The estimation of phase–type related functionals through Markov chain Monte Carlo methodology

Mogens Bladt*
Antonio Gonzalez †
Steffen L. Lauritzen‡

June 22, 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 Markov chain Monte Carlo simulation to reconstruct the Markov jump processes underlying the phase–type distributions and a Gibbs sampling of the completed data in order to obtain a stationary sequence of phase–type probability measures from the posterior distribution of these given the data. We will not be able to fit the paramters of the phase–type distribution itself for various reasons, although we are able to compare performance to the method of maximum likelihood and we find a good coincidence. We illustrate the statistical potential of the method by some examples with ruin probabilities.

Keywords: MARKOV CHAIN MONTE CARLO, GIBBS SAMPLING, PHASE–TYPE DISTRIBUTION, RUIN PROBABILITIES.

---

*IIMAS-UNAM, A.P. 20-726, 01000 Mexico, D.F., Mexico.

†Benemérita Universidad Autónoma de Puebla and Universidad de las Americas, Santa Catarina Martir, C.P. 72820 Puebla, Mexico.

‡Aalborg University, Department of Mathematical Sciences, Fredrik Bajers Vej 7G, DK-9220 Aalborg, Denmark.

1 Introduction

In this article we address the following problem: how does a realization of a finite state–space Markov jump process known to be absorbed by some state at a certain time x look like? An algorithm for constructing such Markov jump processes is developed and applied to statistical issues concerning phase–type distributions. In particular, given data $x_1, \ldots, x_n$ from a phase–type distribution of known dimension and parameters, we can reconstruct some underlying Markov jump processes which generate such data, that is Markov jump processes which get absorbed exactly at times $x_1, x_2, \ldots, x_n$. The generation of such Markov jump processes is based on a Metropolis–Hastings algorithm.

We present applications of the algorithm addressing statistical issues of phase–type distribution. We sample distributions (measures) of Markov jump processes underlying phase–type data $x_1, \ldots, x_n$ with a Gibbs sampler in order to obtain a stationary sequence of phase–type distributions. Ergodicity of this sequence will then enable us to estimate quantities depending on the phase–type parameters. We will not be able to estimate phase–type parameters themselves using this method, as these are typically not identifiable, see Section 2 below. We show some examples where the functional of interest is the ruin probability of a risk–reserve process and where the claims are supposed to be i.i.d. phase–type distributed.

2 A brief review of 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, $\pi_i = \mathbb{P}(J_0 = i), \ \ i = 1, \ldots, p$. 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 distribution of the time until absorption, X, is a phase–type distribution generated by $\boldsymbol{\pi}$ and $\boldsymbol{T}$, and we write $X \sim \mathrm{PH}(\boldsymbol{\pi}, \boldsymbol{T})$. Note that $\boldsymbol{t} = -\boldsymbol{T}\boldsymbol{e}$ where $\boldsymbol{e}$ is the column vector of 1's. The density is given by

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

Phase–type distributions generalize mixtures of sums of exponential distributions as the latter class of distributions is contained in the class of phase–type distributions.

Exact solutions to many problems arising in applied probability like risk–theory and queueing theory can be obtained when distributions involved are assumed to be phase–type. Here we give a few examples. In renewal theory the renewal density $u(x)$ of an arrival at time x has the following form $u(x) = \boldsymbol{\pi} \exp((\boldsymbol{T} + \boldsymbol{t}\boldsymbol{\pi})x)\boldsymbol{t}$ when the interarrival 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 a Markov modulated Poisson process (see e.g. [2] and [3]) and claims being phase–type distributed. 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. One can also obtain asymptotic variances of time averages of such waiting times.

Being very attractive from a probabilistic point of view, phase–type distributions do present difficulties concerning statistical issues. They are heavily over-parametrized using in general $p^2 + p - 1$ freely varying parameters for a p–order phase–type distributions when in principle they could be represented (loosing the probabilistic interpretation which, however, in most applications is essential) by $2p - 1$ parameters. The representation of a phase–type distribution (that is the parameters which implies a certain density $f(x)$) is by no means unique. They are generalized mixture models which implies that the likelihood cannot be optimized explicitly. The EM–algorithm proposed by [4] works very well in lower dimensions, however, methods for statistical inference beyond estimation of parameters has not yet been proposed. For a review of other methods proposed for estimating phase–type distributions, see [4].

In general, due to non–uniqueness and over–parametrisation it does not make sense to construct confidence intervals or tests for specific parameters. When estimating parameters of a phase–type distribution one would often want to apply those estimated parameters for some specific purpose such as calculating a ruin probability related to the parameters. For this purpose it seems more natural to adopt a Bayesian approach, generating a stationary sequence of phase–type distributions (measures) using e.g. the Gibbs sampler and calculating the functional of interest (e.g. the ruin probability) in each iteration directly from the current phase–type distribution.

For further properties and general treatment of phase–type distributions, see e.g. [1], [2], [5] or [6].

3 Generating a Markov process underlying a phase–type distribution

Let X be a random variable with a phase-type distribution. Given a realization of X, x say, simulate a Markov jump process with intensity matrix $\mathbf{\Lambda}$ and initial distribution $(\boldsymbol{\pi}, 0)$ such that the Markov jump process gets absorbed exactly at time x.

The idea is to approximate the Markov jump process that gets absorbed at x with an arbitrary Markov jump process having an absorption time which exceeds x and apply a Metropolis-Hastings algorithm for accepting/rejecting proposals.

In general, the Metropolis–Hastings algorithm works as follows. Suppose that we wish to draw samples from an arbitrary distribution $\boldsymbol{\pi}(\cdot)$. Then we construct a (general state–space) Markov chain which has $\boldsymbol{\pi}$ as its stationary distribution, such that after a certain burn–in period, successive draws from a realization of this Markov chain result in draws from or close to the distribution $\boldsymbol{\pi}$. Such a chain $\{X_n\}$ can be constructed using a proposal distribution $q(X_{n+1}|X_n)$ proposing a new point X_{n+1} given the previous X_n. If we accept the proposal X_n with probability

$$\alpha(X_n, X_{n+1}) = \min(1, \frac{\boldsymbol{\pi}(X_{n+1})/q(X_{n+1}|X_n))}{\boldsymbol{\pi}(X_n)/q(X_n|X_{n+1}))}$$

and reject otherwise, in which case we let $X_{n+1} = X_n$, then the resulting Markov chain indeed has $\boldsymbol{\pi}$ as stationary distribution. When the proposal distribution does not depend on the previous point, that is when $q(X|Y) = q(X)$, then we talk about an independence sampler. In what follows we will construct an independence sampler for drawing Markov process which get absorbed at a known time.

Let $\{J_t\}$ denote the underlying 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. If $i \in \{1, 2, \ldots, n\}$ then

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

Also 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

$$\tilde{\boldsymbol{\pi}}_i := \mathbb{P}\{J_{x-} = i \mid X = x\} = \frac{q_i(x) t_i}{f_X(x)}, \tag{1}$$

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 under interest and $\mathbb{P}_x^* = \mathbb{P}\left(\cdot | X \geq x\right)$ 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 probability measure which we wish to calculate and $\mathbb{P}_x^*$ will serve as an approximating probability measure. The latter is easy to simulate by rejection, 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,

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

for $t, t < x-$ and the two measures only differs at $x-$. The distribution of $\mathbb{P}_x^*$ at time $x-$ is given by

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

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

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

where we have let $j^* = j_{x-}$. These fractions are the importance weights for an importance sampling based on sampling from $\mathbb{P}_x^*$ instead of $\mathbb{P}_x$. The importance sampling scheme indeed provides a solution to our problem, but we shall apply a more efficient way of simulation as the importance weights my be computationally burdensome.

An alternative is to construct a stationary Markov chain with the desired distribution $\mathbb{P}_x$ as equilibrium distribution. To do this we apply the Metropolis-Hastings 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 $(j_t, t < x)$, obtained by sampling from $\mathbb{P}_x^*$. This is done by rejecting if the ratio of importance weights is too small. More precisely, do the following steps

1. choose j from $\mathbb{P}_x^*$;
2. choose $U = u$ uniformly from the unit interval;
3. if
$$u \leq \min\{1, z(j, j')\},$$
then replace j' with j, else keep j'.

The quantity $z(j, j')$ is given by the ratio of importance weights above

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

The stationary distribution of this Markov chain of Markov processes will be that of $\mathbb{P}_x$. Thus applying the scheme above iteratively implies that the Markov chain of Markov processes converges to its stationary distribution, and any of the realizations created from a certain (however unknown) point and forward in the iterations can be seen as a Markov jump process which terminates at time x. Thus we proved:

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

0. *Draw a $Y_0 = \{j'_t, t < x\}$ from $\mathbb{P}^*_x = \mathbb{P}(\cdot|X \geq x)$ by simple acceptance/rejection. Set $i = 0$.*
1. *i=i+1*
2. *Draw a $\{j_t, t < x\}$ from $\mathbb{P}^*_x = \mathbb{P}(\cdot|X \geq x)$ by simple acceptance/rejection.*
3. *Draw $U \sim$ Uniform$[0, 1]$.*
4. *If $U \leq \min(1, t_{j_{x-}}/t_{j'_{x-}})$ then $X_i = \{j_t\}_{t<x}$, else $X_i = \{j'_t\}_{t<x}$.*
5. *GO TO 1.*

4 Gibbs sampling of phase–type distributions

4.1 The Gibbs sampler

From the reconstructed data of the previous section we are now able to sample iteratively from the posterior distribution (which has density proportional to the product of the prior density and the likelihood) using a Gibbs sampler obtaining eventually a stationary sequence of random phase–type measures which incorporates information from the data.

Choose constants α_i, β_{ij} and γ_i such that $(\boldsymbol{\pi}, \boldsymbol{T})$ has a prior density proportional to

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

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

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

Then we obtain a distribution that is conjugate and is easy to sample from since this distribution has π_i Dirichlet distributed, and t_i and t_{ij} are Gamma distributed with the same scale parameter.

Let $y_1, \ldots, y_N$ denote the observed data, that is the times to absorption of a phase–type distribution to be estimated. These absorption times are obviously incomplete data since we have no information of the behaviour of the underlying Markov process. A sufficient statistic for complete data is (see [4])

$$\boldsymbol{S} = ((B_i)_{i=1,\ldots,p}, (Z_i)_{i=1,\ldots,p}, (N_{ij})_{i=1,\ldots,p,j=0,\ldots,p,j\neq i}),$$

where B_i denotes the number of times the Markov process initiated in state i, let Z_i denotes the total time the Markov processes are in state i, and N_{ij} are the number of jumps from state i to j.

From a complete data set $\boldsymbol{x} = (x_1, \ldots, x_N)$, where we do not only observe the lifetimes but also the sample paths of the underlying Markov process, the likelihood function can be written as (see [4] for details)

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

The posterior distribution is then

$$\begin{aligned} p(\boldsymbol{\pi}, \boldsymbol{T} | \boldsymbol{x}) &\propto \phi(\boldsymbol{\pi}, \boldsymbol{T}) L(\boldsymbol{\pi}, \boldsymbol{T}; \boldsymbol{x}) \\ &= \prod_{i=1}^{p} \pi_i^{B_i + \alpha_i - 1} \prod_{i=1}^{p} t_i^{N_{i0} + \beta_{i0} - 1} e^{-t_1(\gamma_i + Z_i)} \prod_{i=1}^{p} \prod_{\substack{j=1 \\ j\neq i}}^{p} t_{ij}^{N_{ij} + \beta_{ij} - 1} e^{-t_{ij}(Z_i + \gamma_i)}. \end{aligned}$$

Thus with

$$\begin{aligned} \alpha_i^* &= \alpha_i + B_i \\ \beta_{ij}^* &= \beta_{ij} + N_{ij} \\ \gamma_i^* &= \gamma_i + Z_i, \end{aligned}$$

the initial distribution is Dirichlet, the exit intensities follow Gamma distributions $\Gamma(1/\gamma_i^*, \beta_{i0}^*)$ and the transition intensities are $\Gamma(1/\gamma_i^*, \beta_{ij}^*)$.

The Gibbs-sampler is now as follows:

Algorithm 4.1 *1. Draw (π, T) from the apriori distribution.*

2. Draw Y (complete data) as described above using the M-H procedure.

3. *Observe data B_i, Z_i and N_{ij} from the data $\boldsymbol{Y}$.*
4. *Calculate*

$$
\begin{aligned}
\alpha_i^* &= \alpha_i + B_i \\
\beta_{ij}^* &= \beta_{ij} + N_{ij} \\
\gamma_i^* &= \gamma_i + Z_i,
\end{aligned}
$$

4. *Draw (π, T) from the Posterior distribution ($*$-distribution)*
5. *GOTO 2;*

In this way we produce after a certain burn–in period a stationary sequence of distributions (measures) drawn from a class of phase–type distributions. This sequence can be used in several ways to obtain information about the parameters or related functionals of the parameters as we shall see later. The main purpose of the Gibbs sampler is not that of estimating the parameters $(\boldsymbol{\pi}, \boldsymbol{T})$. Even more, estimating the parameters directly from the Gibbs sampler is not straightforward as we can see from the following argument.

By ergodicity, the empirical average of the measures drawn in the Gibbs sampler converges to a measure of phase–type which is the object for estimation. As the points drawn in each iteration of the Gibbs sampler are measures, an empirical average of drawn measures can be represented as a mixture of densities. To find a representation of a phase–type distribution of such a mixture is non–trivial. It is *not* the phase–type distribution represented by the empirical average of the paramters. Furthermore, since the draws are dependent, one cannot guarantee a finite average to be of phase–type. Secondly, even if independence were assumed, the order of the representation would be n times larger than the order of the phase–type distribution to be estimated, where n is the number of terms in the empirical average. Thirdly, stationary draws may change representation through the iterations as representation are not unique. Hence, it hardly makes sense to look for a specific representation using this MCMC method. We shall in turn look for related functionals and their properties as illustrated in the following section.

5 Examples and improved mixing

Example 5.1 Consider 250 observations generated from a log–normal distribution LN$(-0.32, 0.8)$. This distributions has been considered earlier in

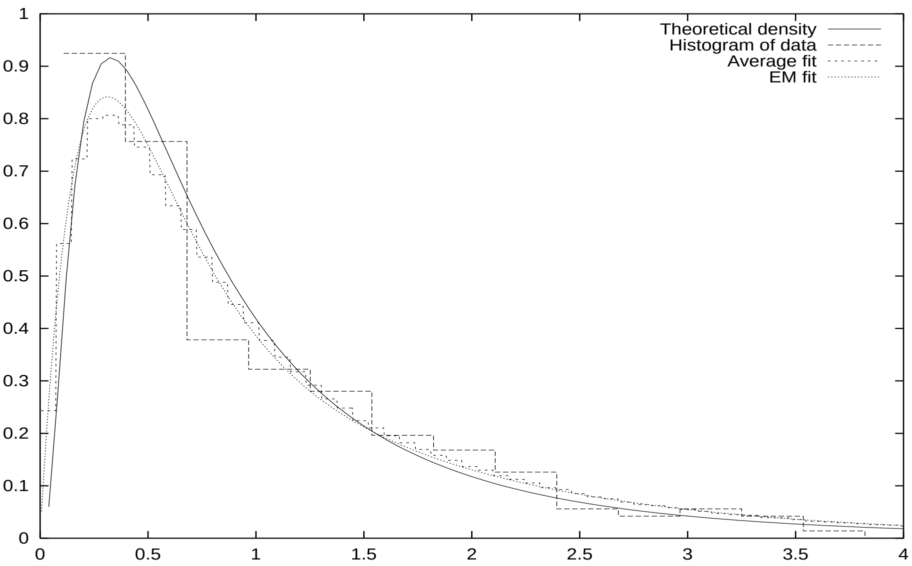

Figure 1 MCMC and EM fits to data

risk theory for adjusting insurance claim data. We adjust a fourth–order phase–type distribution by the EM–algorithm to data 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. In the EM–algorithm 23,000 iterations were necessary to obtain convergence. In the Gibbs sampler mixing appears to be rather quick, even after 20-30 iterations the draws are "on the right track", but we applied an initial burn-in period of 80 iterations. The average fit obtained by averaging the measures drawn are obtained through the last 1000 iterations, hence a total of 1080 iterations.

In Figure 1 we have displayed the fits of the EM–algorithm and the Gibbs sampler which seems to be quite alike. Also the tail behaviour are similar (Figure 2).

Next we estimate the ruin probability $\phi(u)$ (to be specified below) which 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 $\phi(u) = \mathbb{P}(V > u) = \mathbb{P}(W > u)$.

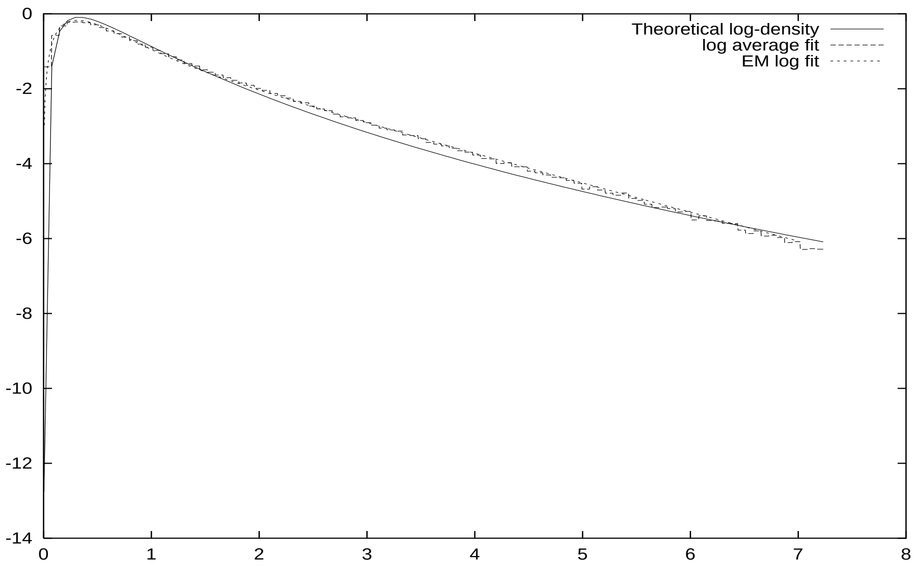

Figure 2 Averaged log-estimator vs. $f(x)$ and log-EM fit

Consider a risk–reserve process R_t defined by,

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

where insurance claims $U_1, U_2, \ldots$ arrive according to a Poisson process with intensity β, $N(t)$ is the number of claims which have arrived up to time t and it is assumed that the claims are independent and follow a p–dimensional phase–type distribution $PH(\boldsymbol{\pi}, \boldsymbol{T})$. Unit rate increase of capital between arrival of claims can be assumed without loss of generality.

Assume for simplicity that the arrival intensity is known and that claims data $X_1, X_2, \ldots, X_n$ are available. We are interested in calculating the ruin probability for this model. Using consecutive draws of $(\boldsymbol{\pi}, \boldsymbol{T})$'s of the Gibbs sampler and evaluating the corresponding ruin probability for each draw, we are provided with a set of ruin probabilities and by ergodicity we may then calculate confidence intervals, percentiles etc. for the ruin probability.

For fixed parameters $(\boldsymbol{\pi}, \boldsymbol{T})$ the ruin probability $\phi(u) = \mathbb{P}(\inf_{t>0} R_t < 0 | R_0 = u)$ that the risk–reserve process will ever become negative is given by

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

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

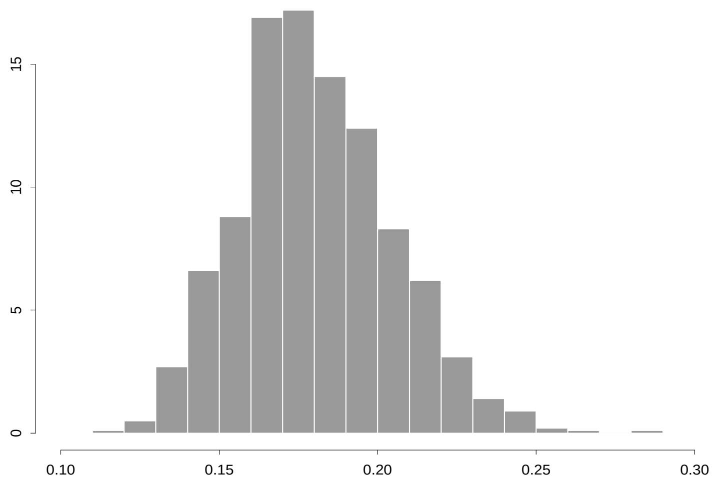

Fig 3 Histogram of ruin probabilities

We used arrival intensity $\beta = 0.3$ and $u = 1$ for the matter of illustration. The mean of the ruin probabilities is 0.180558. The ruin probability calculated from the EM–estimate is 0.180003. The estimated variance is 0.000579126. The distribution of the ruin probabilties is highly non–normal, and we obtained the following quantiles: 0.1377438, 0.1642835, 0.178307, 0.196096 and 0.2316088 for $2.5\%, 25\%, 50\%, 75\%$ and 97.5% respectively. A 95% empirical confidence interval is hence $[0.1377438, 0.2316088]$. A histogram of the ruin probabilities are given in Figure 3.

5.1 Improved mixing via hyper–parameters

Though the estimation presented above was quite succesful, there are problems with insufficient mixing in other cases. In order to improve the mixing we propose to introduce hyper-parameters so that the posteriors becomes less dependent on the initial conditions.

We propose to make the γ_i a hyper-parameter. Through the values of the γ_i's we can also control mean and variance of the β_{ij}. For this reason and for the reason of mathematical simplicity we do not impose further hyper–parameters on the β_{ij} nor the Dirichlet initial. Hence we propose the following model:

$$\begin{aligned}
\boldsymbol{\pi} &\sim \text{Dirichlet}(\boldsymbol{\alpha}) \\
t_i &\sim \Gamma(1/\gamma_i, \beta_{i0}) \\
t_{ij} &\sim \Gamma(1/\gamma_i, \beta_{ij}) \\
\gamma_i &\sim \Gamma(1/\theta_i, r_i) \\
\theta_i &\sim \Gamma(1/d, c).
\end{aligned}$$

Then the posterior is conjugate and the algorithm becomes the following:

Algorithm 5.1 *0. Initialize constants $\alpha_i, \beta_{ij}, r_i, c$ and d.*

1. *Draw $\boldsymbol{\pi}$, θ_i, γ_i and $\boldsymbol{T}$ according to the prior indicated above.*

2. *Draw samples $\boldsymbol{Y} = (\boldsymbol{b}, \boldsymbol{N}, \boldsymbol{z})$ using the Metropolis–Hastings procedure.*

3. *Draw $\boldsymbol{\pi}$, θ_i, γ_i and $\boldsymbol{T}$ from the full conditional:*

$$\begin{aligned}
\boldsymbol{\pi} &\sim \text{Dirichlet}(\boldsymbol{\alpha} + \boldsymbol{b}) \\
\theta_i &\sim \Gamma(1/(\gamma_i + d), r_i + c) \\
\gamma_i &\sim \Gamma(1/(\theta_i - t_{ii}), \sum_{j=0}^{p} \beta_{ij} + r_i) \\
t_i &\sim \Gamma(1/(\gamma_i + z_i), N_{i0} + \beta_{i0}) \\
t_{ij} &\sim \Gamma(1/(\gamma_i + z_i), N_{ij} + \beta_{ij})
\end{aligned}$$

4. *GOTO 2.*

We present three examples of ajustment: a mixture between lognormal and gamma, a Gamma distribution which is potentially demanding due to low probabilities for low values and a repetition of the log–normal fit with hyper–parameters, see figures 4-6.

5.2 A double sampler

As a final example of application we show how to use the MCMC–algoritms constructed above in a situation where the resulting functional is dependent on two data streams. We consider the risk reserve process as mentioned above, but instead of Poisson arrivals we assume that they occur according

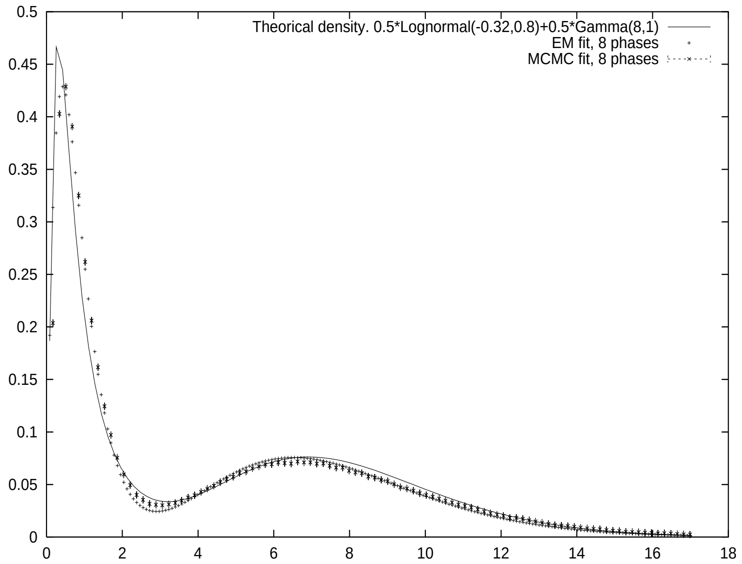

Fig 4 A mixture of lognormal and gamma

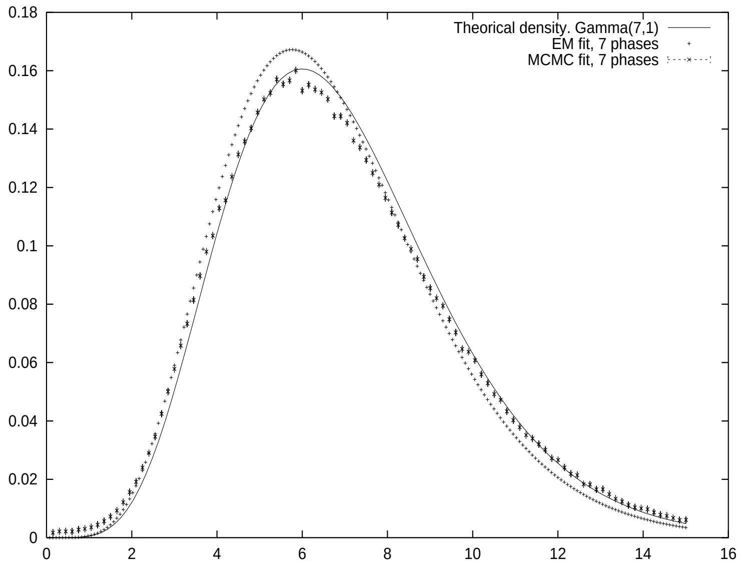

Fig 5 A 7–dimensional phase–type

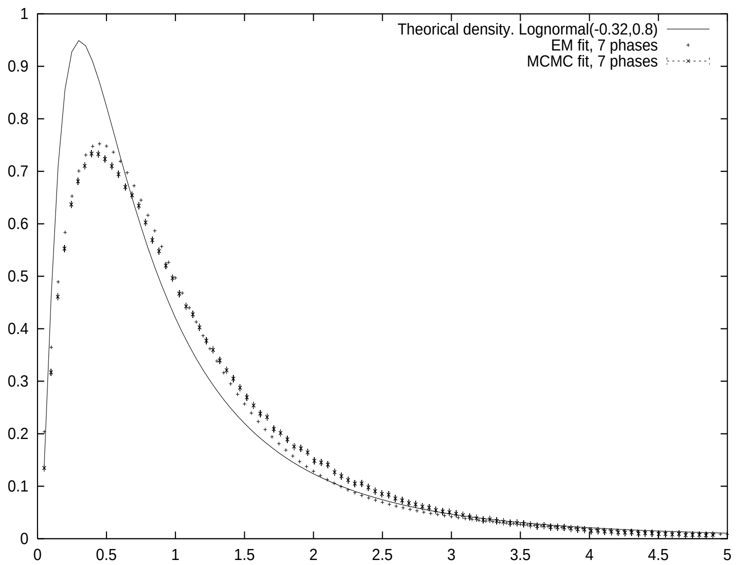

Fig 6 The lognormal distribution, again.

to a phase–type renewal process, that is, interarrival times are i.i.d. with a comon phase–type distribution $\mathrm{PH}(\boldsymbol{\alpha}, \boldsymbol{S})$ while the claims remain i.i.d with common distribution $\mathrm{PH}(\boldsymbol{\pi}, \boldsymbol{T})$ and independent of the arrival process. Then the ruin probability (infinite horizon) is given by (see [2] Proposition 4.4)

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

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

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

We are interested in sampling this ruin probability given we have data of inter-arrivals and claims respectively. 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 interarrival and the claims data. Then for each pair of parameters, we calculate the corresponding ruin probability. Iteration of this method obviously results in a stationary sequence of ruin probabilities.

We considered the following numerical example. Claims were 250 data from a log–normal(-.32,0.8) distribution, interarrivals being 250 data from an Erlang(3) distribution (the number of data in each series need not be the

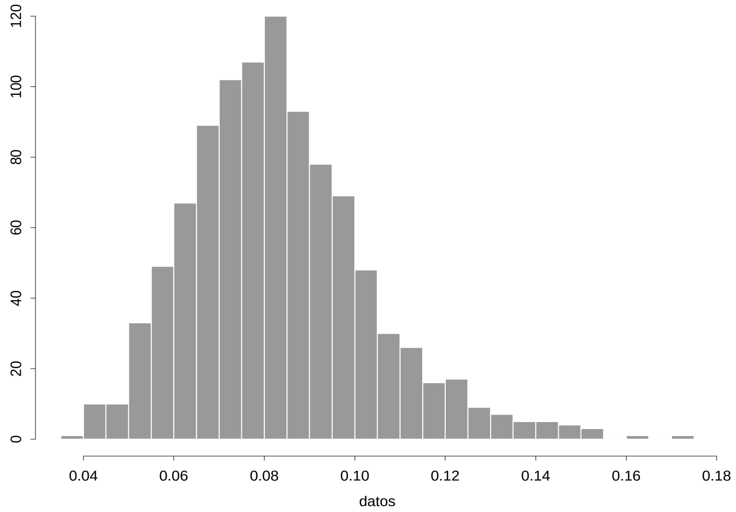

Fig 7 Histogram of ruin probabilities in the double sampler

same). We ran 1500 iterations on the double MCMC with 20 iteration 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 1000 iterations were considered as stationary. The ruin probability was evaluated at $u = 1.0$, and the results are shown in Figure 7 as a histogram.

The mean of the ruin probabilities is 0.08330805. Using an EM fit to both series resulted in a ruin probability of 0.0805026. The estimated variance is 0.0003990814 and we obtained the following quantiles: 0.05160402, 0.0694919, 0.08129265, 0.09457985 and 0.1303623 for $2.5\%, 25\%, 50\%, 75\%$ and 97.5% respectively. A 95% empirical confidence interval is hence $[0.05160402, 0.1303623]$.

6 Conclusion

We have introduced an MCMC–algorithm for simulating Markov jump processes which get absorbed at a certain known time. Using this information we can construct a Gibbs sampler which, in its stationary mode, draw phase–type parameters which contain information about data. The method is extended to involve hyper–parameters to improve mixing. We can check how well the algorithms mix by plotting averages of histograms that corresponds to the densities drawn in the Gibbs sampler. The overall performance is good, and fits are close to the ones obtained through maximum likelihood estimation (EM algorithm). In general hyper–parameters are necessary for an adequate mixing. The application presented to the estimation of ruin probabilities is just one of a wide range of potential statistical application. Also extention to more complicated situations where several MCMC samplers work together should be possible. For example the 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.

References

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

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

[3] S. Asmussen and M. Bladt (1996). Phase-type distributions and risk processes with state–dependent premiums, *Scandinavian Actuarial Journal*, **96**, 19-36.

[4] S. Asmussen, O. Nerman and M. Olsson (1996). Fitting phase–type distributions via the EM–algorithm. *Scandinavian Journal of Statistics*, **23**, 419–441.

[5] M. Bladt & S. Asmussen (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.

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