

A REVIEW ON PHASE-TYPE DISTRIBUTIONS AND THEIR USE IN RISK THEORY

BY

MOGENS BLADT

ABSTRACT

Phase-type distributions, defined as the distributions of absorption times of certain Markov jump processes, constitute a class of distributions on the positive real axis which seems to strike a balance between generality and tractability. Indeed, any positive distribution may be approximated arbitrarily closely by phase-type distributions whereas exact solutions to many complex problems in stochastic modeling can be obtained either explicitly or numerically. In this paper we introduce phase-type distributions and retrieve some of their basic properties through appealing probabilistic arguments which, indeed, constitute their main feature of being mathematically tractable. This is illustrated in an example where we calculate the ruin probability for a rather general class of surplus processes where the premium rate is allowed to depend on the current reserve and where claims sizes are assumed to be of phase-type. Finally we discuss issues concerning statistical inference for phase-type distributions and related functionals such as e.g. a ruin probability.

KEYWORDS

Phase-type distribution, ruin probability, EM-algorithm, Markov chain Monte Carlo.

1. INTRODUCTION

The aim of this paper is to serve as an introduction to the use of phase-type distributions in risk theory and at the same time to outline a recent line of research which includes statistical inference for phase-type distributions and related functionals such as ruin probabilities.

We start with a short bibliographic review. Though phase-type distributions can be traced back to the pioneering work of Erlang (1909) and to Jensen (1953), it was not until the late seventies that Marcel F. Neuts and co-workers established much of the modern theory available today (see Neuts (1981), Neuts (1989), Neuts (1995)). See also Asmussen (2003) for a more recent account. Most of the original applications were in the area of queueing theory but many

applications to risk theory can be found in Asmussen (2000). O’Cinneide (1990) studies theoretical properties of phase-type distributions, such as their characterization. Asmussen and Bladt (1996) generalizes risk models to situations with Markov modulated arrivals and to situations where the premium depends on the current reserve. Asmussen *et al.* (2002) provides an elegant algorithmic solution to the finite time-horizon ruin probability. In Aalen (1995) a proposal of application to survival analysis is outlined. For a more elementary textbook, which mainly draws examples from queueing theory, we refer to Latouche and Ramaswami (1999). Many results using phase-type methodology have been generalized into the broader class of matrix-exponential distributions (distributions with a rational Laplace transform), either by analytic methods (see Asmussen and Bladt (1997)) or more recently using a flow interpretation (see Bladt and Neuts (2003)).

Statistical inference for phase-type distributions is of more recent date, where likelihood estimation was first proposed by Asmussen *et al.* (1996) using the EM-algorithm whereas a Markov chain Monte Carlo (MCMC) based approach was suggested in Bladt *et al.* (2003). For an overview of earlier attempts to the estimation of phase-type distributions by methods other than maximum likelihood or MCMC, see Asmussen *et al.* (1996).

The rest of the paper is organized as follows. In Section 2 we provide the necessary background on the theory of Markov jump processes in order to introduce the concept of phase-type distributions in Section 3. In Section 4 we outline the method of probabilistic reasoning with phase-type distributions by establishing some of their basic properties. Section 5 introduces phase-type renewal theory and a first application in risk theory. In Section 6 we consider a model for the surplus in insurance where the premium is allowed to depend on the current reserve. Probabilistic arguments involving phase-type distributions allow us to establish a system of coupled differential equations, the solution of which is the ruin probability. The solution yields the exact solution to the problem though in practice the differential equations are solved by numerical methods. The last two Sections 7 and 8 are dedicated to the estimation and statistical inference for phase-type distributions and related functionals (such as e.g. the ruin probability in a surplus process).

The style of the paper is expository and technical details will occasionally be omitted but can be traced through appropriate references.

2. MARKOV JUMP PROCESS

Before defining phase-type distributions we shall recall some basic properties of Markov process with finite state spaces (also called continuous time Markov chains or Markov jump processes). Let $\{X(t)\}_{t \geq 0}$ be a Markov process which takes values in the state space $E = \{1, 2, \dots, n\}$. Then $\{X(t)\}_{t \geq 0}$ behaves in the following way. Let T_1, T_2, \dots denote the times where $\{X(t)\}_{t \geq 0}$ jumps from one state to another. Define for convenience $T_0 = 0$. Then the discrete time process $\{Y_n\}_{n \in \mathbb{N}}$, where $Y_n = X(T_n)$, is a Markov chain that keeps track of which states have been visited. Let $\mathbf{Q} = \{q_{ij}\}_{i,j \in E}$ denote its transition matrix. Furthermore,

if $Y_n = i$, then $T_{n+1} - T_n$ is exponentially distributed with a certain parameter λ_i . Also given $Y_0 = i_0, Y_1 = i_1, \dots, Y_n = i_n$, the holding times $T_1 - T_0, T_2 - T_1, \dots, T_{n+1} - T_n$ are independent. The latter property is referred to as conditional independence given $\{Y_n\}_{n \in \mathbb{N}}$.

Since the holding times $T_{n+1} - T_n$ are exponentially distributed with parameter λ_i given that $Y_n = i$, the conditional probability that there will be a jump in the process $\{X(t)\}_{t \geq 0}$ during the infinitesimal time interval $[t, t + dt)$ is $\lambda_i dt$. Given a jump at time t out of state i , the probability that the jump leads to state j is by definition q_{ij} . Hence for $j \neq i$, $\lambda_i dt q_{ij}$ is the probability of a jump from i to j during $[t, t + dt)$. Thus for $j \neq i$,

$$\lambda_{ij} = \lambda_i q_{ij}$$

is interpreted as the intensity of jumping from state i to j . Define $\lambda_{ii} = -\sum_{j \neq i} \lambda_{ij}$ and $\Lambda = \{\lambda_{ij}\}_{i,j \in E}$. The matrix Λ is called the intensity matrix or infinitesimal generator of the process. Let the transition probabilities of the Markov jump process be $p_{ij}^t = \mathbb{P}(X(t) = j | X(0) = i)$ and the corresponding transition matrix $P^t = \{p_{ij}^t\}_{i,j \in E}$. Then we have the following important relation between P^t and Λ ,

$$P^t = \exp(\Lambda t),$$

where $\exp(A)$ denotes the exponential of a matrix A defined in the usual way by series expansion,

$$\exp(A) = \sum_{n=0}^{\infty} \frac{A^n}{n!}.$$

Classification of states for Markov jump processes is as follows: a state i is recurrent (resp. transient) if i is recurrent (resp. transient) for the chain $\{Y_n\}_{n \in \mathbb{N}}$. A state i is absorbing if it is impossible to jump out of it again, that is if $q_{ij} = 0$ for all $j \neq i$ implying $\lambda_{ij} = 0$ for all j .

3. PHASE-TYPE DISTRIBUTIONS

We now let $\{X(t)\}_{t \geq 0}$ be a Markov jump process on the finite state-space $E = \{1, 2, \dots, p, p + 1\}$ where states $1, \dots, p$ are transient and state $p + 1$ is absorbing. Then $\{X(t)\}_{t \geq 0}$ has an intensity matrix on the form

$$\Lambda = \begin{pmatrix} \mathbf{T} & \mathbf{t} \\ \mathbf{0} & 0 \end{pmatrix}, \tag{3.1}$$

where \mathbf{T} is $p \times p$ dimensional matrix, \mathbf{t} is a p dimensional column vector (or $p \times 1$ dimensional matrix) and $\mathbf{0}$ is the p dimensional row vector of zeros. We shall make the following convention: unless otherwise stated matrices are denoted by boldface capital letters (Latin or Greek), boldface lowercase Latin letters refer to column vectors and lowercase boldface Greek letters refer to row vectors.

Since the intensities of rows must sum to zero, we notice that $\mathbf{t} = -\mathbf{T}\mathbf{e}$ where, $\mathbf{e} = (1, 1, \dots, 1)'$. The intensities t_i are the intensities by which the process jumps to the absorbing state and are referred to as exit rates (exit from the transient subset of states). Let $\pi_i = \mathbb{P}(X_0 = i)$, $i = 1, \dots, p$, $\mathbb{P}(X_0 = p + 1) = 0$ denote the initial probabilities. Notice that we are not allowed to initiate in state $p + 1$ (this would cause an atom at zero; extension to this case is, however, straightforward). Let $\boldsymbol{\pi} = (\pi_1, \dots, \pi_p)$ denote the initial distribution of $\{X(t)\}_{t \geq 0}$ over the transient states only.

Definition 3.1. *The time until absorption*

$$\tau = \inf\{t \geq 0 \mid X_t = p + 1\}$$

is said to have a phase-type distribution and we write

$$\tau \sim PH(\boldsymbol{\pi}, \mathbf{T}).$$

The set of parameters $(\boldsymbol{\pi}, \mathbf{T})$ is said to be a representation of the phase-type distribution. The dimension of $\boldsymbol{\pi}, p$, is said to be the dimension of the phase-type distribution.

In the following we give three examples of phase-type distributions.

Example 3.2. Let X_1, \dots, X_n be independent with $X_i \sim \exp(\lambda_i)$. Then $S = X_1 + \dots + X_n$ has a phase-type distribution with representation $\boldsymbol{\pi} = (1, 0, \dots, 0)$ (dimension n), and

$$\mathbf{T} = \begin{pmatrix} -\lambda_1 & \lambda_1 & 0 & \dots & 0 & 0 \\ 0 & -\lambda_2 & \lambda_2 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & 0 & -\lambda_n \end{pmatrix}, \quad \mathbf{t} = \begin{pmatrix} 0 \\ 0 \\ \dots \\ \lambda_n \end{pmatrix}.$$

Indeed the sum $X_1 + \dots + X_n$ may be interpreted as the time until absorption by a Markov jump process with n transient states which initiates in state 1 and always jumps to the next state in the sequence, up to state n from which it jumps to the absorbing state.

Since the sum S_n can be obtained by summing the X_i 's in any order we see that representations are by no means unique. Indeed we might permute the above states and obtain an alternative representation of the same distribution.

Example 3.3. The next distribution is known as hyper-exponential. Let X_1, \dots, X_n be independent and $X_i \sim \exp(\lambda_i)$ and let f_i denote the corresponding exponential density. Let

$$f = \sum_{i=1}^n \alpha_i f_i$$

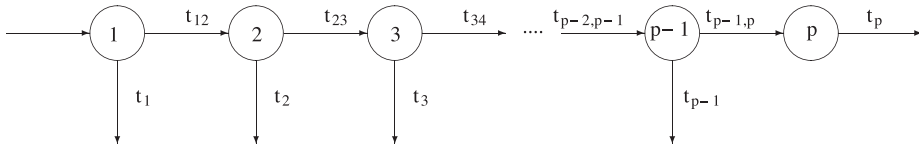


FIGURE 3.1. A flow diagram leading to a Coxian distribution.

where $\alpha_i > 0$, $\sum_{i=1}^n \alpha_i = 1$. Then f is phase-type with representation $\pi = (\alpha_1, \dots, \alpha_n)$,

$$T = \begin{pmatrix} -\lambda_1 & 0 & 0 & \dots & 0 & 0 \\ 0 & -\lambda_2 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & 0 & -\lambda_n \end{pmatrix}.$$

Example 3.4. Coxian distributions arise from the convolution of exponential distributions with a random (yet bounded) number of terms (called phases or stages). This can be interpreted as the time until absorption of a Markov jump process as represented by the flow diagram in Figure 3.1: starting from state 1, there is a total jump rate out of state 1 of $t_1 + t_{12}$. The probability of a jump to state 2 is $t_{12}/(t_1 + t_{12})$ and the probability of a jump to the absorbing state is $t_1/(t_1 + t_{12})$. This is equivalent to the jump rate from state 1 to state 2 being t_{12} while the jump rate to the absorbing state (the exit rate) is t_1 . All other states $j, j = 1, \dots, p - 1$ behave similarly, while the probability of jumping from state p to the absorbing state is 1.

If we let $\lambda_i = t_i + t_{i,i+1}$ for $i = 1, \dots, p - 1$ and $\lambda_p = t_p$ then the following choice of parameters yields a representation for the Coxian distribution described by the flow diagram in Figure 3.1.

$$\pi = (1, 0, 0, \dots, 0)$$

$$T = \begin{pmatrix} -\lambda_1 & t_{12} & 0 & \dots & 0 & 0 \\ 0 & -\lambda_2 & t_{21} & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & 0 & -\lambda_p \end{pmatrix}.$$

4. PROPERTIES OF PHASE-TYPE DISTRIBUTIONS

In this section we derive some basic properties of phase-type distributions by probabilistic arguments which, due to their importance later on, will be spelled out in details.

The following result is of main importance. With the notation (3.1), we have

$$\exp(\Lambda s) = \begin{pmatrix} \exp(Ts) & e - \exp(Ts) e \\ 0 & 1 \end{pmatrix}.$$

The proof uses the definition of the matrix exponential and the facts that $\mathbf{t} = -\mathbf{T}e$. We recall that $\exp(\mathbf{\Lambda}s)$ is the transition matrix \mathbf{P}^s of the Markov jump process $\{X(t)\}_{t \geq 0}$ and the important thing to notice here is that the restriction of \mathbf{P}^s to the transient states is given by $\exp(\mathbf{T}x)$. Hence we are able to compute transition probabilities $p_{ij}^s = \mathbb{P}(X(s) = j | X(0) = i) = \exp(\mathbf{T}s)_{ij}$ for $i, j = 1, \dots, p$.

We now take $\tau \sim \text{PH}(\boldsymbol{\pi}, \mathbf{T})$ and derive the density f of τ . The quantity $f(s)ds$ may be interpreted as the probability $\mathbb{P}(\tau \in [s, s + ds))$. If $\tau \in [s, s + ds)$, then the underlying Markov jump process $\{X(t)\}_{t \geq 0}$ must be in some transient state $j = 1, \dots, p$ at time s . If the process initiates in a state $i = 1, \dots, p$, the probability that $X(s) = j$ is then $p_{ij}^s = \exp(\mathbf{T}s)_{ij}$. The probability that the process $X(t)$ starts in state i is π_i by definition. If $X(s) = j$, the probability of a jump to the absorbing state $p + 1$ during $[s, s + ds)$ is $t_j ds$ (cfr. Section 2).

Conditioning on the initial state of the process and its state by time s we then get that

$$\begin{aligned} f(s)ds &= \mathbb{P}(\tau \in [s, s + ds)) \\ &= \sum_{i=1}^p \pi_i \sum_{j=1}^p p_{ij}^s t_j ds \\ &= \sum_{i=1}^p \sum_{j=1}^p \pi_i \exp(\mathbf{T}s)_{ij} t_j ds \\ &= \boldsymbol{\pi} \exp(\mathbf{T}s) \mathbf{t} ds. \end{aligned}$$

We have thus proved the following theorem:

Theorem 4.1. *If $\tau \sim \text{PH}(\boldsymbol{\pi}, \mathbf{T})$ the density f of τ is given by*

$$f(s) = \boldsymbol{\pi} \exp(\mathbf{T}s) \mathbf{t},$$

where $\mathbf{t} = -\mathbf{T}e$.

We could now obtain an expression for the distribution function by integrating the density, but we shall retrieve this formula by an even simpler argument. If F denotes the distribution function of τ , then $1 - F(s)$ is the probability that $\{X(t)\}_{t \geq 0}$ has not yet been absorbed by time s (i.e. $\tau > s$). But the event $\{\tau > s\}$ is identical to that of $\{X(s) \in \{1, 2, \dots, p\}\}$. Hence, by a similar conditioning argument as above, we get that

$$\begin{aligned} 1 - F(s) &= \mathbb{P}(X(s) \in \{1, 2, \dots, p\}) \\ &= \sum_{i=1}^p \pi_i \sum_{j=1}^p p_{ij}^s \\ &= \sum_{i=1}^p \sum_{j=1}^p \pi_i \exp(\mathbf{T}s)_{ij} \\ &= \boldsymbol{\pi} \exp(\mathbf{T}s) \mathbf{e}. \end{aligned}$$

Thus we have proved:

Theorem 4.2. *If $\tau \sim PH(\boldsymbol{\pi}, \mathbf{T})$, the distribution function F of τ is given by*

$$F(s) = 1 - \boldsymbol{\pi} \exp(\mathbf{T}s)\mathbf{e},$$

where $\mathbf{t} = -\mathbf{T}\mathbf{e}$.

As a corollary we may observe that the tail of a phase-type distribution decays exponentially which makes phase-type distributions thin-tailed.

We have the integration rule

$$\int \exp(\mathbf{T}s) ds = \mathbf{T}^{-1} \exp(\mathbf{T}s) = \exp(\mathbf{T}s) \mathbf{T}^{-1} \quad (4.1)$$

as a direct consequence of

$$\frac{d}{ds} \exp(\mathbf{T}s) = \mathbf{T} \exp(\mathbf{T}s) = \exp(\mathbf{T}s) \mathbf{T} \quad (4.2)$$

and that \mathbf{T} is invertible, being a sub-exponential matrix. Of course (4.2) is also valid for $\boldsymbol{\Lambda}$ but (4.1) is not since $\boldsymbol{\Lambda}$ is singular. Using (4.1) the following analytic properties of phase-type distributions may easily be verified.

Theorem 4.3. 1. *The n -th ($n \geq 1$) moment of τ is given by $\mathbb{E}(\tau^n) = (-1)^n n! \boldsymbol{\pi} \mathbf{T}^{-n} \mathbf{e}$.*
 2. *The moment generating function of τ is given by $\mathbb{E}(e^{s\tau}) = \boldsymbol{\pi}(-s\mathbf{I} - \mathbf{T})^{-1} \mathbf{t}$, where \mathbf{I} denotes the identity matrix of dimension p .*

Apart from being mathematically tractable, phase-type distributions have the additional appealing feature of forming a dense class of distributions within the class of distribution on the positive real axis, that is, for any distribution μ on the positive real axis there exists a sequence of phase-type distributions which converges weakly to μ (see (2003) for details). In other words, phase-type distributions may approximate arbitrarily closely any distribution with support on the positive reals. This means that for thin-tailed distributions one may assume without (too much) loss of generality that distributions are of phase-type. For heavy tailed distributions more care should be taken. Though in principle phase-type distributions are able to approximate also heavy tailed distributions arbitrarily well, the approximations will always be bad in the tails, and the tail behavior is of crucial importance in many situations.

5. PHASE-TYPE RENEWAL THEORY

Consider a phase-type renewal process $\{N(t)\}_{t \geq 0}$, that is, a renewal process where the interarrival times have a phase-type distribution $\sim PH(\boldsymbol{\pi}, \mathbf{T})$. For the subsequent applications we are in particular interested in the renewal density u of $\{N(t)\}_{t \geq 0}$, which has the interpretation that $u(s)ds$ is the probability of a

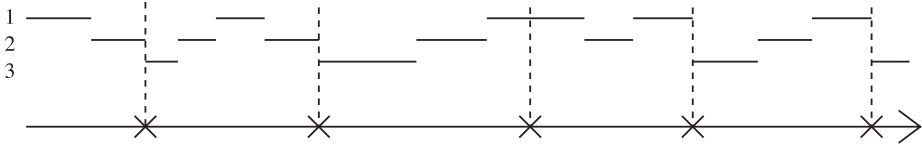


FIGURE 5.1. The sample paths of the Markov jump process constructed by concatenating the Markov processes underlying the phase-type distributions. The “crosses” on the time axis denote arrival times. As we can see there are two type of jumps in the constructed Markov process: as a result of a jump in an underlying phase-type process or as a result of an arrival, where the phase-type process exits from one state and a new phase-type process initiates in a new state (possibly the same).

renewal (an arrival) during the infinitesimal time interval $[s, s + ds)$. Formally, if F and f are the distribution function and density of the interarrival times respectively, then

$$U(s) = \sum_{n=0}^{\infty} F^{*n}(s), \quad u(s) = U'(s) = \sum_{n=1}^{\infty} f^{*n}(s)$$

where $*$ denotes convolution.

By concatenating the sample paths of the Markov jump processes underlying the phase-type distributions between arrivals we obtain a new Markov jump process $\{J(s)\}_{s \geq 0}$ on the state space $\{1, 2, \dots, p\}$ with intensity matrix $\Gamma = \{\gamma_{ij}\}_{i,j=1,\dots,p} = \mathbf{T} + \mathbf{t}\boldsymbol{\pi}$ (see Figure 5.1).

More precisely, we let $\{X^i(t)\}_{0 \leq t < T_i}$ denote the Markov jump process generating the i th inter-arrival time T_i observed only up to the time of absorption. Then we let $J(s) = X^i(s - T_0 - \dots - T_{i-1})$ if $s \in [T_{i-1}, T_i)$. Then $\{J(s)\}_{s \geq 0}$ is a Markov jump process with state space $\{1, 2, \dots, p\}$. For this process, a transition from i to j can take place in either of two mutually exclusive ways: either through a process $\{X^i(t)\}_{0 \leq t < T_i}$ jumping from i to j or by such a process exiting from state i (to the absorbing state) and the next process, $\{X^{i+1}(t)\}_{0 \leq t < T_{i+1}}$, initiating in state j . The probability of the former is $t_{ij}ds$ while the latter has probability $t_i ds \pi_j$. Hence $\gamma_{ij}ds = t_{ij}ds + t_i ds \pi_j$, proving that $\Gamma = \mathbf{T} + \mathbf{t}\boldsymbol{\pi}$.

The transition matrix of $\{J(t)\}_{t \geq 0}$ is hence given by $\mathbf{P}^s = \exp((\mathbf{T} + \mathbf{t}\boldsymbol{\pi})s)$ which is the key to finding an expression for the renewal density u . At time s , the process $\{J(t)\}_{t \geq 0}$ will develop through some process $\{X^i(t)\}_{t \geq 0}$. There is a renewal at time s if and only if the phase-type process $\{X^i(t)\}_{t \geq 0}$ makes a transition to the absorbing state during $[s, s + ds)$. Conditioning on the initial state of $\{X^1(t)\}_{0 \leq t < T_1}$ and the state of the process $\{J(t)\}_{t \geq 0}$ at time s we get that

$$\begin{aligned} u(s)ds &= \sum_{i=1}^p \pi_i \sum_{j=1}^p p_{ij}^s t_j ds \\ &= \sum_{i=1}^p \sum_{j=1}^p \pi_i \exp((\mathbf{T} + \mathbf{t}\boldsymbol{\pi})s)_{ij} t_j ds \\ &= \boldsymbol{\pi} \exp((\mathbf{T} + \mathbf{t}\boldsymbol{\pi})s) \mathbf{t} ds, \end{aligned}$$

which proves the following theorem:

Theorem 5.1. *The renewal density u of a renewal process with interarrival times which are $PH(\boldsymbol{\pi}, \mathbf{T})$ is given by*

$$u(s) = \boldsymbol{\pi} \exp((\mathbf{T} + \mathbf{t}\boldsymbol{\pi})s) \mathbf{t}, \quad s \geq 0.$$

A delayed renewal process is one where the waiting time of the first arrival has a different distribution than the distribution of the rest of the interarrival times. If the waiting time of the first arrival is distributed $PH(\boldsymbol{\beta}, \mathbf{T})$ while the remaining interarrival times have a distribution $PH(\boldsymbol{\pi}, \mathbf{T})$, then the renewal density u_0 of the delayed process is given by

$$u_0(s) = \boldsymbol{\beta} \exp((\mathbf{T} + \mathbf{t}\boldsymbol{\pi})s) \mathbf{t},$$

which follows immediately by noting that the initial distribution of $\{J(s)\}_{s \geq 0}$ is now $\boldsymbol{\beta}$.

The overshoot or the residual waiting time is another example of distributions which are of prime interest in applications. The residual waiting time $\zeta(s)$ at time s is defined as the time until the next arrival. The distribution of $\zeta(s)$ is easily obtained by the following argument. Since $\mathbf{T} + \mathbf{t}\boldsymbol{\pi}$ is the intensity matrix of $\{J(s)\}_{s \geq 0}$ then the distribution of $J(s)$ is given by $\boldsymbol{\beta}_s = \boldsymbol{\pi} \exp((\mathbf{T} + \mathbf{t}\boldsymbol{\pi})s)$ in the non-delayed case. Since $\{J(s)\}_{s \geq 0}$ moves according to \mathbf{T} until the next arrival, we then conclude that $\zeta(s) \sim PH(\boldsymbol{\beta}_s, \mathbf{T})$.

A renewal process is called terminating if the interarrival distribution is defective, that is, if $\int_0^\infty dF(s) = F(\infty-) < 1$. This is usually interpreted as the distribution having an atom at $+\infty$. We notice that all arguments above hold also for terminating renewal processes, and hence all the results previously established also hold true.

We now consider an immediate application of the phase-type renewal theory to the following model for the surplus. Let R_t be the surplus process given by

$$R_t = u + pt - \sum_{n=0}^{N(t)} U_n,$$

where u is the initial capital, p the premium rate, $\{N(t)\}_{t \geq 0}$ a Poisson processes with intensity $\beta > 0$ and U_1, U_2, \dots i.i.d. claims with distribution $PH(\boldsymbol{\pi}, \mathbf{T})$. We assume that R_t drifts toward $+\infty$ which amounts to $-\beta\boldsymbol{\pi}\mathbf{T}^{-1}\mathbf{e} < 1$ (see Asmussen (2000) p. 227 for details).

We are interested in calculating the probability of ruin for an infinite time horizon,

$$\psi(u) = \mathbb{P}(\inf_{0 \leq s < \infty} R_s < 0 \mid R_0 = u). \tag{5.1}$$

Figure 6.1 shows a sample path of such a process, though with a more general premium income $p(R_t)$ rather than pt . In this case we see that the process

that is underlying the concatenated descending ladder heights is a terminating phase-type renewal process with interarrival distribution $\text{PH}(\boldsymbol{\pi}_-, \mathbf{T})$ for some defective distribution $\boldsymbol{\pi}_-$.

It is clear that $(\boldsymbol{\pi}_-)_i$ is the probability that a Markov jump process underlying the phase-type claims downcrosses level u in state i when the surplus process jumps to a level below u for the first time. Since there is a positive probability of $\{R_t\}_{t \geq 0}$ never jumping to a level below u , the distribution $\boldsymbol{\pi}_-$ is defective (does not sum to 1). Since ruin happens if and only if the terminating renewal process is in some state $1, 2, \dots, p$ by time u (see Figure 6.1) we conclude that

$$\psi(u) = \boldsymbol{\pi}_- e^{(\mathbf{T} + \mathbf{t}\boldsymbol{\pi}_-)u} \mathbf{e}.$$

Indeed, $(\boldsymbol{\pi}_- \exp((\mathbf{T} + \mathbf{t}\boldsymbol{\pi}_-)u))_i$ is the probability that the defective renewal process is in state i by time u . Summing over all the states then gives the result. It is a bit more involved to prove that $\boldsymbol{\pi}_- = -\beta \boldsymbol{\pi} \mathbf{T}^{-1} / p$; see Asmussen (2000) for details.

6. SURPLUS PROCESSES OF MORE GENERAL KIND

In this section we shall consider a surplus process where arrivals occur according to a Poisson process with rate $\beta > 0$ and the claim sizes are i.i.d. phase-type $\text{PH}(\boldsymbol{\pi}, \mathbf{T})$. Between jumps the surplus process R_t moves according to the (deterministic) differential equation

$$\frac{d}{dt} R_t = p(R_t)$$

for some “well behaved” premium function p . Hence we have a model where the premium depends on the current reserve. If $p(x) = p$ constant we are back at the “classical” risk model.

We are again interested in calculating the infinite time horizon ruin probability (5.1). The idea is essentially the same as for the case of a constant premium function. We consider the process obtained by projecting the processes underlying the descending ladder height on the vertical axis. The difference from earlier is that this process is no longer a renewal process and we approach the problem in a slightly different way. We shall again be looking at downcrossing probabilities. We consider the first time the surplus process jumps to a level below its initial level u . Such a jump is evidently caused by a claim and we let $v_i(u)$ denote the probability that the underlying Markov jump process of such a claim downcrosses level u in state i (see Figure 6.1). We notice that $v_i(u)$ corresponds to $\boldsymbol{\pi}_-$ in the case of a constant premium function, but with a non-constant premium function this probability will in general depend on the initial capital u .

By a conditioning argument similar to the linear case we obtain that

$$v_i(u) = \beta dt \pi_i + (1 - \beta dt) \sum_{j=1}^p v_j(u + p(u)dt) (\delta_{ij} + t_{ji} p(u)dt + t_j p(u)dt v_i(u)).$$

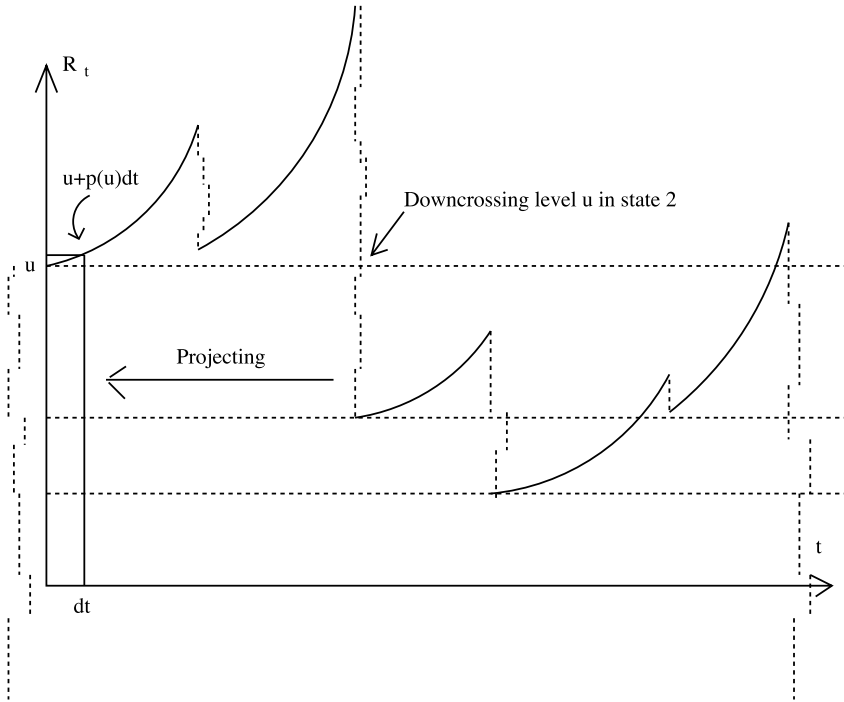


FIGURE 6.1. A surplus process with non-constant premium function and phase-type claims. At jump times the Markov processes underlying the phase-type claims are indicated as dashed lines with three states moving downwards.

Here we condition on the event of an arrival in a small time interval $[0, dt]$ the probability of which is βdt . In case of an arrival, the probability of downcrossing level u in state i is simply π_i . If there are no arrivals, $R_{dt} = u + p(u)dt$ and conditioning on this new level being downcrossed in state j , the probability of which is $v_j(u + p(u)dt)$, either the process which downcrosses level $u + p(u)dt$ in state j continues and downcrosses level u in state i with probability $p(u)dt t_{ji}$ (if $j = i$ $1 + t_{ii}p(u)dt$ is the probability of no change of state), or the downcrossing process exits between $u + p(u)dt$ and u with probability $t_j p(u)dt$, in which case the probability of downcrossing level u amounts to $v_i(u)$.

Then we Taylor-expand $v_j(u + p(u)dt) = v_j(u) + v'_j(u)p(u)dt$, and inserting this expression and putting $(dt)^k = 0$ for $k > 1$, we obtain the following system of non-linear differential equations:

$$-p(u)v'_i(u) = \beta\pi_i + v_i(u)\left(\sum_{j=1}^p v_j(u)t_j p(u) - \beta\right) + \sum_{j=1}^p v_j(u)t_{ji}p(u), \quad i = 1, \dots, p. \quad (6.1)$$

To find the ruin probability we now look at the terminating descending ladder process $\{I_t\}_{t \geq 0}$ which initiates (time zero) at level u . Ruin occurs if and only if

I_t reaches time u (i.e. level zero). Let $\gamma_i(t) = \mathbb{P}(I_t = i)$ and consider $\gamma_i(t + dt)$. Conditioning on the state of I_t at time t we easily get that

$$\gamma_i(t + dt) = \sum_{j=1}^p \gamma_j(t) (\delta_{ij} + t_{ji}dt + t_j dt v_i(u - t))$$

where the term $t_{ji}dt$ corresponds to the process going from state j to i generated by one of the ladder processes (if $i = j$ again $1 + t_{ii}dt$ is the probability that a ladder process being in state i will not change state during $[t, t + dt)$). The term $t_j dt v_i(u - t)$ corresponds to the probability that at time t (level $u - t$) a ladder process will exit from state j and eventually later downcross level $u - t$ in state i .

Again by Taylor expansion $\gamma_i(t + dt) = \gamma_i(t) + \gamma'_i(t)dt$ we obtain the following system of differential equations

$$\gamma'_i(t) = \sum_{j=1}^p \gamma_j(t) v_i(u - t) t_j + \sum_{j=1}^p \gamma_j(t) t_{ji}, \quad i = 1, \dots, p. \quad (6.2)$$

The initial condition for (6.2) is obviously $\gamma_i(0) = v_i(u)$, whereas for (6.1) the initial condition is a non-trivial matter. However, if the premium function p was such that it would be constant (equal to c , say) above a certain level v then the corresponding surplus process R_t^v would be linear above this level and at least we would know that $v_i^v(v) = -\beta \boldsymbol{\pi} \mathbf{T}^{-1} / c$, where $v_i^v(u)$ denotes the corresponding downcrossing probability for the modified surplus process. Letting $v \rightarrow \infty$ then $v_i^v(u) \rightarrow v_i(u)$. In practice one would linearize at e.g. $v = 2u, 3u, 4u, \dots$ and solve for $v_i^v(u)$ until convergence is obtained.

Thus (6.1) and (6.2) constitutes a non-linear system of coupled differential equations, which may be effectively solved by a numerical procedure like for instance a fourth order Runge-Kutta method (see e.g. Press *et al.* (1992)). Explicit solutions can in general not be obtained. At last we are able to calculate the ruin probability by noting that ruin occurs if and only if $I_u \in \{1, 2, \dots, p\}$ yielding

$$\psi(u) = \sum_{i=1}^p \gamma_i(u).$$

An important extension of the above model can be obtained by generalizing the arrival process to a Markov modulated Poisson process where the rates of the Poisson process depend on an underlying Markovian environment, see Asmussen and Bladt (1996) for details. Also an extension of the model to claim sizes having a Matrix-exponential distribution (distributions with rational Laplace transforms) has been proved in Bladt and Neuts (2003).

7. MAXIMUM LIKELIHOOD ESTIMATION

We consider data x_1, x_2, \dots, x_n which we might think of as claim sizes. We now suppose that these data are generated by i.i.d. phase-type distributed random variables of dimension p and representation $\text{PH}(\boldsymbol{\pi}, \mathbf{T})$. All we observe are the

times until absorption of the underlying Markov jump processes and not the underlying trajectories. The data are hence incomplete and in the following we shall describe a method for calculating the maximum likelihood estimator using the EM-algorithm. We follow Asmussen *et al.* (1996) which may be consulted for further details.

Suppose that we observed complete data such that j_1, \dots, j_n are the sample paths of the underlying Markov processes generating the absorption times $X_1 = x_1, \dots, X_n = x_n$. Then the Likelihood function for $\mathbf{j} = (j_1, \dots, j_n)$ is given by

$$p(\mathbf{j} | \boldsymbol{\pi}, \mathbf{T}) = \prod_{i=1}^p \pi_i^{B_i} \prod_{i=1}^p t_i^{N_{i0}} \exp(-t_i Z_i) \prod_{i=1}^p \prod_{\substack{j=1 \\ j \neq i}}^p t_{ij}^{N_{ij}} \exp(-t_{ij} Z_i),$$

where B_i is the number of times the processes of \mathbf{j} initiates in state i , N_{i0} is the number of jumps the processes perform from state i to the absorbing state $p + 1$, N_{ij} is the total number of jumps from state i to state j and Z_i is the total time the Markov processes are in state i . The maximum likelihood estimators for $\boldsymbol{\pi}$ and \mathbf{T} are given by

$$\hat{\pi}_i = \frac{B_i}{n}, \quad \hat{t}_{ij} = \frac{N_{ij}}{Z_i}, \quad i \neq j, \quad \hat{t}_i = \frac{N_{i0}}{Z_i}, \quad \hat{t}_{ii} = -\sum_{j \neq i} \hat{t}_{ij} - \hat{t}_i, \quad i, j = 1, \dots, p. \quad (7.1)$$

The EM-algorithm is an iterative procedure that maximizes in each step the conditional expected value of the log-likelihood function given incomplete data. Hence there are essentially two steps involved in each iteration: the calculation of the conditional expectation of the log-likelihood given absorption times (the E-step) and the maximization (the M-step). Taking the logarithm of the likelihood function gives

$$\begin{aligned} \log p(\mathbf{j} | \boldsymbol{\pi}, \mathbf{T}) &= \sum_{i=1}^p B_i \log(\pi_i) + \sum_{i=1}^p N_{i0} \log(t_i) - \sum_{i=1}^p t_i Z_i \\ &+ \sum_{i=1}^p \sum_{\substack{j=1 \\ j \neq i}}^p N_{ij} \log(t_{ij}) - \sum_{i=1}^p \sum_{\substack{j=1 \\ j \neq i}}^p t_{ij} Z_i. \end{aligned}$$

Hence the log-likelihood is linear in the sufficient statistics B_i, N_{i0}, N_{ij}, Z_i for $i, j = 1, \dots, p, i \neq j$ so the calculation of the conditional expectation of the log-likelihood, given $\mathbf{X} = (X_1, \dots, X_n)$, reduces to the calculation of the conditional expectation of each of the B_i, N_{i0}, N_{ij} for $i, j = 1, \dots, p, i \neq j$. Let $B_i^{(k)}$ be the indicator for whether the k th Markov jump process j_k initiated in state i , $Z_i^{(k)}$ the total time j_k spent in state i , $N_{ij}^{(k)}$ the number of jumps from state i to j in j_k and $N_{i0}^{(k)}$ the indicator for whether the exit to the absorbing state in j_k was caused by a jump from state i . Then $B_i = \sum_{k=1}^n B_i^{(k)}, i = 1, \dots, p, Z_i = \sum_{k=1}^n Z_i^{(k)}, i = 1, \dots, p, N_{ij} = \sum_{k=1}^n N_{ij}^{(k)}, i = 1, \dots, p, j = 0, \dots, p$, and it is sufficient to calculate the conditional expectations of the statistics with superscript k . In Asmussen *et al.* (1996) it is shown that

$$\mathbb{E}_{(\boldsymbol{\pi}, \boldsymbol{\tau})} \left(B_i^{(k)} \mid X_k = x_k \right) = \frac{\pi_i \mathbf{e}'_i \exp(\mathbf{T}x_k) \mathbf{t}}{\boldsymbol{\pi} \exp(\mathbf{T}x_k) \mathbf{t}}, \quad i = 1, \dots, p,$$

$$\mathbb{E}_{(\boldsymbol{\pi}, \boldsymbol{\tau})} \left(Z_i^{(k)} \mid X_k = x_k \right) = \frac{\int_0^{x_k} \boldsymbol{\pi} \exp(\mathbf{T}u) \mathbf{e}_i \mathbf{e}'_i \exp(\mathbf{T}(x_k - u)) \mathbf{t} du}{\boldsymbol{\pi} \exp(\mathbf{T}x_k) \mathbf{t}}, \quad i = 1, \dots, p,$$

$$\mathbb{E}_{(\boldsymbol{\pi}, \boldsymbol{\tau})} \left(N_{ij}^{(k)} \mid X_k = x_k \right) = \frac{t_{ij} \int_0^{x_k} \boldsymbol{\pi} \exp(\mathbf{T}u) \mathbf{e}_i \mathbf{e}'_j \exp(\mathbf{T}(x_k - u)) \mathbf{t} du}{\boldsymbol{\pi} \exp(\mathbf{T}x_k) \mathbf{t}}, \quad i, j = 1, \dots, p, \quad i \neq j,$$

and

$$\mathbb{E}_{(\boldsymbol{\pi}, \boldsymbol{\tau})} \left(N_{i0}^{(k)} \mid X_k = x_k \right) = \frac{\boldsymbol{\pi} \exp(\mathbf{T}x_k) \mathbf{e}_i t_i}{\boldsymbol{\pi} \exp(\mathbf{T}x_k) \mathbf{t}}, \quad i = 1, \dots, p.$$

We see that the E-step above essentially involves matrix-exponentials and an integral of matrix-exponentials which cannot be reduced further. Effective methods for calculating matrix-exponentials are given in Moler and Van Loan (1978) including the Runge-Kutta method, where the matrix-exponential is recognized as a solution to a linear system of differential equations. An alternative effective method, given in Neuts (1995) p. 232, is the uniformization method which works well even in high dimensions. In Asmussen *et al.* (1996) a system of differential equations of dimension $p(p+1)$ is solved using a fourth order Runge-Kutta method.

The EM-algorithm now works as follows. Given any initial parameter $(\boldsymbol{\pi}_0, \mathbf{T}_0)$ we calculate the conditional expected values above and plug them in as the sufficient statistics for complete data. Then we calculate the maximum likelihood estimates (7.1) and repeat the procedure with $(\boldsymbol{\pi}_0, \mathbf{T}_0)$ replaced by the maximum likelihood estimator. This goes on until convergence of the maximum likelihood estimates. While the EM algorithm always converges, it does not necessarily converge to the maximum likelihood estimator as it can get trapped in a local maximum. If that happens a more elaborated search may have to be established by a suitable variation of the initial parameters $\boldsymbol{\pi}_0$ and \mathbf{T}_0 .

In Oakes (1999) it is described how to obtain confidence intervals directly as a by-product of the EM algorithm. It involves derivatives with respect to the parameters of the conditional expectation of the log-likelihood function. Such derivatives are not readily available but can be obtained numerically by variation of the parameters in a small neighborhood around the maximum likelihood estimate and a recalculation of the expected log-likelihood in the neighboring points.

To which extent it makes sense at all to talk about confidence intervals for phase-type distribution is a controversial issue due to the problems of identifiability and over-parameterization.

8. MARKOV CHAIN MONTE CARLO BASED INFERENCE

In this section we present an alternative method for fitting phase-type distributions based on Bladt *et al.* (2003). This method can only estimate functionals of phase-type distributions which are invariant under different representations.

This means that we shall not be able to estimate the parameters themselves, but only quantities such as density function, quantiles or ruin probabilities. The method we present may either be viewed as Bayesian or as penalized likelihood.

The key idea is to produce a stationary sequence of measures (distributions) which has a stationary distribution being that of a conditional phase-type distribution given the data. Let \mathbf{J} denote a Markov jump process with intensity matrix Λ of the form (3.1) and let $\boldsymbol{\alpha} = (\boldsymbol{\pi}, 0)$ denotes its initial distribution. We then construct a stationary Markov chain of probability measures (general state-space type of chain) which has as stationary distribution that of $(\boldsymbol{\pi}, \mathbf{T}, \mathbf{J})$ given the data $\mathbf{x} = (x_1, x_2, \dots, x_n)$. To this end we use a so-called Gibbs sampler which will produce the stationary distribution by alternately drawing $(\boldsymbol{\pi}, \mathbf{T})$ given (\mathbf{J}, \mathbf{x}) and \mathbf{J} given $(\boldsymbol{\pi}, \mathbf{T}, \mathbf{x})$. After a certain number of initial iterations (burn-in) the Markov chain will settle into a stationary mode. Suppose that $g(\boldsymbol{\pi}, \mathbf{T})$ is a function of the phase-type parameters which is invariant under different representations of the same distribution, that is, if $(\boldsymbol{\pi}, \mathbf{T})$ and $(\boldsymbol{\pi}', \mathbf{T}')$ are two different representations of the same phase-type distribution, then $g(\boldsymbol{\pi}, \mathbf{T}) = g(\boldsymbol{\pi}', \mathbf{T}')$. This is for example the case for the density, quantiles or the ruin probabilities in the models above. If $(\boldsymbol{\pi}_i, \mathbf{T}_i)$, $i = 1, \dots, m$, denote m parameter sets obtained from the stationary distribution above, then $g(\boldsymbol{\pi}_1, \mathbf{T}_1), \dots, g(\boldsymbol{\pi}_m, \mathbf{T}_m)$ is again a stationary sequence of numbers. Ergodicity conditions are in general not available but will be merely assumed and empirically verified by the behavior of the sample means. Hence, assuming ergodicity, one may average the numbers in order to produce an estimator for the “true value” and one may calculate quantiles for the parameters from the empirical distribution of $g(\boldsymbol{\pi}_1, \mathbf{T}_1), \dots, g(\boldsymbol{\pi}_m, \mathbf{T}_m)$. If g is not invariant under different representations, it is not possible to average over the different g -values to obtain an estimator because the type of representation may switch through the iterations. As a consequence the present method cannot estimate the parameter values themselves. For parameter estimation we refer to the EM algorithm in the previous section.

We now outline the details for the alternate drawings of $(\boldsymbol{\pi}, \mathbf{T})$ given (\mathbf{J}, \mathbf{x}) and of \mathbf{J} given $(\boldsymbol{\pi}, \mathbf{T}, \mathbf{x})$. We need to impose a probability structure on $(\boldsymbol{\pi}, \mathbf{T})$, the so-called prior distribution:

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

where $\beta_i, \zeta_i, v_{i0}, v_{ij}$ for $i, j = 1, \dots, p$ are constants. Hence $\boldsymbol{\pi}$ has a Dirichlet distribution whereas t_i and t_{ij} are Gamma distributed. Furthermore they are all independent. This makes it easy to sample parameters from this prior.

The posterior distribution, which is the product of the prior and the likelihood, is hence given by

$$p(\mathbf{j} | \boldsymbol{\pi}, \mathbf{T}) = \prod_{i=1}^p \pi_i^{\beta_i + B_i - 1} \prod_{i=1}^p t_i^{N_{i0} + v_{i0} - 1} \exp(-t_i (Z_i + \zeta_i)) \\ \times \prod_{i=1}^p \prod_{\substack{j=1 \\ j \neq i}}^p t_{ij}^{N_{ij} + v_{ij} - 1} \exp(-t_{ij} (Z_i + \zeta_i)).$$

Thus $(\boldsymbol{\pi}, \boldsymbol{T})$ given $(\boldsymbol{J}, \boldsymbol{x})$ (here \boldsymbol{x} is of course unimportant because \boldsymbol{J} contains information about \boldsymbol{x} in particular) is simply drawn from the posterior distribution, which amounts to drawing from the prior with parameters $\beta'_i, \xi'_i, v'_{i0}, v'_{ij}$ for $i, j = 1, \dots, p$ where $\beta'_i = \beta_i + B_i, v'_{i0} = v_{i0} + N_{i0}, v'_{ij} = v_{ij} + N_{ij}$ and $\xi'_i = \xi_i + Z_i$.

Drawing \boldsymbol{J} given $(\boldsymbol{\pi}, \boldsymbol{T}, \boldsymbol{x})$ is much more involved. Given parameters $\boldsymbol{\pi}$ and \boldsymbol{T} and absorption times x_1, \dots, x_n we must produce realizations of Markov jump processes with the specified parameters which get absorbed exactly at times x_1, \dots, x_n . Since the probability of this event is zero it is a non-trivial task. It turns out, however, that we may employ a Metropolis-Hastings (MH) algorithm to simulate such Markov jump processes (see Bladt *et al.* (2003) for technical details). The MH algorithm amounts to the following simple procedure for simulating a Markov jump process j which gets absorbed exactly at time x :

0. Draw a Markov jump process j which is not absorbed by time x . This is done by simple rejection sampling: if a Markov jump process is absorbed before time x it is thrown away and a new Markov jump process is tried. We continue this way until we obtain the desired Markov jump process.
1. Draw a new Markov jump processes j' as in 0.
2. Draw a uniform random number U in $[0, 1]$.
3. If $U \leq \min(1, t_{j'} / t_{j_x-})$ then $j := j'$, otherwise keep j .
4. GO TO 1.

Here $x-$ denotes the limit from the left so j_{x-} is the state of the realization $\{j_s\}_{s \geq 0}$ just prior to exit. We iterate this procedure a number of times (burn-in) in order to get it into stationary mode. After this point and onwards, any j produced by the procedure may be considered as a draw from the desired conditional distribution and hence as a realization of a Markov jump process which gets absorbed exactly at time x .

The full procedure (Gibbs sampler) is then as follows.

0. Draw $(\boldsymbol{\pi}, \boldsymbol{T})$ from prior.
1. Draw $\boldsymbol{j} = (j_1, \dots, j_n)$ underlying Markov trajectories given $(\boldsymbol{\pi}, \boldsymbol{T})$ using the Metropolis-Hastings algorithm.
2. Draw $(\boldsymbol{\pi}, \boldsymbol{T})$ from posterior
3. Calculate ruin probabilities (or other representation-invariant functionals of interest) using the current parameters $(\boldsymbol{\pi}, \boldsymbol{T})$.
4. GO TO 1

After a number of initial iterations (burn-in), the procedure will stabilize into a stationary mode and from this point onwards we may produce a sample of e.g. ruin probabilities to be analyzed.

In order to display the quality of actual fit to the data it may be desirable also to produce a sequence of density values at each draw. Averaging these values through the iterations then produces an estimate for the density which may be compared to a histogram of the original data. For practical details on implementation, extensions, choice of prior or hyper-prior and a concrete statistical analysis of ruin probabilities, we refer to Bladt *et al.* (2003).

The EM and MCMC approaches should not be seen as competitors but rather as complementary methods. While the objective of the EM algorithm is to obtain a maximum likelihood estimator of the parameters, the MCMC approach focuses on inference for related functionals which may be useful when addressing issues concerning more complex models which partly depend on a phase-type assumption.

REFERENCES

- AALEN, O.O. (1995) Phase type distributions in survival analysis. *Scand. J. Statist.*, **22**, 447-463.
- ASMUSSEN, S. (2000) *Ruin Probabilities*, volume 2 of *Advanced Series on Statistical Science & Applied Probability*. World Scientific Publishing Co. Inc., River Edge, N.J.
- ASMUSSEN, S. (2003) *Applied Probability and Queues*. Springer-Verlag, New York.
- ASMUSSEN, S. and BLADT, M. (1996) Phase-type distribution and risk processes with state-dependent premiums. *Scand. Actuar. J.*, **1**, 19-36.
- ASMUSSEN, S. and BLADT, M. (1997) Renewal theory and queueing algorithms for matrix-exponential distributions. In S. Chakravarty and A. Alfa, editors, *Matrix-analytic methods in stochastic models*, volume 183 of *Lecture Notes in Pure and Appl. Math.*, pages 313-341. Dekker, New York.
- ASMUSSEN, S., NERMAN, O. and OLSSON, M. (1996) Fitting phase-type distributions via the EM-algorithm. *Scand. J. Statist.*, **23**, 419-441.
- ASMUSSEN, S., AVRAM, F. and USABEL, M. (2002) Erlangian approximations for finite-horizon ruin probabilities. *Astin Bulletin*, **32(2)**, 267-281.
- BLADT, M. and NEUTS, M.F. (2003) Matrix-exponential distributions: calculus and interpretations via flows. *Stochastic Models*, **19(1)**, 113-124.
- BLADT, M., GONZALEZ, A. and LAURITZEN, S.L. (2002) The estimation of phase-type related functionals through Markov chain Monte Carlo methods. *Scand. Actuar. J.*, **4**, 280-300.
- ERLANG, A. (1909) Sandsynlighedsregning og telefonsamtaler. *Nyt tidsskrift for Matematik*, **20**, 33-39.
- JENSEN, A. (1953) *A distribution model applicable to economics*. Munksgaard, Copenhagen.
- LATOUCHE, G. and RAMASWAMI, V. (1999) *Introduction to matrix analytic methods in stochastic modeling*. ASA-SIAM Series on Statistics and Applied Probability. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.
- MOLER, C. and VAN LOAN, C. (1978) Nineteen dubious ways to compute the exponential of a matrix. *SIAM Rev.*, **20**, 801-836.
- NEUTS, M.F. (1981) *Matrix-geometric solutions in stochastic models.*, volume 2 of *John Hopkins Series in the Mathematical Sciences*. Johns Hopkins University Press, Baltimore, Md.
- NEUTS, M.F. (1989) *Structured stochastic matrices of the M/G/1 type and their applications*, volume 5 of *Probability: Pure and Applied*. Marcel Dekker Inc., New York.
- NEUTS, M.F. (1995) *Algorithmic probability*. Stochastic Modeling Series. Chapman & Hall, London.
- OAKES, D. (1999) Direct calculation of the information matrix via the EM algorithm. *J. Royal Statist. Soc. B*, **61**, 479-482.
- O'CONNOR, C.A. (1990) Characterization of phase-type distributions. *Comm. Statist. Stochastic Models*, **6**, 1-57.
- PRESS, W., TEUKOLSKY, S., VETTERLING, W. and FLANNERY, B. (1992) *Numerical Recipes in Fortran*. Cambridge University Press.

MOGENS BLADT

IIMAS-UNAM

A.P. 20-726

01000 Mexico, D.F. – Mexico

E-mail: bladt@sigma.iimas.unam.mx