq Y < b, \\ b - a, & \text{if } b \leq Y \leq \infty, \end{cases} \tag{22}$$

where a is the deductible, also called the attachment point, and b is the limit, see e.g. Klugman et al. (2004). Thus, the interest is now focussed on the estimation of the following aggregate claim amount,

$$\tilde{S}(t) = \sum_{j=1}^{N(t)} \tilde{Y}_j,$$

where $\tilde{Y}_j$ is obtained from the j-th claim size, Y_j, according to the relation (22).

A Bayesian estimation of the distribution of $\tilde{S}(t)$ can be obtained using a similar approach to the described in the previous section as follows. Firstly, assume that the Coxian inter-arrival parameters, $\boldsymbol{\theta}_\tau = (L, \mathbf{P}, \boldsymbol{\lambda})$, and the claim size parameters, $\boldsymbol{\theta}_Y = (M, \mathbf{Q}, \boldsymbol{\mu})$, are fixed. Note that analogously to (18), the Laplace transform of $\tilde{S}(t)$, is given by,

$$f^*_{\tilde{S}(t)}(s \mid \boldsymbol{\theta}) = g^*_{N(t)}\left[f^*_{\tilde{Y}}(s \mid \boldsymbol{\theta}_Y) \mid \boldsymbol{\theta}_\tau\right], \tag{23}$$

where the Laplace transform of the variable $\tilde{Y}$, defined in (22), which can be obtained using the Coxian density (4) as follows,

$$\begin{aligned}
f^*_{\tilde{Y}}(s \mid \boldsymbol{\theta}_Y) &= \int_0^\infty e^{-s\tilde{y}} f(\tilde{y} \mid \boldsymbol{\theta}_Y)\, d\tilde{y} \\
&= e^{s0} \Pr(Y < a \mid \boldsymbol{\theta}_Y) + e^{sa} \int_a^b e^{-sy} f(y \mid \boldsymbol{\theta}_Y)\, dy + e^{-s(b-a)} \Pr(Y > b \mid \boldsymbol{\theta}_Y) \\
&= \sum_{r=1}^{M} Q_r \sum_{t=1}^{r} C_{tr} \left[\int_0^a \mu_t e^{-\mu_t y} dy + e^{sa} \int_a^b \mu_t e^{-(s+\mu_t)y} dy + e^{-s(b-a)} \int_b^\infty \mu_t e^{-\mu_t y} dy\right] \\
&= \sum_{r=1}^{M} Q_r \sum_{t=1}^{r} C_{tr} \left[1 + \left(e^{-\mu_t a} - e^{-\mu_t b-(b-a)s}\right)\left(\frac{\mu_t}{\mu_{t+s}} - 1\right)\right],
\end{aligned}$$

where C_{tr} are the coefficients given in (5). Then, we can now invert the Laplace transform of $\tilde{S}(t)$ for each set of the parameters, $\boldsymbol{\theta}^{(k)}$, in the MCMC sample,

$$\int_0^\infty e^{-sx} \Pr\left(\tilde{S}(t) \le x \mid \boldsymbol{\theta}^{(k)}\right) dx = \frac{1}{s} \sum_{m=0}^{\infty} \left[f^*_{\tilde{Y}}\left(s \mid \boldsymbol{\theta}^{(k)}\right)\right]^m \Pr\left(N(t) = m \mid \boldsymbol{\theta}^{(k)}\right),$$

and estimate the predictive distribution of $\tilde{S}(t)$ using the following montecarlo approximation as usual,

$$\Pr\left(\tilde{S}(t) \le x \mid data\right) \simeq \frac{1}{B_1} \sum_{k=1}^{B_1} \Pr\left(\tilde{S}(t) \le x \mid \boldsymbol{\theta}^{(k)}\right). \tag{24}$$

As in the previous section, we can also estimate the main characteristics of $\tilde{S}(t)$ such as the mean, variance, quantiles, etc. using the mean of their values for each set of parameters in the MCMC sample. In particular, we use the following formulae analogous to (21) to obtain the mean and variance for each $\boldsymbol{\theta}$,

$$
\begin{aligned}
E\left[\tilde{S}(t) \mid \boldsymbol{\theta}\right] &= E\left[N(t) \mid \boldsymbol{\theta}_\tau\right] \times E\left[\tilde{Y} \mid \boldsymbol{\theta}_Y\right], \\
V\left[\tilde{S}(t) \mid \boldsymbol{\theta}\right] &= E\left[N(t) \mid \boldsymbol{\theta}_\tau\right] \times V\left[\tilde{Y} \mid \boldsymbol{\theta}_Y\right] + E\left[\tilde{Y} \mid \boldsymbol{\theta}_Y\right]^2 V\left[N(t) \mid \boldsymbol{\theta}_\tau\right],
\end{aligned}
$$

where,

$$
\begin{aligned}
E\left[\tilde{Y} \mid \boldsymbol{\theta}_Y\right] &= E\left[Y - a \mid a < Y < b, \boldsymbol{\theta}_Y\right] \Pr\left(a < Y < b \mid \boldsymbol{\theta}_Y\right) + (b-a)\Pr\left(Y > b \mid \boldsymbol{\theta}_Y\right) \\
&= \sum_{r=1}^{M} Q_r \sum_{r=1}^{M} C_{tr} \left[\int_a^b y\mu_t e^{-\mu_t y} dy - a\int_a^b \mu_t e^{-\mu_t y} dy + (b-a)\int_b^\infty \mu_t e^{-\mu_t y} dy\right] \\
&= \sum_{r=1}^{M} Q_r \sum_{r=1}^{M} C_{tr} \left[\frac{e^{-\mu_t a} - e^{-\mu_t b}}{\mu_t}\right]
\end{aligned}
$$

and,

$$
\begin{aligned}
E\left[\tilde{Y}^2 \mid \theta_Y\right] &= E\left[(Y-a)^2 \mid a < Y < b, \theta_Y\right] P\left(a < Y < b \mid \theta_Y\right) + (b-a)^2 P\left(Y > b \mid \theta_Y\right) \\
&= \sum_{r=1}^{M} Q_r \sum_{r=1}^{M} C_{tr} \left[\int_a^b (y-a)^2 \mu_t e^{-\mu_t y} dy + (b-a)^2 \int_b^\infty \mu_t e^{-\mu_t y} dy\right] \\
&= \sum_{r=1}^{M} Q_r \sum_{r=1}^{M} C_{tr} \left[\frac{2e^{-\mu_t a}}{\mu_t^2} - \frac{2e^{-\mu_t b}}{\mu_t^2}\left(1 + (b-a)\mu_t\right)\right].
\end{aligned}
$$

Finally, note that the same procedure can be considered for the estimation of the distribution, main characteristics and confidence intervals for the total claim amount with alternative layer specifications to the given in (22) such as,

$$
\tilde{Y} = \begin{cases} Y, & \text{if } 0 < Y < a, \\ a, & \text{if } a \leq Y < \infty, \end{cases}
$$

which may be of the interest of the insured customer, or,

$$
\tilde{Y} = \begin{cases} 0, & \text{if } 0 < Y < b, \\ Y - b, & \text{if } b \leq Y < \infty, \end{cases}
$$

which may be useful for the reinsurance company.

6 Application to real data

In this section, we illustrate our methodology with a real data set provided by the insurance department of an international company. A large data base was collected containing the dates and amounts of claims in different sectors of activity in the company, such as commerce, transportation, public liability, etc. Claim sizes presented left-censoring because those values smaller than previously fixed deductibles were not recorded by the insurer company as it was not responsible of their payment. To preserve confidentiality, these original data have been rescaled (multiplied by a constant) in this section.

We show here the results concerning two sectors of activity, namely, sector C and sector T, using two random subsamples from the original data. The sample of sector C contains 600 observations, with 200 left-censored claim sizes, and the sample of sector T is given by 400 observations, with 100 left-censored claim sizes, both observed during a past time interval. Assume, for example, that the company is interested in making predictions for a future time period whose length is given by 200 units of time. Thus, we wish to estimate the distributions of the total number of claims, $N(200)$, and the total claim amount, $S(200)$, in this future time period for each sector separately and for the two sectors jointly. In the next subsections, we firstly present the results obtained for sector C and then, the results for sectors C and T jointly, which have been called the global sector.

6.1 Analysis of sector C

From this sector, we have a sample of 600 complete inter-ocurrence times and 600 left-censored claim sizes. Table 1 shows some summary statistics of these data. Note that the statitistics for the censored claim size variable, Y, have been calculated using the Kaplan-Meier weights.

	τ	Y
Mean	0.8381	2059.71
Median	0.5602	641.90
Std. Deviation	0.9892	4390.48
Skewness	2.6269	5.158
Kurtosis	11.7487	33.910
Percentile 95	2.8661	8 228.83
Percentile 99	4.3187	26 563.05

Table 1: Summary statistics for the sample of inter-claim times, τ, and claim sizes, Y, in sector C.

Firstly, we consider the sample of 600 complete inter-ocurrence times, τ, observed in sector C. The MCMC algorithm introduced in Section 3 is run with $B_0 = 10\ 000$ burn-in iterations and $B_1 = 10\ 000$ iterations "in equilibrium". To assess the convergence of the Markov chain, we use the convergence diagnostic proposed in

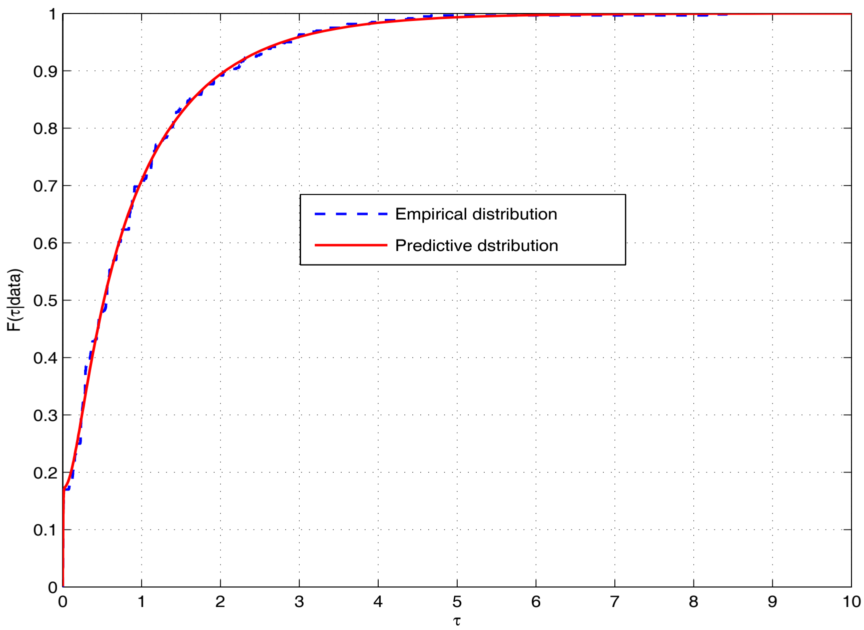

Figure 2: Empirical (dotted) and Bayesian estimation (solid) of the cumulative distribution function of the inter-ocurrence time between claims in sector C.

Geweke (1992). Figure 2 illustrates the empirical and the Bayesian estimation of the cumulative distribution function of the inter-ocurrence time between claims in sector C.

Now, we analyze the claim size distribution using the sample of 600 claim sizes, with 200 left-censored data, observed in sector C. We run the MCMC algorithm described in Section 3 for left-censored data using the same number of iterations as before and checking the convergence with the Geweke's statistic. Figure 3 shows the empirical and Bayesian estimation of the cumulative distribution function for the claim size in sector C. The empirical distribution have been obtained using the usual Kaplan-Meier estimator for right-censored data after multiplying each observation by -1 plus the maximum of the data sample.

Next, we develop Bayesian prediction for the total claim count and the total amount random variables in a future time period. Using the MCMC output and following the approach described in Section 4, we firstly estimate the cumulative distribution function of the total number of claims, $N(200)$, that will occur in sector C up to time $t = 200$. Figure 4 shows the posterior mean, obtained with (13), the posterior median and 95% predictive intervals, obtained as described in Section 4. Observe that the predictive intervals are quite symmetric for the values of m which are close to the mean of the variable, and are left and right-skewed for values of m that are quite smaller and larger, respectively, than the mean of $N(200)$. Nevertheless, their amplitudes are not very wide, meaning that the estimated probabilities are fairly accurate.

Table 2 shows the posterior means and 95% credible intervals for the main characteristic measures

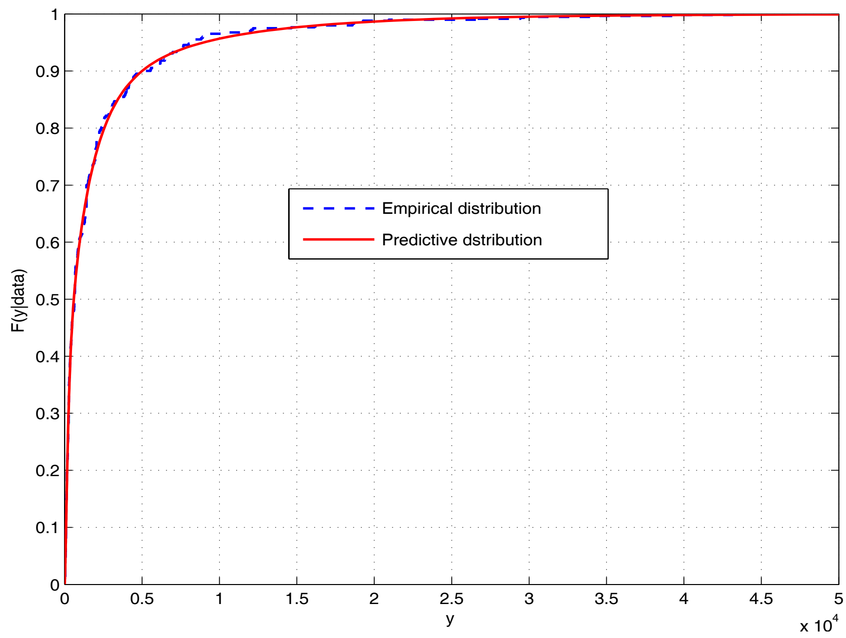

Figure 3: Empirical (dotted) and Bayesian estimation (solid) of the cumulative distribution function for the claim sizes in sector C.

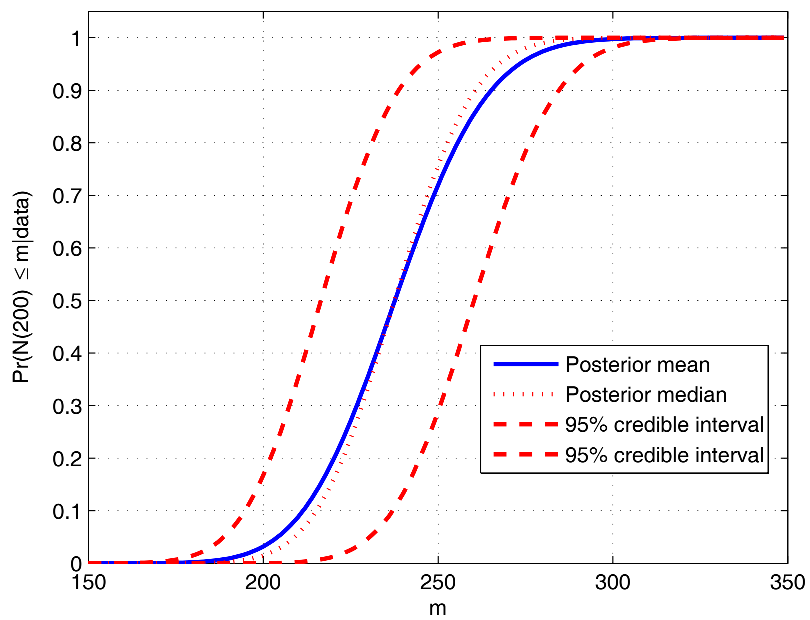

Figure 4: Posterior mean, median and 95% credible interval for the cumulative distribution function of the number of claims, $N(200)$, up to time $t = 200$ in sector C.

associated with $N\,(200)$ using the proposed Bayesian approach described in Section 4. These results are compared with those obtained with a nonparametric approach developed Gonález-Fragueiro et al. (2006) using the same real data set. Note that the results are comparable and the Bayesian credible intervals always include the nonparametric point estimations and viceversa.

Measure	BY estimation	95% BY interval		NP estimation	95% NP interval	
Mean	238.53	217.17	260.99	231.72	208.73	251.17
Median	238.31	217.00	261.00	231.00	207.00	250.00
Std. dev.	17.81	16.25	19.63	17.37	15.67	19.21
0.95 quantile	268.19	246.00	292.00	261.00	238.00	282.00
0.99 quantile	280.91	258.00	305.00	272.00	247.00	293.00

Table 2: Bayesian estimates (BY) and 95% credible intervals of some characteristic measures of $N\,(200)$, compared with the equivalent nonparametric (NP) point estimates and 95% confidence intervals in sector C.

Now, we present the results obtained for the total claim amount random variable. Figure 5 illustrates the Bayesian estimations of the cumulative distribution function of the total claim amount, $S\,(200)$, that will be paid in sector C in the next 200 units of time. Again we obtain the posterior mean, obtained with (14), the posterior median and 95% credible intervals as described in Section 4. As before, the credible intervals are quite symmetric when x is close to the mean of $S\,(200)$, and left and right-skewed when x is rather smaller and larger, respectively, than it. However, as before, there is not a large uncertainty in the estimated probabilities.

Table 3 shows the Bayesian (BY) and nonparametric (NP) point estimations and 95% confidence intervals for the main characteristic measures associated with $S\,(200)$. Estimation results are again comparable.

Measure	BY estimation	95% BY interval		NP Estimation	95% NP interval	
Mean	496 818.42	406 103.26	605 286.56	539 260.85	344 325.73	609 162.04
Median	492 454.10	402 153.98	599 364.54	532 485.13	347 352.02	604 282.41
Std. dev.	77 860.13	61 817.98	102 269.34	96 871.05	27 508.60	110 454.63
0.95 quantile	631 955.73	516 584.37	775 230.65	707 480.58	392 685.38	788 830.90
0.99 quantile	696 602.77	569 205.56	858 758.21	794 422.38	393 147.98	887 291.18

Table 3: Bayesian estimates (BY) and 95% credible intervals of some characteristic measures of $S\,(200)$, compared with the equivalent nonparametric (NP) point estimates and 95% confidence intervals in sector C.

Finally, we analyze the same problem when the coverage is restricted by a deductible and maximum limit. Assume for example that the insurance policy for sector C includes a deductible amount of $a = 12\ 000$ and a maximum limit of $b = 16\ 000$ monetary units. Then, we can estimate the distribution of total claim amount $\tilde{S}(t)$ as described in Section 5. Figure 6 illustrates the posterior mean, obtained with (24), the posterior median and 95% credible intervals for the cumulative distribution function of $\tilde{S}(t)$. Table 4 shows

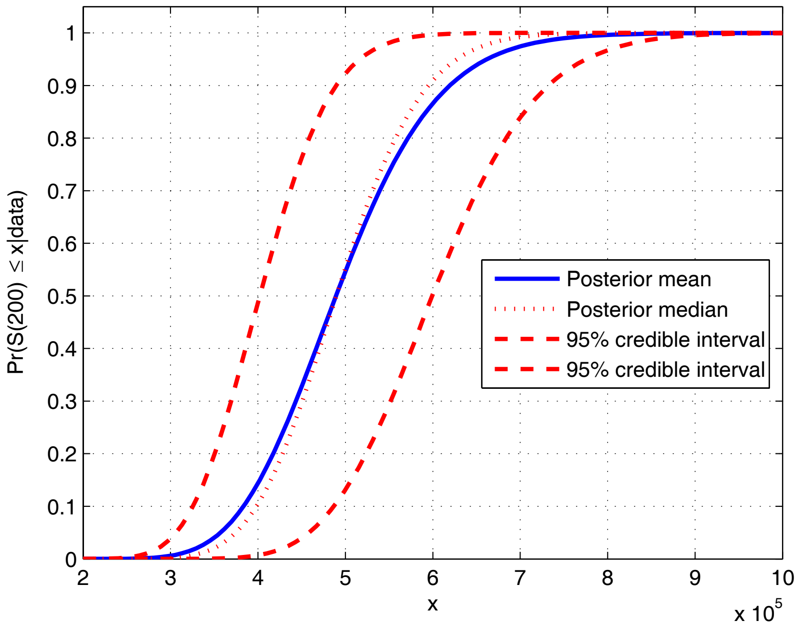

Figure 5: Posterior mean, median and 95% credible interval for the cumulative distribution function of the total claim amount, $S(200)$, up to time $t = 200$ in sector C.

the estimated characteristics of this distribution with 95% credible intervals. Again in this case, these are we compared with the estimations obtained in Gonález-Fragueiro et al. (2006) leading to similar results.

Measure	BY estimation	95% BY interval		NP Estimation	95% NP interval	
Mean	25 313.12	15 589.21	37 412.36	36 504.26	15 277.85	44 843.92
Median	24 660.22	15 048.96	36 716.66	35 796.55	14 451.90	43 927.78
Std. dev.	9 640.26	7 546.71	11 885.85	11 618.72	8 895.56	13 283.07
0.95 quant.	42 230.53	28 970.45	58 104.65	56 754.88	31 315.10	67 986.48
0.99 quant.	50 403.21	35 813.18	67 637.69	66 162.25	38 920.00	78 510.05

Table 4: Bayesian estimates (BY) and 95% credible intervals of some characteristic measures of $\tilde{S}(200)$, compared with the equivalent nonparametric (NP) point estimates and 95% confidence intervals in sector C.

6.2 Analysis of the global sector

We now present the results obtained for the two sectors jointly. Thus, we have a complete sample of 1000 inter-occurence times between claims and a sample of 1000 claim sizes, with 300 left-censored data. For each of the two data samples, we run the corresponding MCMC algorithm described in Section 4. Using the MCMC output, we are able to estimate the distributions and main characteristic measures of the total claim count, $N_G(200)$, and the total claim amount, $S_G(200)$, that will be observed in the global sector using the

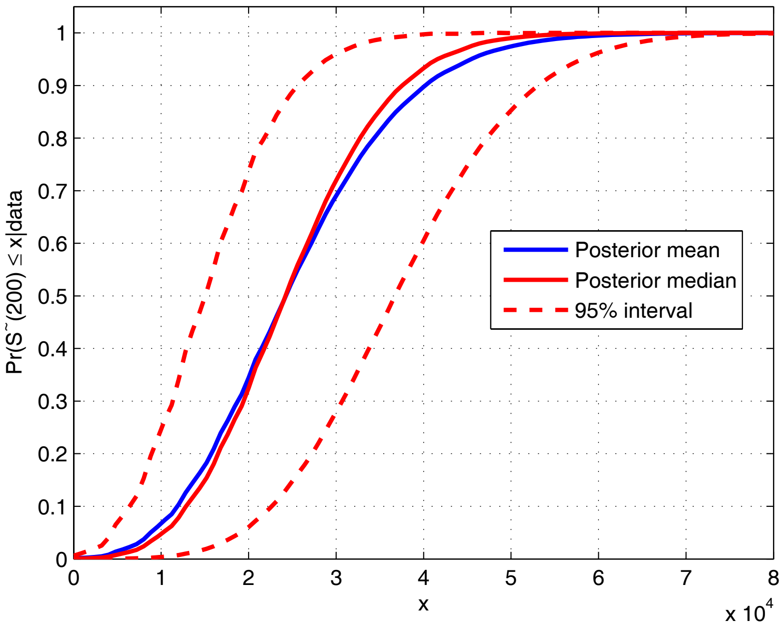

Figure 6: Posterior mean, median and 95% credible interval for the cumulative distribution function of the total claim amount, $\tilde{S}\,(200)$, up to time $t = 200$ with a deductible amount, $a = 12\ 000$, and a limit amount, $b = 16\ 000$, in sector C.

proposed procedure described in Section 4.

Tables 5 and 6 presents the Bayesian estimation results and predictive intervals with $\alpha = 0.05$ for the main characteristic measures associated to $N_G\,(200)$ and $S_G\,(200)$, respectively. These are compared with the equivalent nonparametric point estimates and confidence intervals. Observe that also for the global sector the estimation results obtained with both approaches are rather comparable.

Measure	BY estimation	95% BY interval		NP Estimation	95% NP interval	
Mean	399.28	369.71	429.49	385.23	354.34	411.85
Median	399.14	370.00	429.00	385.00	354.00	412.00
Std. dev.	24.12	22.11	26.73	23.24	21.21	25.48
0.95 quantile	439.18	409.00	471.00	423.00	391.00	449.00
0.99 quantile	456.03	425.00	489.00	438.00	405.00	464.00

Table 5: Bayesian estimates (BY) and 95% credible intervals of some characteristic measures of $N_G\,(200)$, compared with the equivalent nonparametric (NP) point estimates and 95% confidence intervals for the global sector.

Measure	BY estimation	95% BY interval		NP Estimation	95% NP interval	
Mean	1 211 905.35	1 005 404.50	1 484 856.03	1 282 345.55	806 030.49	1 424 752.83
Median	1 197 517.98	994 806.52	1 461 266.93	1 262 393.83	812 525.45	1 403 843.83
Std. dev.	191 285.88	140 767.39	279 941.64	229 816.85	35 309.81	275 564.40
0.95 quant.	1 548 429.41	1 259685.00	1 963 150.76	1 696 326.32	882 639.23	1 914 863.11
0.99 quant.	1 716 911.07	1 383 157.37	2 206 583.17	1 901 194.91	820 948.52	2 149 454.59

Table 6: Bayesian estimates (BY) and 95% credible intervals of some characteristic measures of $S_G(200)$, compared with the equivalent nonparametric (NP) point estimates and 95% confidence intervals for the global sector.

7 Comments and extensions.

We have developed a Bayesian approach to make inference for insurance aggregate loss models. A semiparametric density approximation based on Coxian distributions have been proposed for the estimation of the claim inter-arrival and claim size distributions. We have constructed an MCMC algorithm to obtain samples from the posterior distribution of the model parameters and then, we have combined this with Laplace inversion methods to make predictions about the total number of claims and total claim amount in future time periods. We have illustrated the proposed procedure with a real data set from the insurance department of a commercial company. The estimation results have shown to be comparable with those obtained with a non parametric approach developed in Gonález-Fragueiro et al. (2006).

A notable difference between the results obtained with the Bayesian and the nonparametric approach is that point estimations of individual and aggregate claim sizes are in general larger with the proposed Bayesian method. This can be due to differences in the tail estimation with both approaches. Note that the Bayesian procedure is based on the Coxian distribution which is a parametric model and then, assigns a positive (small) probability for very large claim sizes. In contrast, the nonparametric procedure gives zero probability for those values which are larger than the maximum observed claim size. A second difference between the results obtained with both approaches is that confidence intervals are in general narrower with the Bayesian approach. The most probable reason for this is that the uncertainty is usually smaller using a parametric method and then, provided that the parametric model is adequate, it leads to more accurate estimations. Finally, the computational cost is sensibly larger using the proposed Bayesian approach than the nonparametric method. For example, the total computational cost required to obtained all predictive distributions, main characteristic measures and predictive intervals for three sectors individually and globally was approximately 18 hours, while the nonparametric approach required approximately 11 hours, using MATLAB (The MathWorks, Inc.) with both procedures.

Although the proposed Bayesian MCMC algorithm has been constructed for possibly left-censoring data, it is straightforward to modify it for the case of right-censoring or for the case that there are both right and left-censored data in the sample. In these cases, we simply update the missing data in each MCMC iteration by simulating from the corresponding Coxian distribution truncated to the non-censored region.

We have found that eventually a large number of mixture components are obtained with some over-fitting problems such as giving a single mixture component for a small data subset of close to zero values. One possibility could be assuming a Poisson prior distribution on the mixture size in order to penalize a large number of components in the mixture.

Both claim arrival and claim size processes have been assumed to be renewal sequences of i.i.d. random variables. This assumptions could not be very realistic in some practical situations where, for example, time of day effects produce non-stationarity. Thus, more generally, we could assume for example a Markov modulated claim arrival process and use the Bayesian procedure proposed in Scott and Smyth (2003) for this process, such that we could extend our approach for this case and make inference for the claim count and claim size random variables.

Finally, we could also extend our approach to make inference about other quantities of interest in insurance aggregate loss models, such as the probability of ruin. An advantage of the Coxian model is that it is a phase-type distribution and then, explicit expression for the ruin probabilities can be obtained when the model parameters are known. Using this results, we could apply a similar approach to the proposed in this article to make Bayesian inference on these probabilities. Related ideas are developed in Bladt et al. (2003).

References

Abate, J., and Whitt,W., (1992). The Fourier-series method for inverting transforms of probability distributions. *Queueing Systems*, **10,** 5-88.

Asmussen, S., (2000). *Ruin probabilities.* World Scientific Publishing, Singapore.

Ausín, M.C., Wiper, M.P., and Lillo, R.E., (2004). Bayesian estimation for the M/G/1 queue using a phase type approximation. *Journal of Statistical Planning and Inference*, **118**, 83-101.

Bladt, M., Gonzalez, A., and Lauritzen, S.L., (2003). The estimation of phase-type related functionals using Markov Chain Monte Carlo methods. *Scandinavian Actuarial Journal*, **2003**, 280-300.

Cairns, A.J.G., (2000). A discussion of parameter and model uncertainty in insurance. *Insurance: Mathematics and Economics*, **27**, 313-330.

Cizek, P., Hardle, W., Weron, R., (2005). *Statistical Tools for Finance and Insurance*, Springer, New York.

Dickson, D.C., Tedesco, L.M., Zehnwirth, B., (1998). Predictive aggregate claim distributions. *Journal of Risk and Insurance*, **65**, 689-709.

Geweke, J., (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In: Bernado, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds.), *Bayesian Statistics 4*. Clarendon Press, Oxford.

Gilks, W., Richardson, S., and Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in practice*. Chapman and Hall, Londres.

González-Fragueiro, C., Vilar, J.M., Cao, R., and Ausín, M.C., (2006). Analysis of an aggregate loss model. Discussion paper.

Gruet, M.A., Philippe, A., and Robert, C.P., (1999). MCMC control spreadsheets for exponential mixture estimation. *Journal of Computational and Graphical Statistics*, 8, 298-317.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika,* **57**, 97-109.

Johnson, N.L. and Kotz, S. (1970). *Distributions in statistics. Continuous univariate distributions.* John Wiley and Sons, New York.

Klugman, S.A., (1992). *Bayesian Statistics in Actuarial Science*, Kluwer Academic Publisher, Norwell, MA.

Klugman, S.A., Panger, H.H., and Willmot, G.E., (2004). *Loss models: From data to decisions*, (2nd ed.) John Wiley.

Neuts, M.F., (1981). *Matrix-Geometric Solutions in Stochastic Models*. Johns Hopkins University Press, Baltimore, MD.

Richardson, S., and Green, P.J., (1997). On Bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society, Series B,* **59**, 731-792.

Robert, C.P., and Mengersen, K.L., (1999). Reparameterisation Issues in Mixture Modelling and their bearing on MCMC algorithms. *Computational Statistics and Data Analysis*, **29**, 325-343.

Rolski, T., Schmidli, H., Schmidt, V., Teugels, J., (1999). *Stochastic processes for insurance and finance.* John Wiley and Sons, New York.

Scott, S.L., and Smyth, P., (2003). The Markov Modulated Poisson Process and Markov Poisson Cascade with applications to web traffic data. In Bayarri, M. J., Berger, J. O., Bernardo, J. M., Dawid, A. P., Heckerman, D., Smith, A. F.M., and West,M., eds., *Bayesian Statistics 7*, pp. 671-680. Oxford University Press.

Wiper, M.P., Rios, D., Ruggeri, F., (2001). Mixtures of gamma distributions with applications. *Journal of Computational and Graphical Statistics*, **10**, 440-454.