ABC methods for phase-type distributions with applications in insurance risk problems

Concepcion Ausin,
Department of Statistics,
Universidad Carlos III de Madrid

Joint work with:
Pedro Galeano, Universidad Carlos III de Madrid
Simon Wilson, Trinity College Dublin

23rd September, 2016

Introduction

- The main features of loss data sets are strong skewness, heavy-tails, heterogeneity and the presence of extremes.
- The class of phase-type distributions (PH) is a very large and flexible family of distributions defined on the positive real line (Neuts, 1981).
- PH distributions are dense on $[0, \infty)$ and therefore, any positive distribution can be theoretically well approximated by a PH distribution.
- Basically, a positive random variable is PH if each realization can be expressed as a sum of exponential random variables.
- Using PH distributions, it is possible to obtain ruin probabilities in insurance risk, system failure times in reliability and waiting times in queuing systems.

Introduction

- Classical estimation methods for PH distributions are traditionally based on the method of moments (Johnson and Taaffe, 1990) and maximum likelihood estimation (Asmussen, 1996).
- However, using these approaches, it is not easy how to derive confidence intervals for quantities of interest depending on the estimated PH distributions such as the ruin probability in a risk reserve process with PH claim sizes.
- This can be done in a natural way from the Bayesian perspective using MCMC methods (Bladt et al., 2003).
- Nevertheless, classical and Bayesian methods for PH distributions are very time consuming, mainly due to the difficulties in the evaluation of the likelihood.

Introduction

- Our proposal is to make use of the recently developed Approximate Bayesian Computation (ABC) methods (Marin et al., 2012) to make Bayesian inference on PH distribution and further, estimate quantities of interest such as ruin probabilities.
- ABC methods provide the advantage of avoiding the evaluation of the likelihood and are mainly based on simulation.
- Therefore, ABC methods seems to be very suitable for PH distributions whose likelihood is difficult and computationally expensive to evaluate but rather easy to simulate.
- We will apply our proposed ABC method to estimate credible intervals for the ruin probability of an insurance company.

Definition of PH distribution

A continuous PH(α, T) distribution of order m is defined as the distribution of the time until absorption in a finite Markov process on the states $\{1, ..., m+1\}$ with infinitesimal generator:

$$Q = \begin{bmatrix} T & \mathbf{T}^0 \\ \mathbf{0} & 0 \end{bmatrix},$$

where:

- T is a non-singular $m \times m$ matrix with $T_{ii} < 0$ and $T_{ij} \geq 0$.
- $\mathbf{T}^0 = -T\mathbf{1}$.
- α is a $m \times 1$ vector with the initial probabilities in each of the m transient states.
- The distribution function is given by,

$$F(x) = 1 - \alpha \exp\{Tx\}\mathbf{1}, \qquad \text{for } x \geq 0.$$

.

Examples of PH distributions

Exponential distribution, $Exp(\lambda)$: This is the simplest PH distribution with $m = 1$, $\alpha = 1$ and $T = -\lambda$.

Erlang distribution, $Er(k, \lambda)$: It is defined as the sum of k exponentials with the same rate, λ, and then, it is a PH distribution with $m = k$, $\alpha = (1, 0, \ldots, 0)_{(1\times k)}$ and

$$T = \begin{bmatrix} -\lambda & \lambda & & \\ & \ddots & \ddots & \\ & & -\lambda & \lambda \\ & & & -\lambda \end{bmatrix}_{(k\times k)}.$$

Examples of PH distributions

Mixture of Exponential distributions, $H_k(\alpha, \lambda)$, whose density is given by:

$$f(x \mid \alpha, \lambda) = \sum_{i=1}^{k} \alpha_i \lambda_i \exp(-\lambda_i x), \quad x > 0,$$

such that the variable is an $Exp(\lambda_i)$ with probability α_i.

This model is also a PH distribution with $m = k$, $\alpha = (\alpha_1, \ldots, \alpha_k)$ and

$$T = \begin{bmatrix} -\lambda_1 & & & \\ & -\lambda_2 & & \\ & & \ddots & \\ & & & -\lambda_k \end{bmatrix}_{(k \times k)}.$$

Examples of PH distributions

Coxian distributions, $Cox(k, p, \lambda)$: This is a generalization of the Erlang by having different rates and being able to reach the absorbing state from any phase with $m = k$, $\alpha = (1, 0, \ldots, 0)$ and

$$T = \begin{bmatrix} -\lambda_1 & \lambda_1 p_1 & & & \\ & -\lambda_2 & \lambda_2 p_2 & & \\ & & \ddots & \ddots & \\ & & & -\lambda_{k-1} & \lambda_{k-1} p_{k-1} \\ & & & & -\lambda_k \end{bmatrix}_{(k \times k)}.$$

Coxian distributions represent a wide subclass of PH distributions since any acyclic PH distribution can be expressed as a Coxian distribution.

Some more properties

Assume a random variable X following a PH distribution, $X \sim PH(\alpha, T)$, then:

- The density function is given by:

$$f(x) = \alpha \exp\{Tx\}\mathbf{T}^0, \qquad \text{for } x \geq 0.$$

- The r-th moment is given by:

$$E[X^r] = (-1)^r r! \alpha T^{-r} \mathbf{1}.$$

Lack of identifiability: Unfortunately, a PH representation (m, α, T) is not unique. For example, given an invertible matrix S such that $S\mathbf{1} = \mathbf{1}$, both representations (m, α, T) and $(m, \alpha S, S^{-1}TS)$ lead to the same PH distribution.

Interpretation of phases

Sometimes the number of phases, m, and the structure of the vector-matrix (α, T) is known. This is the case in many reliability problems as the following.

Repairmen problem: Assume we have a system with k independent machines that may eventually fail and be repaired. Machines lifetimes and repair times are independent exponential variables with rates λ and μ, respectively.

Therefore, the whole system failure time follow a PH distribution $m = k$, $\alpha = (1, 0, \ldots, 0)$ and T given by

$$\begin{bmatrix} -k\lambda & k\lambda & & & \\ \mu & -(k-1)\lambda - \mu & (k-1)\lambda & & \\ & 2\mu & -(k-2)\lambda - 2\mu & (k-2)\lambda & \\ & & \ddots & \ddots & \ddots \end{bmatrix}_{(k \times k)}.$$

Interpretation of phases

- Also, in survival analysis, there are many examples where PH distributions with known number of phases and known structures can be applied.
- One popular example is the compartmental kinetics model of pharmacokinetics which consists in describing the movements of a drug in the body.
- On the contrary, in finance and insurance risk, PH models are usually considered for describing loss distributions and in this case, the number of phases, m, and the structures of α and T are unknown.

Estimation of PH distributions

Assume that we have observed a sample of positive observations, $\{x_1, \ldots, x_n\}$, from a PH distribution with unknown (α, T).

- Traditional approximation methods (mostly in engineering) are based on the method of moments, (Johnson and Taaffe, 1990). These are fast and easy to implement but possible solutions are limited by the moment bounds.
- Maximum likelihood estimation methods can be developed through the Expectation-Maximization algorithm, (Asmussen, 1996), where the observed data is augmented such that for each transient state, $i = 1, \ldots, m$, we have:
  - B_i: The number of observations starting in state i, .
  - Z_i: The total time spent in state i.
  - N_{ij}: The total number of jumps from state i to state j, for $j \neq i$ and for $j = 1, \ldots, m+1$.

Estimation of PH distributions

Thus, the likelihood of the complete data set, $\mathbf{x}_c$, is simplified to:

$$l(\alpha, T \mid \mathbf{x}_c) = \prod_{i=1}^{m} \alpha_i^{B_i} \prod_{i=1}^{m} \exp(t_{ii} Z_i) \prod_{i=1}^{m} \prod_{\substack{j=1 \\ j \neq i}}^{m+1} t_{ij}^{N_{ij}}$$

$$= \prod_{i=1}^{m} \alpha_i^{B_i} \prod_{i=1}^{m} t_{i,m+1}^{N_{i,m+1}} \exp(-t_{i,m+1} Z_i) \prod_{\substack{j=1 \\ j \neq i}}^{m} t_{ij}^{N_{ij}} \exp(-t_{ij} Z_i)$$

where $t_{i,m+1}$ is the i-th element of the exit vector, T^0.

Given the complete data, the MLE of the PH parameters can be obtained:

$$\hat{\alpha}_i = \frac{B_i}{n}; \quad \hat{t}_{i,m+1} = \frac{N_{i,m+1}}{Z_i}; \quad \hat{t}_{ij} = \frac{N_{ij}}{Z_i}; \quad \hat{t}_{ii} = -\left(\hat{t}_{i,m+1} + \sum_{j=1; j \neq i}^{m} \hat{t}_{ij} \right)$$

Estimation of PH distributions

- The EM algorithm alternates these steps:
  - E-step: Computes the expectation of the missing quantities: B_i, Z_i, N_{ij}, for $i = 1, \ldots, m$, and for $j = 1, \ldots, m+1$ with $j \neq i$.
  - M-step: Given the expected missing quantities, the MLE of (α, T) are directly obtained.
- The E-step is computationally heavy since these expectations depend on exponential matrices.
- Alternatively, Bladt et al. (2001) develop a Bayesian algorithm to make inference on (α, T).
- They propose a Gibbs sampling method based on the same data augmentation strategy considered in the EM algorithm.

Estimation of PH distributions

Bladt et al. (2001) assume the following semi-conjugate priors:

$$
\begin{aligned}
(\alpha_1, \ldots, \alpha_m) &\sim \textit{Dirichlet}\,(\phi_1, \ldots, \phi_m) \\
t_{i,m+1} &\sim \textit{Gamma}\,(\nu_{i,m+1}, \zeta_i)\,, \ \text{for } i = 1, \ldots, m. \\
t_{ij} &\sim \textit{Gamma}\,(\nu_{ij}, \zeta_i)\,, \ \text{for } i = 1, \ldots, m;\ j \neq i.
\end{aligned}
$$

Estimation of PH distributions

Given a sample of observed data, $\{x_1, \ldots, x_n\}$, the following Gibbs sampling algorithm is developed:

Set some initial values for (α, T).

1. Generate a complete sample $\{y_1, \ldots, y_n\}$ where each y_i is a realization of a continuous time Markov process which get absorbed at time x_i.

2. Given the complete data, obtain the missing quantities: B_i, Z_i, N_{ij}, for $i = 1, \ldots, m$, and for $j = 1, \ldots, m+1$ with $j \neq i$.

3. Generate a sample from the conditional posterior:

$$(\alpha_1, \ldots, \alpha_m) \sim \textit{Dirichlet}\,(\phi_1 + B_1, \ldots, \phi_m + B_m)$$

$$t_{i,m+1} \sim \textit{Gamma}\,(\nu_{i,m+1} + N_{i,m+1}, \zeta_i + Z_i)\,, \text{ for } i = 1, \ldots, m$$

$$t_{ij} \sim \textit{Gamma}\,(\nu_{ij} + N_{i,j}, \zeta_i) + Z_i, \text{ for } i = 1, \ldots, m;\ j \neq i.$$

Estimation of PH distributions

- The main difficult part is step 1 since a Metropolis-Hastings method is used to simulate the missing sample with the exact observed absorbtion times.
- To do that, proposal underlying processes are simulated such that the absorption times are larger than those observed in the sample. This is done by rejection sampling.
- Frequently, this algorithm produces low acceptance rates which implies a bad mixing of the MCMC chain.
- Therefore, it seems reasonable to study the performance of ABC methods to make Bayesian inference for general PH models.

ABC methods

- ABC methods, (Marin et al., 2012), replace the calculation of the likelihood function with a simulation of the model that produces an artificial data set.
- The simulated data is then compared with the observed data using some kind of distance to approximate the posterior distribution of the model parameters.
- The idea is similar to approximation methods based on simulation which is standard in computer models.
- ABC methods are becoming popular in genetics, epidemiology and in population biology, where likelihood functions can usually not be calculated explicitly but stochastic simulation is straightforward.

ABC methods for discrete data

Assume that we have observed a data sample $\mathbf{x} = \{x_1, \ldots, x_n\}$ from a discrete variable, $X \mid \theta \sim f(x \mid \theta)$ and consider a prior, $f(\theta)$.

Suppose that the likelihood, $f(\mathbf{x} \mid \theta)$, is unknown (or difficult to evaluate) but it is easy to sample from $X \mid \theta$.

We may obtain a sample from the posterior distribution $f(\theta \mid \mathbf{x})$ with:

Repeat:

1. Simulate a value θ^* from the prior, $f(\theta)$.
2. Simulate an iid sample $\mathbf{x}^* = \{x_1^*, \ldots, x_n^*\}$ from $X \mid \theta^*$.

The pairs $(\theta^*, \mathbf{x}^*)$ are values from the joint distribution $f(\theta, \mathbf{X})$.

Now, we reject those sampled pairs such that $\mathbf{x}^* \neq \mathbf{x}$. Then, the values of θ^* that remain come from the posterior distribution $f(\theta \mid \mathbf{x})$.

ABC methods for discrete data

A problem with the previous approach is that if the sample size is large, it may take a very large number of iterations to generate artificial samples such that $\mathbf{x}^* = \mathbf{x}$

If there is a sufficient statistic, $s(\mathbf{x})$, for θ, we know that $f(\theta \mid \mathbf{x}) = f(\theta \mid s(\mathbf{x}))$.

Therefore, we can use the same idea as previously, but accepting samples $\mathbf{x}^*$ such that $s(\mathbf{x}^*) - s(\mathbf{x})$.

ABC methods for continuous data

When X is continuous, for any θ, the time taken to generate $\mathbf{x}^* = \mathbf{x}$ (or $s(\mathbf{x}^*) = s(\mathbf{x})$) will be infinite.

In this case, we may accept samples not too far away from $\mathbf{x}$.

Defining a distance measure, $\|\cdot\|$ and a tolerance, ϵ, we accept samples such that $\|\mathbf{x}^* - \mathbf{x}\| < \epsilon$ (or $\|s(\mathbf{x}^*) = s(\mathbf{x})\| < \epsilon$).

In practice, the value of ϵ is not fixed, but instead it is accepted just a certain proportion of the sampled values (e.g. 5%, 1%, 0.5%) with the smallest differences from the real data.

Trivial example

Exponential data: Suppose we have a sample $\mathbf{x} = \{x_1, \ldots, x_n\}$ from an exponential r.v., $X \sim Exp(\lambda)$.

Assume a conjugate prior $\lambda \sim Gamma(a, b)$ such that the true posterior is known: $\lambda \mid \mathbf{x} \sim Gamma(a + n, b + n\bar{x})$.

For this case, a sufficient statistic for λ is $s(\mathbf{x}) = \bar{x}$. Then, we may construct the following ABC algorithm by repeating:

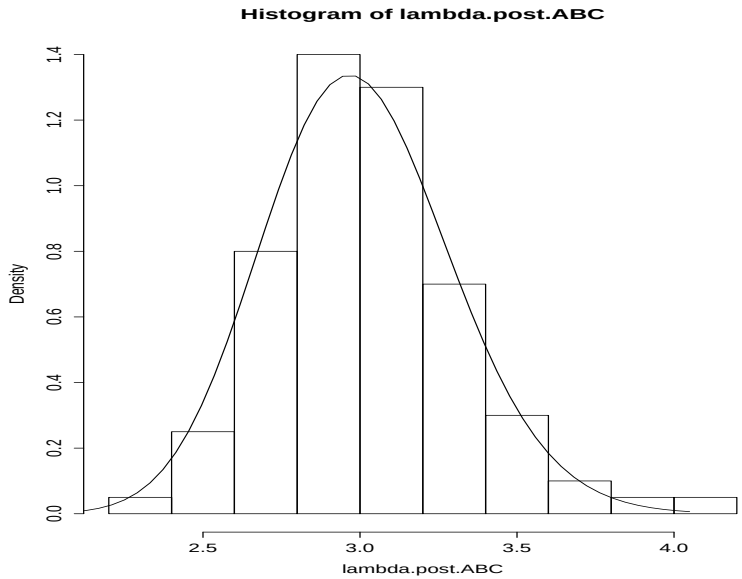
1. Simulate a value $\lambda^* \sim Gamma(a, b)$.
2. Simulate an iid sample $x_i^* \sim Exp(\lambda^*)$, for $i = 1, \ldots, n$.
3. Compute the mean of the simulated sample, $\bar{x}^*$.
4. Reject if $|\bar{x}^* - \bar{x}| > \epsilon$.

Trivial example

We generate $n = 100$ observations from an $Exp(\lambda = -3)$.

We implement the previous ABC method for 10000 iterations using a non informative prior, $\lambda \sim Gamma(0.01, 0.01)$.

The figure shows a histogram of the approximated posterior sample using the 1% of the simulated samples with smallest mean differences in abs. The solid line shows the true posterior density.

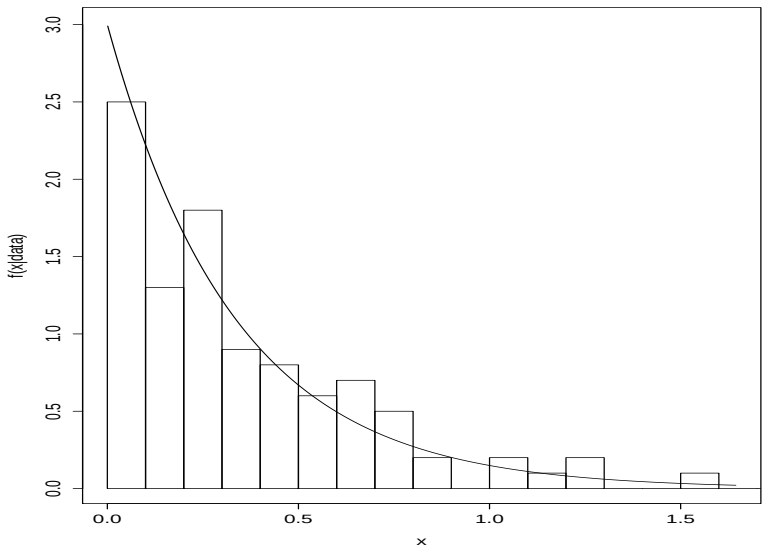


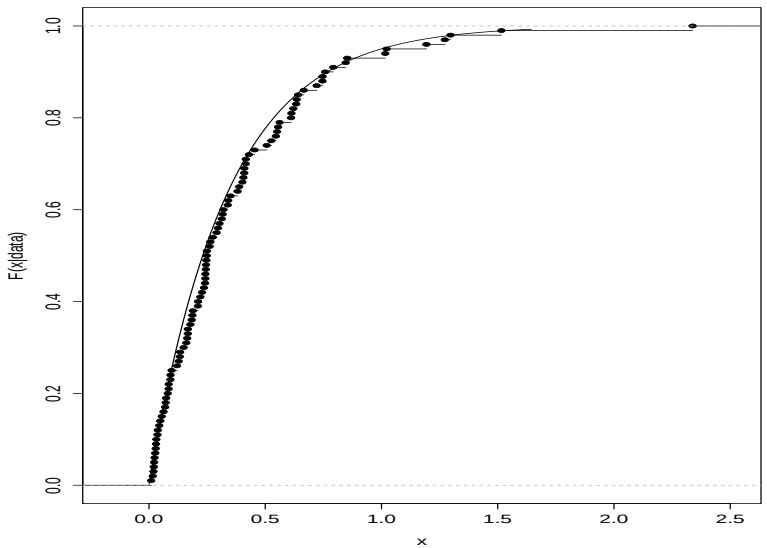
Trivial example

Clearly, we may obtain an approximated sample of the predictive density, $f(x_{n+1} \mid \mathbf{x})$, by just including in the ABC algorithm the step:

- Sample $x^*_{n+1} \sim Exp(\lambda^*)$

The figures show a histogram and the empirical cdf of the approximated predictive sample together with the true exponential pdf and cdf, resp.





ABC for PH distributions

Unfortunately, for any other PH distribution different from the single exponential, sufficient statistics are not available.

Then, we propose the use of the first m sample moments:

$$r_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k, \quad \text{for } k = 1, \ldots, m,$$

which we hope are close to be sufficient.

Theoretically, as $\epsilon \to \infty$, we are then approximating $f(\alpha, T \mid \mathbf{r})$, where $\mathbf{r} = (r_1, \ldots, r_m)$, instead of $f(\alpha, T \mid \mathbf{x})$.

ABC for PH distributions

Given a data sample $\mathbf{x}$ from a general PH(α, T) distribution, we propose the following ABC algorithm to approximate $f(\alpha, T \mid \mathbf{x})$.

Repeat:

1. Simulate a value (α^*, T^*) from the prior:

$$(\alpha_1^*, \ldots, \alpha_m^*) \sim \textit{Dirichlet}\,(\phi_1, \ldots, \phi_m)$$
$$t_{i,m+1}^* \sim \textit{Gamma}\,(\nu_{i,m+1}, \zeta_i)\,, \text{ for } i = 1, \ldots, m.$$
$$t_{ij}^* \sim \textit{Gamma}\,(\nu_{ij}, \zeta_i)\,, \text{ for } i = 1, \ldots, m;\ j \neq i.$$

   Reject and repeat if $|\det(T^*)| < \varepsilon$

2. Simulate an iid sample $x_i^* \sim PH(\alpha^*, T^*)$, for $i = 1, \ldots, n$.

3. Compute the sample moments of the simulated sample:

$$r_k^* = \frac{1}{n} \sum_{i=1}^{n} (x_i^*)^k\,, \quad \text{for } k = 1, \ldots, m,$$

ABC for PH distributions

Once we have the simulated samples, we propose to accept e.g. the 1% of them with the smallest Mahalabobis distance:

$$d_M(\log \mathbf{r}^*, \log \mathbf{r}) = \sqrt{(\log \mathbf{r}^* - \log \mathbf{r})^T S_r^{-1} (\log \mathbf{r}^* - \log \mathbf{r})}$$

where:

- $\log \mathbf{r}^*$ are the log moments of each simulated sample.
- $\log \mathbf{r}$ are the log moments of the observed sample.
- S_r is the sample covariance matrix of the moments, $\log \mathbf{r}^*$ using the whole set of simulations.

As usual, we may then obtain a predictive sample by sampling for each accepted pair (α^*, T^*):

- Sample $x_{n+1}^* \sim PH(\alpha^*, T^*)$

Example of a general PH distribution

We generate $n = 100$ observations from a PH distribution with $m = 4$ phases, $\alpha = (1, 0, 0, 0)$ and

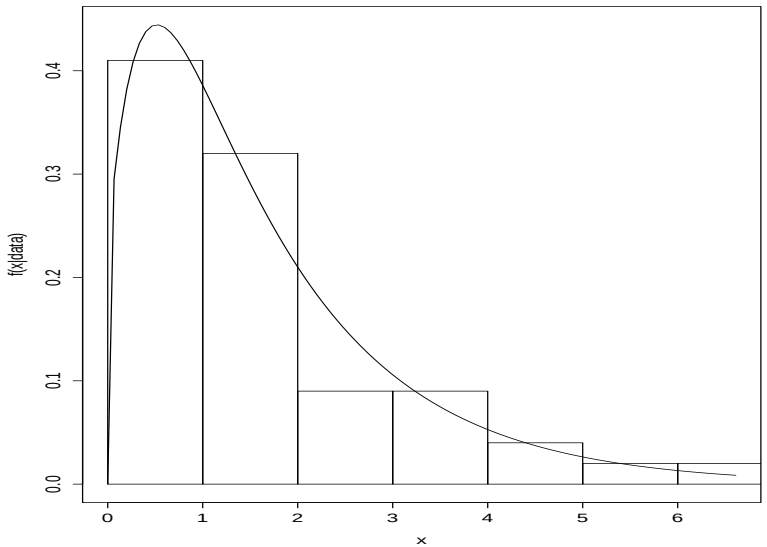
$$T = \begin{bmatrix} -1.29 & 0.1 & 0.67 & 0.29 \\ 0.88 & -5.84 & 2.71 & 2.11 \\ 0.42 & 0.54 & -2.86 & 0.05 \\ 2.97 & 0.12 & 1.09 & -4.73 \end{bmatrix}$$

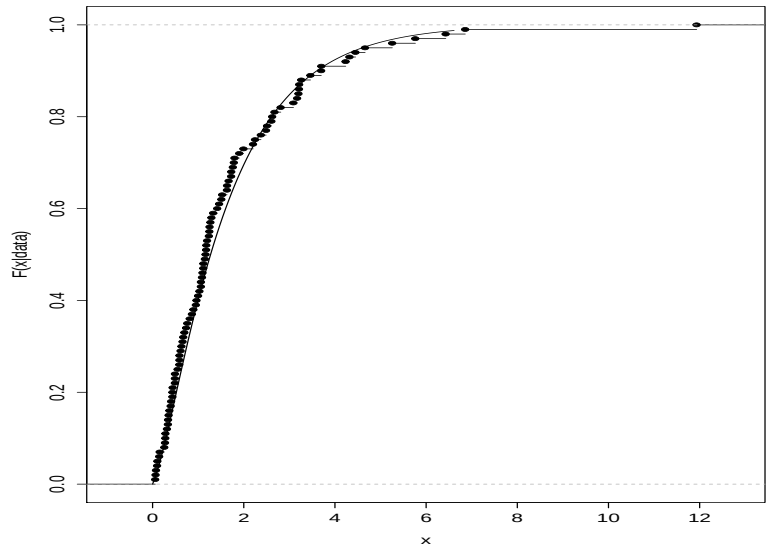
We implement the proposed ABC method for 10000 iterations using a non informative prior:

$$\begin{aligned} (\alpha_1^*, \ldots, \alpha_m^*) &\sim \textit{Dirichlet}\,(1, \ldots, 1) \\ t_{i,m+1}^* &\sim \textit{Gamma}\,(0.01, 0.01)\,, \text{ for } i = 1, \ldots, m. \\ t_{ij}^* &\sim \textit{Gamma}\,(0.01, 0.01)\,, \text{ for } i = 1, \ldots, m;\ j \neq i. \end{aligned}$$

Example of a general PH distribution

The figures show a histogram and the empirical cdf of the approximated predictive sample together with the true exponential pdf and cdf, resp.





Ruin probabilities

Assume a risk reserve process, R_t, defined by:

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

where:

- u is the initial capital of the insurance company.
- c is the premium income per unit time.
- $N(t)$ is number of claims arrived up to time t.
- $X_1, X_2, \ldots$ are iid claim sizes independent from $N(t)$.

Further, we assume that:

- $N(t)$ follows a Poisson process of rate λ.
- Claim sizes are iid following a PH(α, T) distribution

Ruin probabilities

Given the equilibrium condition,

$$-\lambda\alpha T^{-1}\mathbf{1} < c$$

the limit probability of ruin of the company is given by:

$$\Pr\left(\inf_{t>0} R_t < 0 \mid R_0 = u\right) = \alpha_+ \exp\left\{(T + \mathbf{T}^0\alpha_+)u\right\}\mathbf{1}$$

where $\alpha_+ = -\lambda\alpha T^{-1}$.

Otherwise, the limit ruin probability is one.

Note that the problem of estimating the ruin probability for this model is equivalent to estimating the stationary waiting time in a M/PH/1 queueing system with Poisson arrival rate λ and PH(α, T) distributed service times.

Ruin probabilities

Assume now that we have a sample, $\mathbf{t} = \{t_1, \ldots, t_n\}$, of interarrival claim times from a Poisson process of rate λ.

And we also have the corresponding sample of claim sizes, $\mathbf{x} = \{x_1, \ldots, x_n\}$, that we assume to be generated from a PH(α, T).

We may assume the standard conjugate prior, $\lambda \sim Gamma(a, b)$ such that the posterior is,

$$\lambda \mid \mathbf{t} \sim Gamma(a + n, b + \sum_{i=1}^{n} t_i).$$

Therefore, given a sample from $\lambda \mid \mathbf{t}$ and a sample from $(\alpha, T) \mid \mathbf{x}$ obtained from the ABC algorithm, we can obtain a sample of ruin probabilities that can be used to approximate credible intervals for the ruin probabilities.

Example

Assume we have a sample of $n = 100$ claim sizes simulated in the previous example from a PH distribution with $m = 4$ phases, $\alpha = (1, 0, 0, 0)$ and

$$T = \begin{bmatrix} -1.29 & 0.1 & 0.67 & 0.29 \\ 0.88 & -5.84 & 2.71 & 2.11 \\ 0.42 & 0.54 & -2.86 & 0.05 \\ 2.97 & 0.12 & 1.09 & -4.73 \end{bmatrix}$$

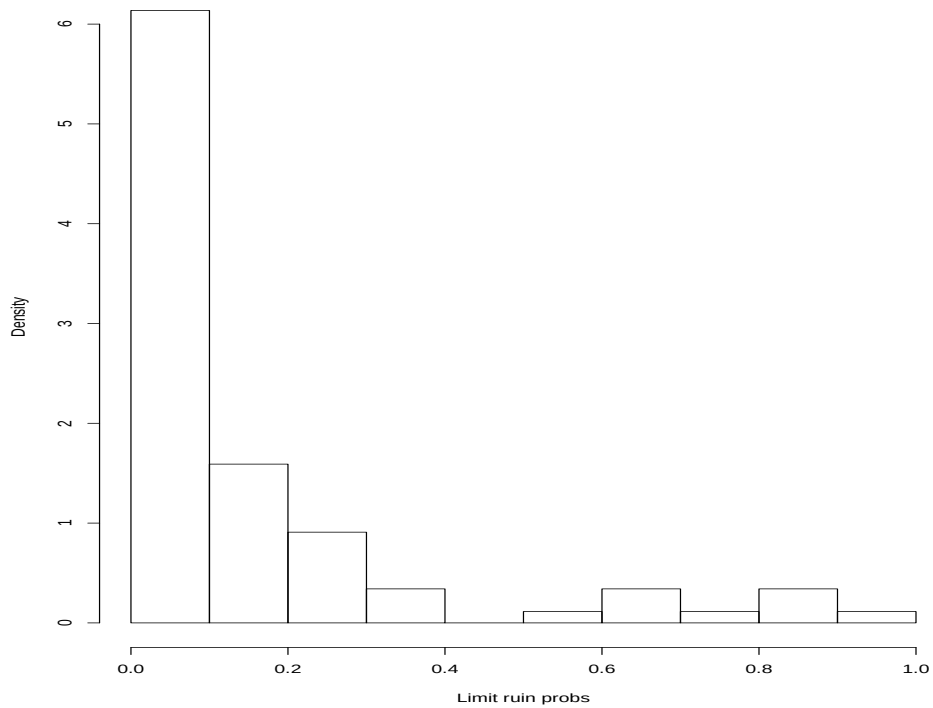
Also, we simulate $n = 100$ interarrival times from an exponential distribution with, $\lambda = 0.5$.

Given the simulated data, a premium income, $c = 1$ and initial capital $u = 10$, we want to estimate the posterior distribution for the limit ruin probability.

Example for ruin probabilities

For the true parameters, the equilibrium condition holds: $-\lambda\alpha T^{-1}\mathbf{1} = 0.8407 < 1$ and the true ruin probability is 0.2884.

The figure shows the posterior sample from ruin probabilities:



A 95% credible interval is $[0, 0.7213]$.

Extensions

- Up to now, we have assumed that the number of phases, m, is known. However, we could impose a prior distribution on m and use ABC to sample from the joint posterior of (m, α, T).
- We have observed that problems may appear when there are many zeros in the intensity matrix, T. One possibility could be to impose positive prior mass on zero values for the off diagonal elements of T.
- Another alternative could be to consider canonical representations of PH distributions.
  - As mentioned, the Coxian family represents the whole set of acyclic PH distributions.
  - There is also a canonical representation for the whole set of PH distributions called monocyclic representation. However, the order of this PH representation is often larger than that of the original representation.

Extensions

- The efficiency of the proposed approach could be also improved by considering the ABC-MCMC approach.

For $s = 1$ to S:

1. Simulate a value $\left(\alpha^{(s)}, T^{(s)}\right) \sim q\left(\alpha, T \mid \alpha^{(s-1)}, T^{(s-1)}\right)$. Reject and repeat if $|\det(T^{(s)})| < \varepsilon$
2. Simulate an iid sample $x_i^{(s)} \sim PH(\alpha^{(s)}, T^{(s)})$, for $i = 1, \ldots, n$.
3. Compute the sample moments, $r_k^{(s)}$, of $\mathbf{x}^{(s)}$, for $k = 1, \ldots, m$.
4. If $d_M(\log \mathbf{r}^{(s)}, \log \mathbf{r}) < \epsilon$, accept $\left(\alpha^{(s)}, T^{(s)}\right)$ with probability:

$$\min\left\{1, \frac{\pi(\alpha^{(s)}, T^{(s)}) q\left(\alpha^{(s-1)}, T^{(s-1)} \mid \alpha^{(s)}, T^{(s)}\right)}{\pi(\alpha^{(s-1)}, T^{(s-1)}) q\left(\alpha^{(s)}, T^{(s)} \mid \alpha^{(s-1)}, T^{(s-1)}\right)}\right\}$$

References

- Asmussen, S., Nerman, O., Olsson, M. (1996). Fitting phase-type distributions via the EM algorithm. *Scandinavian Journal of Statistics*, 23, 419–441.
- Bladt, M., Gonzalez, A., Lauritzen, S.L. (2003). The estimation of phase-type related functionals using Markov Chain Monte Carlo methods. *Scandinavian Actuarial Journal*, 2003, 280–300.
- Johnson, M.A., Taaffe, M.R. (1991). An investigation of phase-distribution moment-matching algorithms for use in queueing models. *Queueing Systems*, 8, 129–148.
- Marin, J.M., Pudlo, P., Robert, C.P., Ryder, R.J. (2012) Approximate Bayesian computational methods. *Statistics and Computing*, 22, 1167-1180.
- Neuts, M.F. (1981). Matrix Geometric Solutions in Stochastic Models. Johns Hopkins University Press, Baltimore.