The estimation of phase-type related functionals using Markov chain Monte Carlo methods*

Mogens Bladt
Universidad Nacional Autónoma de México

Antonio Gonzalez
Benemérita Universidad Autónoma de Puebla

Steffen L. Lauritzen
Aalborg University

November 2001

Abstract

In this paper we present a method for estimation of functionals depending on one or several phase-type distributions. This could for example be the ruin probability in a risk reserve process where claims are assumed to be of phase-type. The proposed method uses a Markov chain Monte Carlo simulation to reconstruct the Markov jump processes underlying the phase-type variables in combination with Gibbs sampling to obtain a stationary sequence of phase-type probability measures from the posterior distribution of these given the observations. This enables us to find quantiles of posterior distributions of functionals of interest, thereby representing estimation uncertainty in a flexible way. We compare our estimates to those obtained by the method of maximum likelihood and find a good agreement. We illustrate the statistical potential of the method by estimating ruin probabilities in simulated examples.

Keywords: Bayesian analysis, Markov chain Monte Carlo, phase-type distribution, ruin probabilities.

*This is Research Report R-01-2027, Department of Mathematical Sciences, Aalborg University.

1 Introduction

In this article we develop a method for estimating functionals of phase-type distributions. This could for example be the ruin probability of a risk-reserve process where claim sizes have phase-type distributions and arrivals occur according to a phase-type renewal process, or the steady state waiting times (virtual and actual) in a $PH/PH/1$ queue. When any such functional is estimated from data on stochastic systems (like above), the estimates are necessarily subject to uncertainty. We are interested in representing this uncertainty and will do so using a Bayesian simulation approach, which in particular enables us to obtain credibility intervals and quantiles for the posterior distribution of such functionals in a simple way.

The main idea is to generate a stationary sequence of random phase-type measures from a distribution incorporating the information in the observations and use ergodicity of such a sequence to estimate the quantities of interest by the empirical averages of corresponding functionals of the measures in the sequence.

A crucial part of this task turns out to be the simulation of a Markov jump process underlying a phase-type variable. More precisely, for an observation $X = x$ from a phase-type distribution, we establish a method for simulating from the conditional distribution of the underlying Markov jump process given the absorption time $X = x$. This simulation is made as a Metropolis–Hastings (MH) algorithm. The final algorithm used for inference is a Gibbs sampler which in each iteration uses this MH algorithm to reconstruct the Markov jump processes.

The article is organized as follows. In Section 2 we provide some relevant background on phase-type distributions, Bayesian analysis, and Markov chain Monte Carlo methodology, whereas Section 3 is devoted to construction of the algorithms. In Section 3.1 we develop an algorithm for generating Markov jump processes which get absorbed at a fixed time. Section 3.2 is devoted to the specification of prior distributions over phase-type measures.

In Section 3.3 we construct a Gibbs sampler using the reconstruction from Section 3.1 and provide some examples of application with simulated data. Section 3.4 describes the use of hyper-priors for obtaining a less informative prior and improving the mixing of the Markov chain of Markov jump processes. A more challenging example involving a double sampler is described in Section 3.5. Finally the work is summarized and further perspectives are discussed in Section 4.

2 Some relevant background

2.1 Phase-type distributions

Consider a Markov jump process $\{J_t\}_{t\geq 0}$ with p transient states and one absorbing state (state $p+1$). Then the intensity matrix $\mathbf{\Lambda}$ can be written in a block-partitioned way as

$$\mathbf{\Lambda} = \begin{pmatrix} \boldsymbol{T} & \boldsymbol{t} \\ \mathbf{0} & 0 \end{pmatrix},$$

where $\boldsymbol{T}$ is a $p \times p$ sub-intensity matrix, $\boldsymbol{t}$ is the column vector (dimension p) of exit intensities (exit to the absorbing state) and $\mathbf{0}$ is the row vector (dimension p) of zeroes. Furthermore, let $\boldsymbol{\pi}$ be the p-dimensional row vector giving the initial distribution of the Markov jump process, $\pi_i = \mathbb{P}(J_0 = i), \quad i = 1, \ldots, p$. Note that $\boldsymbol{\pi}$ is concentrated on the first p states, i.e. we are not allowed to start in the absorbing state. Then we say that the distribution of the time until absorption, X, is a *phase-type distribution of order* p (or with p phases) generated by $\boldsymbol{\pi}$ and $\boldsymbol{T}$, and we write $X \sim \mathrm{PH}(\boldsymbol{\pi}, \boldsymbol{T})$. The pair $(\boldsymbol{\pi}, \boldsymbol{T})$ is referred to as a *representation* of the phase-type distribution. Note that $\boldsymbol{t} = -\boldsymbol{T}\boldsymbol{e}$ where $\boldsymbol{e}$ is the column vector of 1's. The density is given by

$$f_X(x) = \boldsymbol{\pi} e^{\boldsymbol{T}x} \boldsymbol{t}. \tag{1}$$

Phase-type distributions generalize mixtures of convolutions of exponential distributions.

Exact solutions to many problems arising in areas of applied probability such as risk theory and queueing theory can be obtained when distributions involved are assumed to be of phase-type. Here we give a few examples:

In renewal theory the renewal density $u(x)$ of an arrival at time x has the form $u(x) = \boldsymbol{\pi} \exp((\boldsymbol{T} + \boldsymbol{t}\boldsymbol{\pi})x)\boldsymbol{t}$ when the inter-arrival distribution is $\mathrm{PH}(\boldsymbol{\pi}, \boldsymbol{T})$.

In risk theory, the ruin probability of a risk reserve process can be calculated exactly when arrivals of claims are Poissonian, phase-type renewal or follow a Markov modulated Poisson process, and claims are phase-type distributed, (Asmussen, 2000a; Asmussen & Bladt, 1996). For finite-time ruin probabilities with phase-type distributed claims we refer to Stanford (1994).

In queueing theory asymptotic (stationary) waiting time distributions can be calculated when arrivals occur according to a phase-type renewal process and service times are of phase-type (Asmussen, 1987). One can also obtain asymptotic variances of time averages of such waiting times (Asmussen & Bladt, 1994; Bladt, 1996).

Being very attractive from a probabilistic point of view, phase-type distributions present difficulties concerning statistical issues. The representation above is heavily over-parametrized using in general $p^2 + p - 1$ freely varying parameters for p-order phase-type distributions when in principle they could be represented by $2p - 1$ parameters when considered as matrix-exponential distributions, i.e. distributions having a rational Laplace transform; see Asmussen & Bladt (1992) for details on representing matrix-exponential distributions. Although (1) yields an explicit expression for the likelihood function, this cannot easily be directly maximized. The EM algorithm developed for this problem in Asmussen et al. (1996) works very well in lower dimensions. However, methods for statistical inference beyond estimation of parameters has not yet been proposed. For a review of methods for estimating phase-type distributions, see e.g. Asmussen (2000b).

Due to the over-parameterization of phase-type distributions one might question the relevance of obtaining specific estimates for these parameters.

In most cases it will not be possible or feasible to give a physical interpretation of the corresponding underlying Markov model, and the main reason for applying phase-type distributions to specific problems are usually dictated by some specific application in stochastic modeling. For example in risk theory one may be interested in estimating ruin probabilities based on data concerning claim sizes and their arrivals; modeling claims as phase-type distributions is convenient both because phase-type distributions are dense in the class of distributions on the positive real axis and because ruin probabilities in many cases can be explicitly calculated under the phase-type assumption. The ruin probability of course depends on the parameters of the phase-type distribution, but it is unchanged under different representations of the same phase-type distributions. Hence statistical inference for such functionals as the ruin probability is more sensible than for the parameters of the phase-type distribution themselves.

For further properties and general treatment of phase-type distributions we refer to Asmussen (1987, 2000a), Asmussen & Olsson (1998), Asmussen & Bladt (1992) or Neuts (1981). Hipp (1989a, 1989b) discusses estimation and confidence intervals in the special case where the functional is a ruin probability. Rolski et al. (1999) give a general treatment of phase-type methodology in insurance risk.

2.2 Bayesian analysis

This section gives a brief review of basic concepts in the Bayesian approach as used in this paper. We refer the reader to Bernardo & Smith (1994) for the complete story.

We consider a situation with observations $X_1 = x_1, \ldots, X_N = x_N$ of i.i.d. random variables X_i from a distribution with density $f(\cdot \mid \theta)$, where $\theta \in \Theta$ is an unknown parameter (possibly of high or even infinite dimension). We let $x = (x_1, \ldots, x_N)$ so that

$$f(x \mid \theta) = f(x_1 \mid \theta) \cdots f(x_N \mid \theta).$$

In the case considered in the present paper, θ is the representation $(\boldsymbol{\pi}, \boldsymbol{T})$ of a phase-type distribution.

Rather than devising methods for estimation of the unknown parameter θ, a Bayesian analysis first specifies a *prior* distribution G over the parameter space Θ, ideally representing the initial (uncertain) knowledge about θ. We shall later return to the problem of specification of this prior distribution. The density $f(\cdot \mid \theta)$ is now interpreted as a conditional distribution given θ, so that f and G together define a joint distribution P over $\mathcal{X} \times \Theta$. The inference is then summarized through the *posterior* distribution G^* obtained from P by conditioning with the data x so that

$$g^*(\theta) = \frac{dG^*}{dG}(\theta) = \frac{dP(\cdot \mid x)}{dG}(\theta) \propto L(\theta \mid x) = f(x \mid \theta) = f(x_1 \mid \theta) \cdots f(x_N \mid \theta).$$

The density of the posterior with respect to the prior is thus proportional to the likelihood function L.

The posterior distribution G^* represents the complete inference about θ, combining the prior knowledge, represented by G, with the information in the data, represented by L.

In specific examples, such as those in the present paper, one may focus interest on one or several particular functionals $\psi = \psi(\theta)$ of the parameter. In our case, ψ could be the ruin probability in a system associated with the phase-type distribution, or the phase-type distribution itself, e.g. expressed through its cumulative distribution function.

Inference about ψ is represented by the posterior distribution of ψ or specific characteristics of this distribution such as, for example, its *mean*

$$\psi^* = \boldsymbol{E}\{\psi(\theta) \mid x\} = \int_{\Theta} \psi(\theta)\, G^*(d\theta) \tag{2}$$

or the quantiles of the posterior distribution of ψ if ψ is one-dimensional. The posterior expectation ψ^* in (2) is often referred to as the *Bayes estimate* of ψ, although this is somewhat inaccurate, because there are many interesting characteristics of the posterior distribution other than its mean. An interval such as $[u_{.025}, u_{.975}]$, where u_α is the α-quantile of the posterior distribution of ψ, is a 95% *credibility interval* for ψ. Beware that the interpretation of a credibility interval is quite different from that of a traditional confidence interval, which has a much more complex genesis.

The remaining difficult issues of Bayesian inference are concerned with specification of the prior distribution G, representation of the posterior G^*, and computation of integrals w.r.t. G^* such as e.g. (2).

For reasons of simplicity one often uses conjugate families of prior distributions. A family $\mathcal{G}$ of distributions is said to be *conjugate* for a Bayesian inference problem, if it is closed under the prior-to-posterior analysis, i.e. if $G \in \mathcal{G}$ implies that $G^* \in \mathcal{G}$ for any possible data x. Conjugate families are sometimes conveniently indexed by its own *hyper-parameter* η, i.e. $\mathcal{G} = \{G_\eta, \eta \in H\}$. The prior-to-posterior analysis can then be summarized by specifying how the posterior hyper-parameter η^* depends on the prior hyper-parameter η and the data x. For a general discussion of conjugate distributions, see e.g. Diaconis & Ylvisaker (1979) and Peña & Smith (1995).

In general it may be impossible to find conjugate families where this dependence is so simple that it is useful, and even when this can be done, calculation of integrals with respect to the posterior distribution may still present difficulties which in most cases can be overcome only by using Markov chain Monte Carlo Methods.

2.3 Markov chain Monte Carlo methods

Modern Markov chain Monte Carlo methods (MCMC) have their origin in statistical physics (Metropolis et al., 1953), where they have been used to simulate the behaviour of large systems of particles and complex molecules. Their primary use in modern statistics has been the computation of integrals with respect to Bayesian posterior distributions in complex problems (Gelfand and Smith, 1990; Gilks et al., 1996), although they can also be used for a traditional likelihood analysis (Geyer and Thompson, 1992). See Green (2001) for a recent tutorial on MCMC methods.

There is an abundance of related but different MCMC algorithms and here we shall explain and exploit the (systematic) *Gibbs sampler*, introduced in this form in Geman & Geman (1984), and the *Metropolis–Hastings* (MH) algorithm (Hastings, 1970).

The basis of the Gibbs sampler is a collection $(Y_v)_{v\in V}$ of random variables with a joint *target* distribution $\boldsymbol{\tau}$. It then proceeds as follows. First, pick an arbitrary starting configuration $y^0 = (y_v^0)_{v\in V}$. Then number the elements of V as $V = \{1, \dots, |V|\}$ and make successive random drawings from the *full conditionals* $\mathcal{L}(Y_v \mid Y_{V\setminus\{v\}})$ as:

pick y_1^1 from $\mathcal{L}(Y_1 \mid y^0_{V\setminus\{1\}})$;

pick y_2^1 from $\mathcal{L}(Y_2 \mid y^0_{V\setminus\{1,2\}}, y_1^1)$;

pick y_3^1 from $\mathcal{L}(Y_3 \mid y^0_{V\setminus\{1,2,3\}}, y_{1,2}^1)$;

continue in this manner until you pick $y^1_{|V|}$ from $\mathcal{L}(Y_{|V|} \mid y^1_{V\setminus\{|V|\}})$.

Each of the steps above is referred to as a *site visit.* This reflects the origin of the ideas in physics where variables correspond to states of particles placed at different sites. When all sites have been visited, a transition from $y^0 = (y_v^0)_{v\in V}$ to $y^1 = (y_v^1)_{v\in V}$ has taken place. Iteration of the procedure creates successive values $y^0, y^1, \dots, y^n, \dots$. The point of the method is that under very general conditions these successive values form a realization from a Markov chain which has the target $\boldsymbol{\tau}$ as its equilibrium distribution. By ergodicity, integrals of a function h with respect to $\boldsymbol{\tau}$ can then be approximated by averages of the Gibbs sample

$$\int h(y)\, \boldsymbol{\tau}(dy) \approx \frac{1}{n}\sum_{\nu=1}^{n} h(y^\nu). \tag{3}$$

In Bayesian inference the target is often the conditional distribution of Y given an observed subset of the variables $Y_v = y_v^*, v \in A$ where $A \subseteq V$. The only modifications needed to obtain a sample from this conditional distribution are that the starting configuration must satisfy $y_v^0 = y_v^*$ for all $v \in A$ and that sites in A are not updated.

The Metropolis–Hastings algorithm is not necessarily related to specific sites and we therefore suppress the site index and write the iteration number as subscript instead of superscript. The MH algorithm constructs a Markov chain $\{Y_n\}$ by drawing $Z = z$ for each n from a *proposal distribution* $\boldsymbol{\rho}^{y_n}$ proposing to let the chain move to $Y_{n+1} = z$, and then accepting or rejecting this proposal with a suitable probability $a(z, y_n)$. In principle the proposal distribution is almost arbitrary, but the efficiency of the algorithm depends on the acceptance probabilities not being too small. Thus the MH algorithm is as follows

- Initiate at any point $Y_0 = y_0$.
- In the nth step, draw $Z = z$ from $\boldsymbol{\rho}(\cdot \mid y_n)$ and set

$$Y_{n+1} = \begin{cases} z & \text{with probability } a(y_n, z) \\ y_n & \text{with probability } 1 - a(y_n, z), \end{cases}$$

  where the acceptance probability $a(y_n, z)$ is determined as

$$a(y_n, z) = \min\left\{1, \frac{\dfrac{d\boldsymbol{\tau}}{d\boldsymbol{\rho}^{y_n}}(z)}{\dfrac{d\boldsymbol{\tau}}{d\boldsymbol{\rho}^{z}}(y_n)}\right\}. \tag{4}$$

If $\boldsymbol{\rho}^y = \boldsymbol{\rho}$ we talk about an *independence sampler.* In this paper all MH algorithms are independence samplers.

For both the Gibbs and MH algorithms it is common to discard the observations in an initial *burn-in* period and only let the average (3) extend over values obtained after this burn-in. Problems with MCMC methods are associated with the fact that convergence to equilibrium can be very slow if the Markov chain is not mixing well and it can be quite difficult to assess such convergence in a practical situation.

A variant of the algorithms is known as 'Metropolis-within-Gibbs', where an MH-step is used to replace the Gibbs updating at a single site. Indeed, the final algorithm of the present paper is such a variant, where observations from a burn-in period of the MH sampler are discarded before the Gibbs updating is made. The Gibbs sampler is in the present paper used in a case where $|V| = 2$ and the target distribution is the conditional distribution of $(\theta, \boldsymbol{Y})$ given data $\boldsymbol{x}$, where $\theta = (\boldsymbol{\pi}, \boldsymbol{T})$ is the phase-type representation and $\boldsymbol{Y}$ denotes the collection of full trajectories of the underlying Markov jump processes. We then sample θ given $\boldsymbol{y}$ (and $\boldsymbol{x}$) using simple conjugate distributions and $\boldsymbol{Y}$ given $(\theta, \boldsymbol{x})$ using a Metropolis–Hastings algorithm. Strictly speaking, we update each component of θ separately, and each of the N trajectories separately but, as we shall see below, appropriate conditional independencies allow us to consider the joint updating of these groups as single updating steps.

3 Sampling of phase-type distributions

3.1 Sampling from the Markov process associated with a phase-type variable

Let X be a random variable with a phase-type distribution and J be the associated Markov jump process. We would like to simulate the Markov jump process J from the conditional distribution of J given $X = x$, where X is the time to absorption of a Markov jump process with intensity matrix $\boldsymbol{\Lambda}$ and initial distribution $(\boldsymbol{\pi}, 0)$.

The idea is to substitute this Markov jump process, which gets absorbed exactly at time x, with another Markov jump process having absorption time beyond x and use the substitute as a proposal in a Metropolis-Hastings algorithm.

Let $\{J_t\}$ denote the Markov jump process with intensity matrix $\boldsymbol{\Lambda}$ and initial distribution $\boldsymbol{\alpha} = (\boldsymbol{\pi} \;\; 0)$. Then the distribution of J_s is $\boldsymbol{\alpha} e^{\boldsymbol{\Lambda} s}$, and hence

$$q_i(s) := \mathbb{P}_{\boldsymbol{\alpha}}(J_s = i) = \boldsymbol{\alpha} e^{\boldsymbol{\Lambda} s} \boldsymbol{e}_i,$$

is the probability that the Markov jump process is in state i at time s, where $\boldsymbol{e}_i$ is the column vector which i'th coordinate is 1, all other zeroes. Since

$$e^{\boldsymbol{\Lambda} s} = \begin{pmatrix} e^{\boldsymbol{T} s} & \boldsymbol{e} - e^{\boldsymbol{T} s}\boldsymbol{e} \\ \boldsymbol{0} & 1 \end{pmatrix},$$

it holds for $i \in \{1, 2, \ldots, n\}$ that

$$q_i(s) = \boldsymbol{\pi} e^{\boldsymbol{T} s} \boldsymbol{e}_i.$$

Also, since $\boldsymbol{t} = \sum_i t_i \boldsymbol{e}_i$, the density itself can be expressed simply by the function q as

$$f_X(x) = \sum_i q_i(x) t_i.$$

The distribution of the Markov jump process exactly prior to absorption is

$$\tilde{\pi}_i := \mathbb{P}\{J_{x-} = i \mid X = x\} = \frac{q_i(x) t_i}{f_X(x)}, \tag{5}$$

which follows from

$$\begin{aligned} q_i(x) t_i dx &= \mathbb{P}\left(J_{x-} = i, X \in [x, x + dx)\right) \\ &= \mathbb{P}(J_{x-} = i \,|\, X = x) f_X(x) dx. \end{aligned}$$

Let $\mathbb{P}_x = \mathbb{P}\left(\cdot \,|\, X = x\right)$ denote the distribution of interest and let $\mathbb{P}^*_x = \mathbb{P}\left(\cdot \,|\, X \geq x\right)$ be the distribution of $J_{t-}, 0 \leq t \leq x$ conditionally on the event that $X \geq x$. Hence $\mathbb{P}_x$ is the target distribution and $\mathbb{P}^*_x$ will serve as a proposal distribution. The latter is easy to simulate, as we just simulate the original process and reject if absorption happens prior to x. Otherwise we accept.

The strong Markov property of $\{J_t\}$ implies that $J_{t-}, 0 \leq t \leq x$ and X are conditionally independent given J_{x-}. Hence,

$$\mathbb{P}_x(\{J_{t-}\}_{t<x} = \{j_{t-}\}_{t<x} \,|\, J_{x-}) = \mathbb{P}^*_x(\{J_{t-}\}_{t<x} = \{j_{t-}\}_{t<x} \,|\, J_{x-}).$$

The distribution $\mathbb{P}_x^*$ at time $x-$ is given by

$$\pi_i^* := \mathbb{P}\{J_{x-} = i \mid X \geq x\} = \frac{q_i(x)}{\sum_j q_j(x)}$$

because $\mathbb{P}(X \geq x) = \sum_{j=1}^n \mathbb{P}(J_{x-} = j) = \sum_{j=1}^n q_j(x)$. Thus we get that

$$\begin{aligned}
\frac{d\mathbb{P}_x}{d\mathbb{P}_x^*}\{j_t, t < x\} &= \frac{\mathbb{P}_x(\{J_t : t < x\} = \{j_t : t < x\})}{\mathbb{P}_x^*(\{J_t : t < x\} = \{j_t : t < x\})} \\
&= \frac{\mathbb{P}_x(\{J_t : t < x\} = \{j_t : t < x\} \mid J_{x-} = j_{x-})\mathbb{P}_x(J_{x-} = j_{x-})}{\mathbb{P}_x^*(\{J_t : t < x\} = \{j_t : t < x\} \mid J_{x-} = j_{x-})\mathbb{P}_x^*(J_{x-} = j_{x-})} \\
&= \frac{\mathbb{P}_x\{J_{x-} = j_{x-}\}}{\mathbb{P}_x^*\{J_{x-} = j_{x-}\}} = \frac{\tilde{\pi}_{j_{x-}}}{\pi^*_{j_{x-}}}.
\end{aligned}$$

These fractions could in principle be used as weights for an importance sampling procedure based on sampling from $\mathbb{P}_x^*$ instead of $\mathbb{P}_x$. However, it may be computationally burdensome to calculate the importance weights, since they involve matrix-exponentials.

Instead we construct a stationary Markov chain with $\mathbb{P}_x$ as equilibrium distribution based on the MH-algorithm with $\mathbb{P}_x^*$ as proposal distribution.

This involves the repeated potential replacement of a given sample $j' = (j'_t, t < x)$ with a new sample $j = (j_t, t < x)$, obtained by sampling from $\mathbb{P}_x^*$. The MH acceptance probability $a(j, j')$ involves the ratio of the importance weights, cf. (4)

$$a(j, j') = \frac{\pi^*_{j'_{x-}}}{\tilde{\pi}_{j'_{x-}}} \frac{\tilde{\pi}_{j_{x-}}}{\pi^*_{j_{x-}}} = \frac{t_{j_{x-}}}{t_{j'_{x-}}}.$$

Note that this ratio does not involve the numerically troublesome matrix-exponentials and therefore is easily computed.

The stationary distribution of the Markov chain of Markov processes constructed in this way will be that of $\mathbb{P}_x$. Thus applying the scheme iteratively implies that, after a suitable burn-in period, the realizations can be seen as Markov jump processes drawn approximately from the desired conditional distribution x. Formally the algorithm becomes:

Algorithm 3.1 *Sampling from* $\mathbb{P}_x = \mathbb{P}(\cdot \mid X = x)$ *can be done as follows:*

0. *Generate* $\{j'_t, t < x\}$ *from* $\mathbb{P}^*_x = \mathbb{P}(\cdot \mid X \geq x)$ *by rejection sampling.*

1. *Generate* $\{j_t, t < x\}$ *from* $\mathbb{P}^*_x = \mathbb{P}(\cdot \mid X \geq x)$ *by rejection sampling.*

2. *Draw* $U \sim$ *Uniform*$[0, 1]$.

3. *If* $U \leq \min(1, t_{j_{x-}}/t_{j'_{x-}})$ *then replace* $\{j'_t\}_{t<x}$ *with* $\{j_t\}_{t<x}$.

4. *GO TO 1.*

3.2 Conjugate priors for Markov jump processes

The next step in the construction of our Gibbs sampler is the specification of prior distributions on the parameter space for the phase-type distributions under study.

We then choose constant hyper-parameters β_i, ν_{ij} and ζ_i and specify the prior density of $(\boldsymbol{\pi}, \boldsymbol{T})$ as being proportional to

$$\phi(\boldsymbol{\pi}, \boldsymbol{T}) = \prod_{i=1}^{p} \pi_i^{\beta_i - 1} \prod_{i=1}^{p} \exp\left(t_{ii}\zeta_i\right) \prod_{i=1}^{p} \prod_{\substack{j=0 \\ j \neq i}}^{p} t_{ij}^{\nu_{ij}-1},$$

where 0 denotes the absorbing state and $t_{i0} = t_i$ is the exit intensity. Using $t_{ii} = -\sum_{j=1:j\neq i}^{p} t_{ij} - t_i$ we obtain

$$\phi(\boldsymbol{\pi}, \boldsymbol{T}) = \prod_{i=1}^{p} \pi_i^{\beta_i - 1} \prod_{i=1}^{p} t_i^{\nu_{i0}-1} \exp(-t_i\zeta_i) \prod_{i=1}^{p} \prod_{\substack{j=1 \\ j \neq i}}^{p} t_{ij}^{\nu_{ij}-1} \exp(-t_{ij}\zeta_i).$$

It is easy to sample from this distribution since $\boldsymbol{\pi}$, t_i, and t_{ij} are all independent with $\boldsymbol{\pi}$ Dirichlet distributed and t_i and t_{ij} are gamma distributed with the same scale parameter.

Let now $x_1, \ldots, x_N$ denote the observed data from the phase-type distribution to be estimated. As in Asmussen et al. (1996), these can be seen as absorption times and incomplete observations of underlying Markov jump processes, the complete observations being $\boldsymbol{y} = (y_1, \ldots, y_N)$, where we now have let $y_i = \{j_{it}, t < x_i\}$ denote the entire sample paths of the Markov jump processes.

The likelihood function for the complete data y can be written as (Asmussen et al., 1996)

$$\begin{aligned} p(\boldsymbol{y} \mid \boldsymbol{\pi}, \boldsymbol{T}) &= \prod_{i=1}^{p} \pi_i^{B_i} \prod_{i=1}^{p} e^{t_{ii} Z_i} \prod_{i=1}^{p} \prod_{\substack{j=0 \\ j \neq i}}^{p} t_{ij}^{N_{ij}} \\ &= \prod_{i=1}^{p} \pi_i^{B_i} \prod_{i=1}^{p} t_i^{N_{i0}} e^{-t_i Z_i} \prod_{\substack{j=1 \\ j \neq i}}^{p} t_{ij}^{N_{ij}} e^{-t_{ij} Z_i}, \end{aligned}$$

where B_i is the number of times the Markov processes are initiated in state i, Z_i the total time the Markov processes have spent in state i, and N_{ij} the number of jumps from state i to j. Thus the posterior distribution of $(\boldsymbol{\pi}, \boldsymbol{T})$ becomes

$$\begin{aligned} p^*(\boldsymbol{\pi}, \boldsymbol{T}) &= p(\boldsymbol{\pi}, \boldsymbol{T} \mid \boldsymbol{y}) \\ &\propto \phi(\boldsymbol{\pi}, \boldsymbol{T}) p(\boldsymbol{y} \mid \boldsymbol{\pi}, \boldsymbol{T}) \\ &= \prod_{i=1}^{p} \pi_i^{B_i + \beta_i - 1} \prod_{i=1}^{p} t_i^{N_{i0} + \nu_{i0} - 1} e^{-t_i(\zeta_i + Z_i)} \prod_{i=1}^{p} \prod_{\substack{j=1 \\ j \neq i}}^{p} t_{ij}^{N_{ij} + \nu_{ij} - 1} e^{-t_{ij}(Z_i + \zeta_i)}. \end{aligned}$$

and the family of prior distributions is therefore conjugate for complete data with the following updating formulae for the hyper-parameters:

$$\begin{aligned} \beta_i^* &= \beta_i + B_i \\ \nu_{ij}^* &= \nu_{ij} + N_{ij} \\ \zeta_i^* &= \zeta_i + Z_i. \end{aligned}$$

This leads to a posterior distribution which again has $\boldsymbol{\pi}$ Dirichlet distributed, the exit intensities following gamma distributions $\mathrm{Gamma}(1/\zeta_i^*, \nu_{i0}^*)$ and the transition intensities $\mathrm{Gamma}(1/\zeta_i^*, \nu_{ij}^*)$, all independently. This conjugate property is computationally very convenient, since it now becomes trivial to sample from the conditional distribution of $(\boldsymbol{\pi}, \boldsymbol{T})$ given complete data $\boldsymbol{y}$.

3.3 The Gibbs sampler

The Gibbs sampler used for inference alternates between sampling from the conditional distribution of the Markov jump processes $\boldsymbol{Y}$ given $(\boldsymbol{\pi}, \boldsymbol{T})$ and $x_1, \ldots, x_N$, and the conditional distribution of $(\boldsymbol{\pi}, \boldsymbol{T})$ given complete data $\boldsymbol{y}$.

For the first step, we use the Metropolis–Hastings algorithm derived in the previous section, and for the second we use the conjugacy property of the prior distribution for complete data. In summary, we obtain the following algorithm.

Algorithm 3.2 *The full Gibbs sampler: Specify β_i, ν_{ij} and ζ_i, $i, j = 1, \ldots, p$ and let $\boldsymbol{\beta} = \{\beta_i, i = 1, .., p\}$.*

0. *Generate $\boldsymbol{\pi}$, $t_{ij},\ i \neq j$ and $t_i,\ i = 1, \ldots, p$ from the prior distribution.*

1. *Generate $\boldsymbol{Y} = (Y_1, \ldots, Y_N)$, where each Y_i is a Markov jump process which gets absorbed at time x_i obtained using a fixed number of steps of Algorithm 3.1.*

2. *Calculate the statistics $\boldsymbol{b} = \{B_i, i = 1, \ldots, p\}$, $\boldsymbol{z} = \{Z_i, i = 1, \ldots, p\}$ and $\boldsymbol{N} = \{N_{ij}, i, j = 1, \ldots, p\}$ from the data $\boldsymbol{Y}$.*

3. *Draw $\boldsymbol{\pi}$, $t_{ij},\ i \neq j$ and $t_i,\ i = 1, \ldots, p$ from the full conditional:*

$$\begin{aligned}
\boldsymbol{\pi} &\sim \mathrm{Dir}(\boldsymbol{\beta} + \boldsymbol{b}) \\
t_i &\sim \mathrm{Gamma}(1/(\zeta_i + z_i), N_{i0} + \nu_{i0}),\ \ i = 1, \ldots, p \\
t_{ij} &\sim \mathrm{Gamma}(1/(\zeta_i + z_i), N_{ij} + \nu_{ij}),\ \ i \neq j
\end{aligned}$$

4. *GOTO 1;*

In this way, after a certain burn-in period, we produce an approximately stationary sequence of distributions (measures) drawn from the given class of phase-type distributions. This sequence can be used in several ways to obtain information about functionals of the unknown phase-type measure, as we shall see later.

For Algorithm 3.2 it is in principle sufficient to use a single MH-iteration in step 1. This did, however, not provide a satisfactory mixing for the Gibbs sampler, so we use a larger number of initial steps of the MH algorithm for each site visit.

By ergodicity, the empirical average of the phase-type measures with representations drawn in the Gibbs sampler converges to the posterior mean of the phase-type distribution which is the object for estimation. However, it should be noted that although this average is a mixture of phase-type distributions and therefore itself of phase-type, the representation of this mixture would need many more phases than the original phase-type distribution.

Example 3.1 To test the algorithms developed we initially analyse the following simple example.

We consider 250 simulated observations from a log-normal distribution $\mathrm{LN}(-0.32, 0.8)$. This distribution has been considered earlier in risk theory for adjusting insurance claim data (Asmussen & Rolski, 1991; Asmussen & Bladt, 1996). We adjust a fourth-order phase-type distribution to data by the EM-algorithm and compare with an empirical average of draws obtained from the Gibbs sampler. The empirical average is obtained through averaging histograms generated from each of the draws. An initial burn-in of 80 iterations was given (though 20-30 appeared to have been sufficient) and an additional of 1000 iterations were performed for displaying the following graphs.

In Figure 1 we have displayed the fits of the EM-algorithm and the averaged histograms of Gibbs samples from the posterior distributions and they seem to be quite alike. Also the tail behaviour of the estimated distributions are similar, as can be seen in Figure 2.

Next we estimate the ruin probability for this distribution in a model with Poisson arrivals. More precisely, we consider a risk reserve process R_t

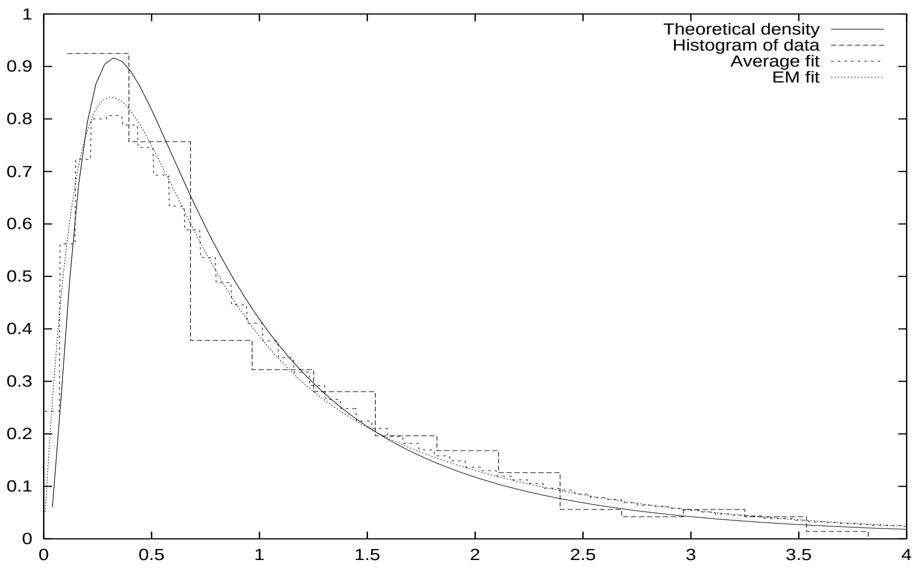

Figure 1: True density, histograms of data, and densities estimated by maximum likelihood via the EM algorithm and posterior mean via MCMC.

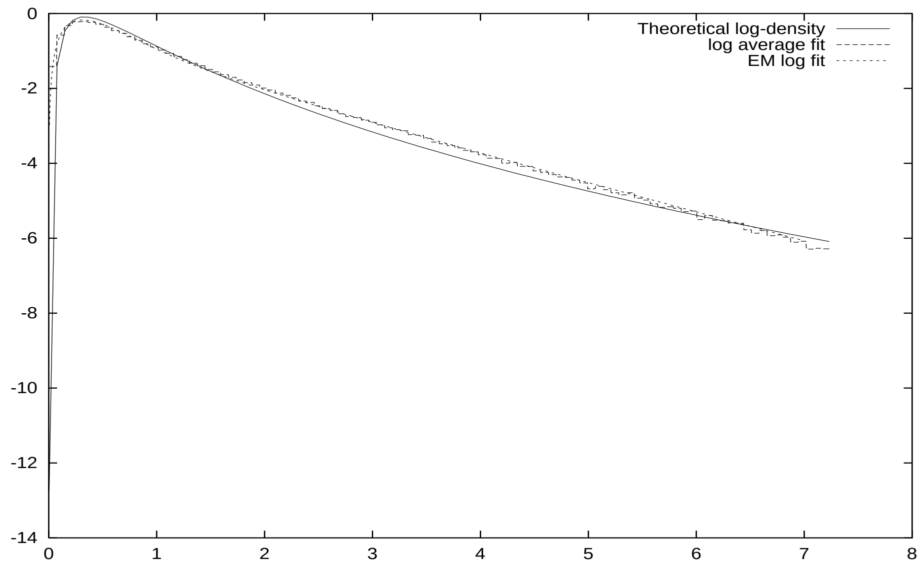

Figure 2: Logarithmic plot of theoretical and estimated densities

defined by

$$R_t = C + t - \sum_{i=1}^{N(t)} U_i,$$

where insurance claims $U_1, U_2, \ldots$ arrive according to a Poisson process with intensity λ, C is the initial capital, $N(t)$ is the number of claims which have arrived up to time t, and it is assumed that the claims are independent and follow a phase-type distribution $PH(\boldsymbol{\pi}, \boldsymbol{T})$. Unit rate increase of capital between arrival of claims has be assumed without loss of generality. For fixed parameters $(\boldsymbol{\pi}, \boldsymbol{T})$ the ruin probability $\rho(C)$ is given by

$$\rho(C) = \mathbb{P}(\inf_{t>0} R_t < 0 \,|\, R_0 = C) = \boldsymbol{\pi}_+ e^{(\boldsymbol{T}+\boldsymbol{t}\boldsymbol{\pi}_+)C}\boldsymbol{e},$$

where $\boldsymbol{e} = (1\ 1\ \ldots\ 1)'$ and $\boldsymbol{\pi}_+ = -\lambda\boldsymbol{\pi}\boldsymbol{T}^{-1}$.

Note that the problem of estimating the ruin probability for this model with Poisson arrivals is equivalent to estimating M/PH/1 waiting time distributions for the stationary virtual waiting time (residual workload) V and actual waiting time W, since $\rho(C) = \mathbb{P}(V > C) = \mathbb{P}(W > C)$ (Asmussen, 1987). We assume for simplicity that the arrival intensity is known and that claims data $x_1, x_2, \ldots, x_N$ are available.

Using consecutive draws of $(\boldsymbol{\pi}, \boldsymbol{T})$'s of the Gibbs sampler and evaluating the corresponding ruin probability for each draw, we obtain a time series of ruin probabilities. By ergodicity we may use this to calculate credibility intervals for the ruin probability and quantiles for its posterior distribution.

We used arrival intensity $\lambda = 0.3$ and $C = 1$, obtaining the following results:

- Estimate of ruin probability using estimated parameters of the EM-algorithm: 0.1800
- Posterior mean of ruin probability, obtained as average in MCMC: 0.1806
- Quantiles of posterior distribution of ruin probability:

2.5 %	25 %	50 %	75 %	97.5 %
0.1377 ± 0.005	0.1643 ± 0.002	0.1783 ± 0.002	0.1961 ± 0.002	0.2316 ± 0.008

- 95% credibility interval for ruin probability: $[0.137, 0.235]$.

The bounds given above on the estimated quantiles were ordinary 95% confidence intervals referring to the Monte Carlo error associated with the computation of these quantiles. The bounds were calculated by the formula (Hald, 1952, p. 138)

$$1.96\sqrt{\frac{p(1-p)}{n\hat{f}(\hat{u}_p)^2}} \tag{6}$$

as the series of ruin probabilities appeared to be uncorrelated by the inspection of auto-correlation functions and partial auto-correlation functions. Here n is the length of the series used to calculate the quantile, $\hat{f}(\hat{u}_p)$ is the posterior density at the estimated p-quantile $\hat{u}_p$, as estimated by the histograms of ruin probabilities displayed in Figure 3. The 95% credibility interval for the ruin probability is obtained by also taking into account the Monte Carlo error when calculating the quantiles of the posterior distribution.

3.4 Improved mixing through hyper-priors

Although Example 3.1 seemed to give a good fit, there may be problems for the Gibbs sampler to tune in on data when e.g. data are bimodal or with low probabilities of small values. In these cases the sampler will have problems pulling away from a too dominant prior. In order to improve on this we may choose to make the prior less dominant and propose a more flexible class of priors by adding randomness to the hyper-parameters in the form of hyper-priors.

In other words, we let the prior depend on hyper-parameters which are random variables instead of fixed, and these random variables are now assumed to have distributions with known fixed parameters. For some surveys on how to choose hyper-priors we refer to Gilks et al. (1996).

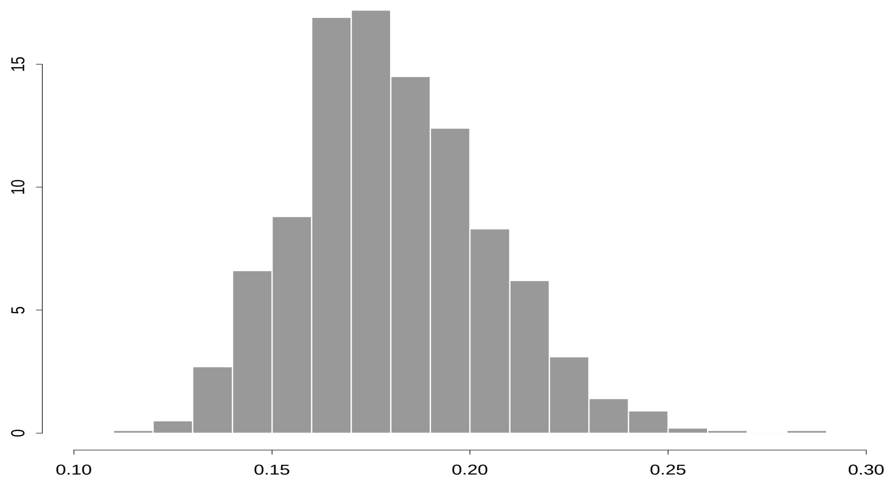

Figure 3: Histogram of ruin probabilities, sampled from their posterior distribution.

In our case we propose to randomize the hyper-parameters ζ_i. Since the values of ζ_i also control mean and variance of the t_{ij}'s, we do not impose further randomization of the ν_{ij}'s nor of $\boldsymbol{\beta}$. We then propose the following family of prior distributions:

$$\begin{aligned}
\boldsymbol{\pi} &\sim \mathrm{Dir}(\boldsymbol{\beta}) \\
t_i &\sim \mathrm{Gamma}(1/\zeta_i, \nu_{i0}) \\
t_{ij} &\sim \mathrm{Gamma}(1/\zeta_i, \nu_{ij}) \\
\zeta_i &\sim \mathrm{Gamma}(1/\theta_i, r_i) \\
\theta_i &\sim \mathrm{Gamma}(1/d, c).
\end{aligned}$$

This family of prior distributions is again conjugate for complete data and the full Gibbs sampling algorithm becomes:

Algorithm 3.3 *Gibbs sampler with hyper-prior: Specify $\beta_i, \nu_{ij}, r_i, c$ and d.*

0. *Draw $\boldsymbol{\pi}$, θ_i, ζ_i, t_{ij}, $i \neq j$ and t_i, $i = 1, \ldots, p$ according to the prior indicated above.*

1. *Generate* $\boldsymbol{Y} = (Y_1, \ldots, Y_N)$*, where each* Y_i *is a Markov jump process which gets absorbed at time* x_i *obtained using a fixed number of steps of Algorithm 3.1.*

2. *Calculate the statistics* $\boldsymbol{b} = \{B_i, i = 1, \ldots, p\}$*,* $\boldsymbol{z} = \{Z_i, i = 1, \ldots, p\}$ *and* $\boldsymbol{N} = \{N_{ij}, i, j = 1, \ldots, p\}$ *from the data* $\boldsymbol{Y}$*.*

3. *Draw* $\boldsymbol{\pi}$*,* θ_i*,* ζ_i*,* t_{ij}*,* $i \neq j$ *and* t_i*,* $i = 1, \ldots, p$ *from the full conditional:*

$$\begin{aligned}
\boldsymbol{\pi} &\sim \mathrm{Dir}(\boldsymbol{\beta} + \boldsymbol{b}) \\
\theta_i &\sim \mathrm{Gamma}(1/(\zeta_i + d), r_i + c) \\
\zeta_i &\sim \mathrm{Gamma}(1/(\theta_i - t_{ii}), \sum_{j=0}^{p} \nu_{ij} + r_i) \\
t_i &\sim \mathrm{Gamma}(1/(\zeta_i + z_i), N_{i0} + \nu_{i0}) \\
t_{ij} &\sim \mathrm{Gamma}(1/(\zeta_i + z_i), N_{ij} + \nu_{ij})
\end{aligned}$$

4. *GOTO 1.*

Example 3.2 We consider two examples of potentially difficult data: Data from a distribution obtained through mixing a log-normal distribution with a gamma distribution, and data from an Erlang(7) distribution. The former is challenging due to bimodality and the latter due to the fact that values close to zero have small probabilities. The bimodal distribution was fitted with 8 phases while the Erlang distribution was fitted with the same number of phases as the Erlang distribution itself (=7). In both cases 250 simulated observations were used. The resulting fits are displayed on Figures 4 and 5.

3.5 A double sampler

Finally we show how to use the MCMC-algorithms constructed above in a situation where the functional of interest depends on two data streams. We consider the risk reserve process from before, but instead of Poisson arrivals we assume that arrivals occur according to a phase-type renewal process, i.e.

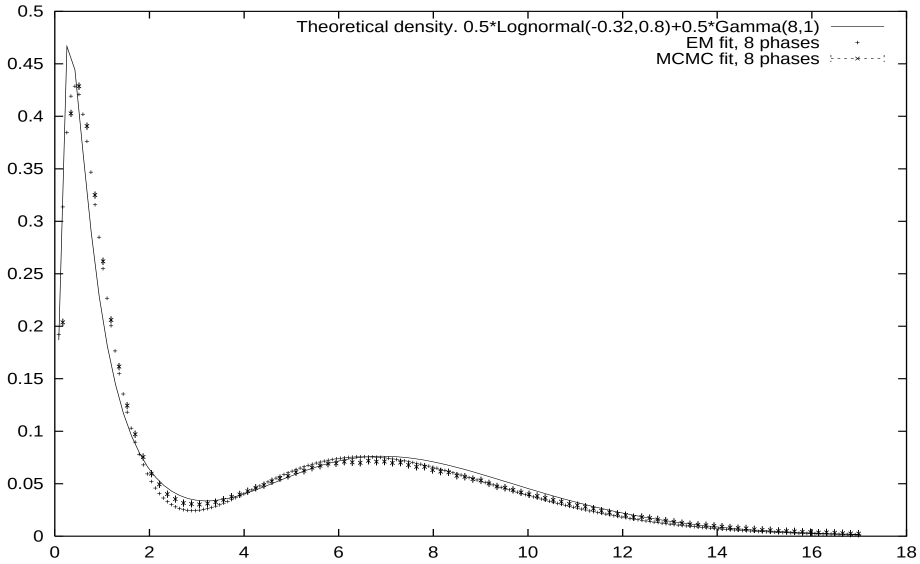

Figure 4: Estimates using 8 phases for a density which is a mixture of a log-normal and a Gamma density.

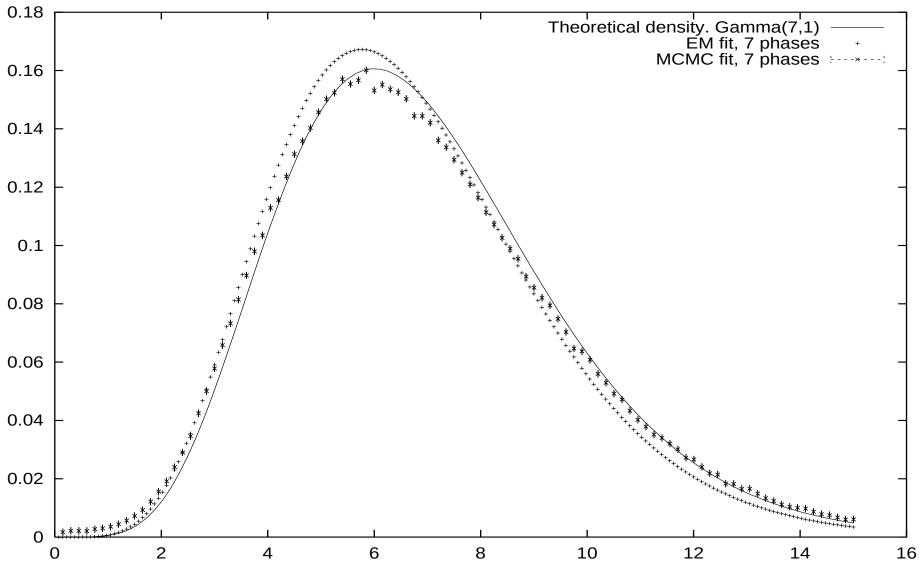

Figure 5: An Erlang(7) distribution, fitted with 7 phases.

inter-arrival times are i.i.d. with a common phase-type distribution PH($\boldsymbol{\sigma}, \boldsymbol{S}$), while the claims remain i.i.d. with common distribution PH($\boldsymbol{\pi}, \boldsymbol{T}$) and independent of the arrival process. Then the (infinite horizon) ruin probability is given by (see Asmussen (2000b), Proposition 4.4)

$$\rho(C) = \boldsymbol{\pi}_+ e^{(\boldsymbol{T}+\boldsymbol{t}\boldsymbol{\pi}_+)C}\boldsymbol{e},$$

where $\boldsymbol{\pi}_+ = \lim_{n\to\infty} \boldsymbol{\pi}_+^{(n)}$ and

$$\boldsymbol{\pi}_+^{(n)} = \boldsymbol{\pi} \int_0^\infty e^{(\boldsymbol{T}+\boldsymbol{t}\boldsymbol{\pi}_+^{(n-1)})y} \boldsymbol{\sigma} e^{\boldsymbol{S}y} \boldsymbol{s} dy.$$

We are interested in the posterior distribution of this ruin probability given data of inter-arrivals and claims. Since they are all assumed independent, we can run two MCMC-samplers as above, but instead of one sample of parameters in each draw we now get two samples from respectively the inter-arrival and the claims data. Then for each pair of parameters, we calculate the corresponding ruin probability. Iterations of this method obviously results in a stationary sequence of ruin probabilities.

Example 3.3 We considered the following numerical example with simulated data. Claims were 250 observations from a log-normal LN($-.32, 0.8$) distribution, inter-arrivals being 250 observations from an Erlang(3) distribution (the number of observations in each series need of course not be the same). We ran 10000 iterations on the double MCMC sampler with 20 iterations in each MH step and hyper-parameters as indicated above, where the dimensions of the phase-type distributions were 3 for the inter-arrivals and 5 for the claims distribution. The last 9500 iterations were considered as stationary.

The ruin probability was evaluated at $C = 1$. Figure 6 illustrates the burn-in period and the settling down of the ruin probabilities to a stationary sequence.

The mixing of the Gibbs sampler can be seen in Figure 7, showing a trace plot of the 9500 ruin probabilities. We notice that the ruin probabilities seem

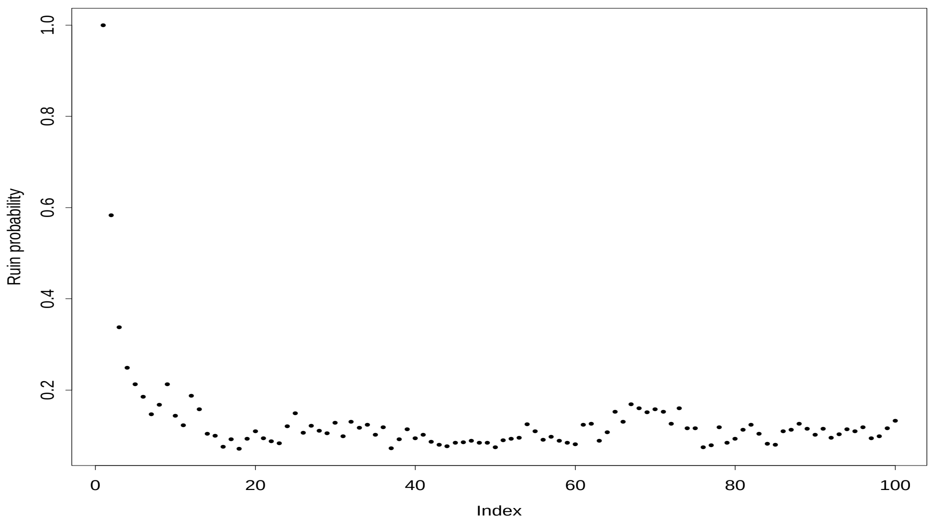

Figure 6: Trace plot of the ruin probabilities during the burn-in period as obtained from the double sampler

to move nicely around in an apparently stationary sequence, although they make a single extreme excursion to the value .2474 at index 7373.

A histogram of the ruin probabilities, representing the posterior distribution of these, is shown in Figure 8. In conclusion we inferred the following:

- Ruin probability calculated using estimated parameters of EM: 0.0805
- Posterior mean of ruin probability: 0.0862
- Quantiles of posterior distribution:

2.5 %	25 %	50 %	75 %	97.5 %
0.05320 $\pm$ 0.0008	0.07141 $\pm$ 0.00076	0.08363 $\pm$ 0.00092	0.09835 $\pm$ 0.0012	0.13354 $\pm$ 0.002

- 95% credibility interval for ruin probability: $[0.0531, 0.1355]$.

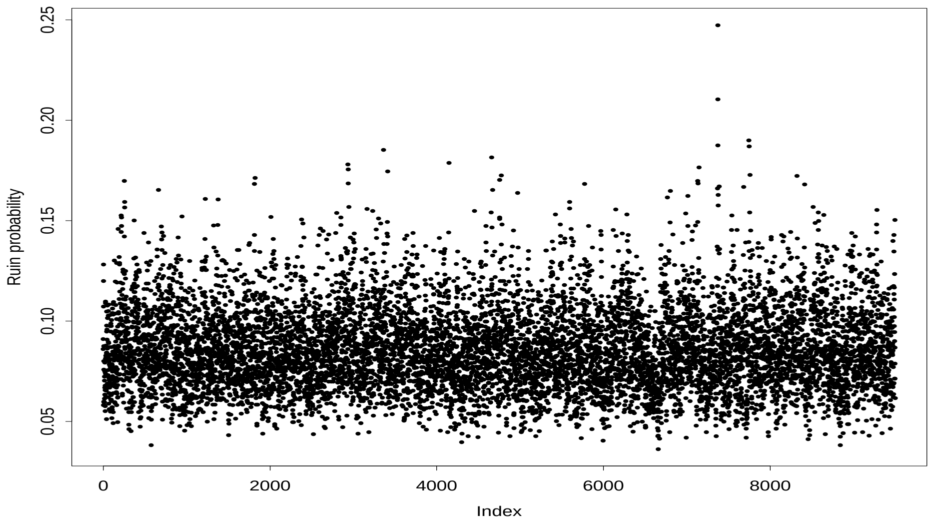

Figure 7: Trace-plot of ruin probabilities as obtained from the double sampler

The confidence bounds on the quantiles were essentially calculated in the same way as in Example 3.1 apart from the fact that here the plots of autocorrelation functions showed that the ruin probabilities could not be considered as being approximately independent. We therefore proceeded as follows. For each estimated p-quantile we constructed a series of data by $\{1_{\{r_i \le \hat{u}_p\}}\}_{i=1,\ldots,9500}$, where $\{r_i\}$ denotes the 9500 ruin probabilities. This generated five time series of binary data, counting 1 if the corresponding ruin probability is below $\hat{u}_p$ and 0 otherwise. For each of these binary time series we estimated their auto-covariance functions and calculated the variance of the average of the series using the initial monotone sequence estimator proposed by Geyer (1992) (see formula (3.3) and following comment on p. 477 and Gilks et al. (1996), pp. 247–248). The confidence limits were then obtained by formula (6) replacing $p(1-p)/n$ by the estimated variance of the averages.

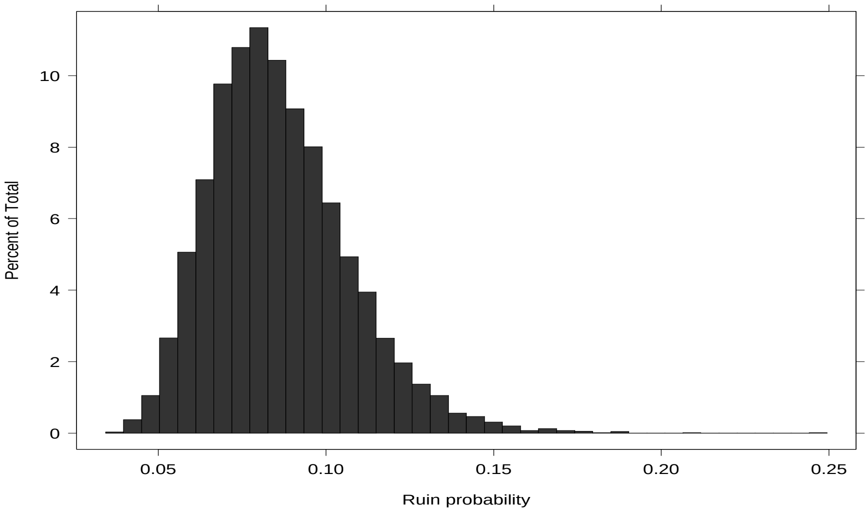

Figure 8: Histogram of ruin probabilities as obtained from the double sampler

4 Discussion

We have introduced an MCMC algorithm for simulating Markov jump processes which get absorbed at a fixed time. Using this and suitable prior distributions, we have constructed a Gibbs sampler which draws phase-type parameters from their posterior distribution. The method was extended to involve hyper-priors to improve mixing. The overall performance of the algorithm was good, and fits were comparable to those obtained through maximum likelihood estimation (EM algorithm) when these applied. A main advantage of the method is that the uncertainty of estimates of complex functionals of the phase-type distribution, such as e.g. ruin probabilities, can be easily represented.

The estimation of ruin probabilities is just one of a wide range of potential statistical applications. Extension to more complicated situations where several MCMC samplers work together should be possible, for example to a situation where the underlying arrival process is a Markov modulated Poisson process. The case of phase-type renewal arrivals was presented in the last example and involves two independent MCMC samplers working together.

Fitting higher order models is numerically feasible but may be statistically unstable. While the EM algorithm was significantly slowed by 15-20 phases and almost detained by 30, it continued to be precise. This is due to the fact that it is based on a numerical maximization by a deterministic algorithm and ignores the sampling uncertainty of the estimates. In principle, the MCMC algorithm presented can be run for quite high dimensions (100, say) without the program being detained, but they then become statistically unstable when the information in the data is too modest.

An interesting direction of further research would be the use of the MH algorithm as a part of a Monte Carlo EM algorithm (Wei & Tanner, 1990), which replaces the time consuming numerical integration involved in the E-step with a MH simulation of the corresponding means. This is an interesting alternative since the time-complexity of the MH is much lower than the numerical integration needed for the EM-algorithm, so the MCEM-algorithm might work faster in higher dimensions.

A stochastic EM-algorithm, see Ch. 15 of Gilks et al. (1996), where the E-step is replaced by a single MH-step will not work. Eventually samples will appear with no initiation in some state i, or no transitions from some states i to j, or no exit from some state i. The corresponding statistics will remain zero from that iteration and onwards due to the M-step. Thus the the limiting phase-type distribution of the stochastic EM-algorithm will be an exponential distribution.

The extension of the algorithm to an unknown number of active phases, possibly exploiting reversible jump processes (Green, 1995), would be of considerable interest.

Acknowledgement

A. Gonzales is indebted to Universidad de las Americas, Puebla, for supporting this work.

References

Asmussen, S. (1987). *Applied probability and queues.* John Wiley & Sons, Chichester and New York.

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

Asmussen, S. (2000b). Matrix-analytic models and their analysis. *Scand. J. Statist.*, **27**, 193–226.

Asmussen, S. & Bladt, M. (1992). Renewal theory and queueing algorithms for matrix-exponential distributions. *Matrix-analytic methods in stochastic models* (A. S. Alfa & S. Chakravarty, eds.), 313–341. Marcel Dekker, New York.

Asmussen, S. & Bladt, M. (1994). Poisson's equation for queues driven by a Markovian marked point process, *Queueing Systems*, **17**, 235–274.

Asmussen, S. & Bladt, M. (1996). Phase-type distributions and risk processes with state-dependent premiums, *Scand. Actuar. J.*, **96**, 19–36.

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

Asmussen, S. & Olsson, O. (1998). Phase-type distributions. *Encyclopedia of statistical sciences, update volume*, **2** (Kotz, Read, Banks eds.), John Wiley and Sons.

Asmussen, S. & Rolski, T. (1991). Computational methods in risk theory: A matrix-algorithmic approach. *Insurance: Mathematics and Economics*, **10**, 259–274.

Bernardo, J. M. & Smith, A. F. M. (1994). *Bayesian Theory.* John Wiley & Sons, Chichester and New York.

Bladt, M. (1996). The variance constant for the PH/PH/1 actual waiting time, *Ann. Appl. Probab.*, **6**, 766–777.

Diaconis, P. & Ylvisaker, D. (1979). Conjugate priors for exponential families. *Ann. Statist.*, **7**, 269–281.

Gelfand, A. E. & Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.*, **85**, 398–409.

Geman, S. & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721–41.

Geyer, C. J. (1992). Practical Markov chain Monte Carlo. *Statist. Sci.*, **7**, 457–511.

Geyer, C. J. & Thompson, E. A. (1992). Constrained Monte Carlo maximum likelihood for dependent data (with discussion). *J. Roy. Statist. Soc. Ser. B*, **54**, 657–99.

Gilks, W. R., Richardson, S. & Spiegelhalter, D. J. (1996). *Markov chain Monte Carlo in practice.* Chapman & Hall, London.

Green, P. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika,* **82**, 711–732.

Green, P. (2001). A primer on Markov chain Monte Carlo. In *Complex Stochastic Systems.* (Barndorff-Nielsen et al. (eds)), pp. 1–62. Chapman and Hall/CRC, London/Boca Raton.

Guitérrez-Pẽna, E. & Smith, A. F. M. (1995). Conjugate parametrizations for natural exponential families. *J. Amer. Statist. Assoc.*, **90**, 1347–1356.

Hald, A. (1952). *Statistical Theory with Engineering Applications.* Wiley, New York.

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

Hipp, C. (1989a). Efficient estimates for ruin probabilities. *Proc. fourth Prague symp. on asymptotic statistics* (P. Mandel & M. Huskova, eds.), 259–268.

Hipp, C. (1989b). Estimates and bootstrap confidence intervals for ruin probabilities. *Astin Bulletin*, **19**, 57–70.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, **21**, 1087–92.

Neuts, M. F. (1981). *Matrix-geometric solutions in stochastic models.* Johns Hopkins University Press, Baltimore and London.

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

Stanford, D. A. & Stroinski, K. J. (1994). Recursive methods for computing finite–time ruin probabilities for phase-type claim sizes. *Astin Bulletin*, **24**, 235–254.

Wei, G. C. G. & Tanner, M. A. (1990). A Monte Carlo implementation of the EM-algorithm and the poor man's augmentation algorithms. *J. Amer. Statist. Assoc.*, **85**, 699–704.

Address for correspondence: Steffen L. Lauritzen, Department of Mathematical Sciences, Aalborg University, Fredrik Bajers Vej 7G, DK-9220 Aalborg, Denmark. `steffen@math.auc.dk`