Rare-event Simulation with Markov Chain Monte Carlo

YUYA SUZUKI

Master of Science Thesis
Stockholm, Sweden 2013
Rare-event Simulation with Markov Chain Monte Carlo

YUYA SUZUKI

Degree Project in Mathematical Statistics (30 ECTS credits)
Degree Programme in Engineering Physics (300 credits)
Royal Institute of Technology year 2013
Supervisor at KTH was Henrik Hult
Examiner was Henrik Hult

TRITA-MAT-E 2013:59
ISRN-KTH/MAT/E--13/59--SE

Royal Institute of Technology
School of Engineering Sciences

KTH SCI
SE-100 44 Stockholm, Sweden

URL: www.kth.se/sci
Abstract

In this thesis, we consider random sums with heavy-tailed increments. By the term random sum, we mean a sum of random variables where the number of summands is also random. Our interest is to analyse the tail behaviour of random sums and to construct an efficient method to calculate quantiles. For the sake of efficiency, we simulate rare-events (tail-events) using a Markov chain Monte Carlo (MCMC) method. The asymptotic behaviour of sum and the maximum of heavy-tailed random sums is identical. Therefore we compare random sum and maximum value for various distributions, to investigate from which point one can use the asymptotic approximation. Furthermore, we propose a new method to estimate quantiles and the estimator is shown to be efficient.
## Contents

1 **Introduction**  
2 **Background**  
   2.1 Markov chain Monte Carlo  
      2.1.1 Markov chain  
      2.1.2 Gibbs sampler  
      2.1.3 Metropolis Hastings  
      2.1.4 Convergence of MCMC  
   2.2 Heavy-tailed distributions  
      2.2.1 Regularly varying distributions  
      2.2.2 Subexponential distributions  
   2.3 Applications of the problem  
      2.3.1 Application for insurance risk  
      2.3.2 Application for queueing theory  
3 **The estimator construction and sampler algorithm**  
   3.1 Construction of the estimator for small probabilities  
   3.2 Description of the MCMC algorithm  
4 **Assessing convergence**  
   4.1 Gelman&Rubin’s method  
   4.2 Geweke’s method  
5 **Numerical experiments**  
6 **Quantile estimation**  
   6.1 Procedure of estimation  
   6.2 Numerical experiments  
   6.3 Example with real data: Danish Fire Insurance 1980-1990  
7 **Conclusion**
Chapter 1

Introduction

SIMULATION methods with intent to estimate small probabilities are of practical importance in various areas such as telecommunication systems, risk management for insurance and operational risk, etc.

Consider the problem to estimate the probability \( p = \mathbb{P}(S_n > a_n) \) where \( a_n \) is some large constant, \( S_n = X_1 + X_2 + \cdots + X_n \) and \( X_i \)'s are non-negative, independent and identically distributed (iid). Here we assume that \( p \to 0 \) as \( a_n \to \infty \) so that we consider events are rare. When \( n = N \) is not constant but has Poisson distribution, then the setting becomes equivalent to the waiting time of an M/G/1 queue system. In the context of queueing theory, especially for the application with the case \( X_i \) having Pareto distribution, has been studied substantially, see Sees Jr and Shortle [2002], Gross et al. [2002]. On the other hand, \( S_N \) is sometimes mentioned as compound sum, this term is mainly used in the context of insurance risk. In the Cramér-Lundberg model, \( X_i \)'s are interpreted as amount of claim and \( N \) is number of claims, see McNeil et al. [2010]. This type of statistical model, assuming a heavy-tail, is also applied to the quantification of operational risk in finance, Embrechts et al. [2003].

The simplest solution for this problem is to use the standard Monte Carlo method that makes \( T \) iid copies \( \{S_n^t\} \) of \( S_n \) and estimates \( p \) by

\[
\hat{p} = \frac{1}{T} \sum_{t=1}^{T} I\{S_n^t > a_n\}.
\]

However, this method is inefficient to compute small probabilities with high threshold \( a_n \). Let us see the relative error of the estimate:

\[
\frac{\text{Std}(\hat{p})}{\hat{p}} = \sqrt{\frac{1 - \hat{p}}{T\hat{p}}} \to \infty,
\]

as \( a_n \to \infty \). This shows the difficulty of computing using standard Monte Carlo.

There are alternative methods mainly based on importance sampling. In
importance sampling, \( \{\tilde{S}_n^t\} \) instead of \( \{S_n^t\} \), is sampled from a new probability measure \( \tilde{P} \), see Rubino et al. [2009], Blanchet and Liu [2008].

However, we use the method based on Makov chain Monte Carlo proposed by Gudmundsson and Hult [2012]. The reason is two-fold: one is the ease of use, importance sampling methods require to compute an appropriate change of measure. If we wish to simulate with another type of distribution, we need to compute analytically and implement it. The other reason is that Gudmundsson and Hult [2012] showed the MCMC based method is compatible or rather better than existing importance sampling methods. The MCMC estimator in Gudmundsson and Hult [2012], is constructed to be unbiased for the reciprocal probability, and having vanishing normalized variance. They showed the efficiency of the proposed method with the setting that \( X_i \)'s have Pareto distribution and the number of sum \( n = N \) having Geometric distribution. In this paper, we show numerical results with other parameters and distributions such as Weibull, Pareto, and log-normal for \( X_i \)'s and Poisson and Geometric distribution for \( N \). With that results, it is shown that the efficiency of the MCMC based method also holds for those parameters and distributions.

In addition, to estimate a quantile of \( S_n \) is also an essential problem. For instance, quantile is especially called “Value at Risk” in finance, and this risk measure is used in the regulatory framework of Basel. In the context of queueing theory there exist some papers on estimating quantiles, for example Fischer et al. [2001] assuming Pareto service times in simulating Internet queues. We propose a new method to estimate high quantiles and it is shown that our proposal method is remarkably accurate.

The rest of this paper is organized as follows. In Chapter 2, we provide a background theory of this paper. In Chapter 3, the method to estimate small probability proposed by Gudmundsson and Hult [2012] is described. In Chapter 3, we assess the convergence of the MCMC. In Chapter 4, numerical results with various distributions are shown. In Chapter 5, we propose the new method to estimate high quantiles and the numerical results are shown.
Chapter 2

Background

In this section, general views of background are taken, such as MCMC, heavy-tailed distributions, and applications for our problem formulation.

2.1 Markov chain Monte Carlo

MCMC is a computer-intensive sampling method based on Bayesian statistics. It is originally proposed by Metropolis et al. [1953] and generalized by Hastings [1970]. MCMC enables to simulate from a distribution by embedding it as a limiting distribution of Markov chain and simulating from the chain until it approaches equilibrium. There exist large classes of MCMC algorithm. Especially, Gibbs sampler and Metropolis Hastings algorithm are the most applied MCMC algorithms.

In this section, we give details of Markov chain, and the algorithm of Gibbs sampler and Metropolis Hastings.

2.1.1 Markov chain

Markov chain $\{\theta^{(j)}\}_{j=1}^{\infty}$ on a state space $S$ is a stochastic process, which satisfies the Markov property,

$$P(\theta^{(j)} = x|\theta^{1} = x_1, ..., \theta^{(j-1)} = x_{j-1}) = P(\theta^{(j)} = x|\theta^{(j-1)} = x_{j-1}).$$

(2.1)

This property means that the state $\theta^{(j)}$ depends on only the previous state $\theta^{(j-1)}$. When it satisfies the following, then the Markov chain is said to be (time-) homogeneous:

$$P(\theta^{(j)} = x|\theta^{(j-1)} = y) = P(\theta^{(j+1)} = x|\theta^{(j)} = y) \forall j.$$

(2.2)

In this case, the transition kernel $P(x, A)$ is defined as:

1. $\forall x \in S, P(x, \cdot)$ is a probability function over $S$. 

2. \( \forall A \subseteq S \), the function \( x \mapsto P(x, A) \) is measurable.

Transition probabilities from state \( x \) to state \( y \) over \( m \) steps, denoted by \( P^m(x, y) \), is obtained as

\[
P^m(x, y) = \sum_{x_1} \cdots \sum_{x_{m-1}} P(\theta^{(m)} = y, \theta^{(m-1)} = x_{m-1}, \ldots, \theta^{(1)} = x_1 | \theta^{(0)} = x)
\]

From the above, the Chapman-Kolmogorov equation can be derived:

\[
P^{n+m}(x, y) = \sum_{z} P(\theta^{(n+m)} = y, \theta^{(n)} = z | \theta^{(0)} = x) P(\theta^{(n)} = z | \theta^{(0)} = x) = \sum_{z} P^n(x, z) P^m(z, y).
\]

Note that all summations are with respect to the elements of state space \( S \).

The distribution \( \pi \) is called a stationary distribution if

\[
\sum_{x \in S} \pi(x) P(x, y) = \pi(y), \forall y \in S.
\] (2.3)

The main interest of MCMC is to sample from the stationary distribution \( \pi(y) \) of a Markov chain, therefore \( \pi(y) \) is sometimes called a target distribution. If there exists a distribution satisfying the following,

\[
\lim_{n \to \infty} P^n(x, y) = \pi'(y),
\] (2.4)

then \( \pi'(y) = \pi(y) \). Therefore limiting results are important matters for MCMC, and we shall mention the property of ergodicity. To define ergodicity, three different properties are required:

- The state \( y \) is said to be positive recurrent if the Markov chain, starting \( \theta^1 = y \) and returning to \( \theta^k = y \) with probability 1, and the returning time, say \( T_y \), has finite mean, \( \mathbb{E}(T_y) < \infty \). The chain is said to be positive recurrent if every state \( y \in S \) is positive recurrent.

- The state pair \( x, y \) is said to be irreducible if the Markov chain, starting \( \theta^1 = x \) and returning to \( \theta^k = y \) with probability 1, and the returning time, say \( T_{x \to y} \), has finite mean, \( \mathbb{E}(T_{x \to y}) < \infty \). The chain is said to be irreducible if every state pair \( x, y \in S \) is irreducible.

- The state \( y \) is said to be aperiodic if

\[
g.c.d\{n|P^n(x, x) > 0\} = 1,
\] (2.5)
where $g.c.d$ is the largest common divisor. The chain is said to be aperiodic if every state $y \in S$ is aperiodic. When the chain is aperiodic, then there exist the limit, $\lim_{n \to \infty} P^n(x, y) = \pi(y)$ for all $x, y \in S$. Furthermore, [Nummelin 1984] has shown that irreducibility and ergodicity are equivalent to the condition,

$$\lim_{n \to \infty} \|P^n(x, \cdot) - \pi(\cdot)\| = 0 \quad \forall x \in S.$$  \hspace{1cm} (2.6)

Then finally, the state $y$ is said to be ergodic if $y$ is positive recurrent and aperiodic. The chain is said to be ergodic if every state $y \in S$ is ergodic. And also the chain is said to be geometrically ergodic if there exists a constant $0 \leq \lambda < 1$, and a real integrable function $M(x)$ such that

$$\|P^n(x, \cdot) - \pi(\cdot)\| \leq M(x)\lambda^n. \hspace{1cm} (2.7)$$

If the function $M$ does not depend on $x$, then the chain is said to be uniformly ergodic. Ergodicity is an important property for MCMC. The following theorem makes the base of main motivation to use MCMC.

If the consisted chain has ergodicity and a real-valued function $t(\theta)$ has a finite expectation, then

$$\frac{1}{n} \sum_{i=1}^{n} t(\theta^{(i)}) \to \mathbb{E}_\pi [t(\theta)] \quad \text{as} \quad n \to \infty, \quad \text{with probability 1.} \hspace{1cm} (2.8)$$

### 2.1.2 Gibbs sampler

Gibbs sampling is an MCMC algorithm where the transition kernel is fully formed by conditional distributions. Assume that the target distribution is $\pi(\theta)$ where $\theta = (\theta_1, ..., \theta_d)$. Assume also that the conditional distributions $\pi_i(\theta_i) = \pi(\theta_i | \theta_1, ..., \theta_{i-1}, \theta_{i+1}, ..., \theta_d)$ for $i = 1, ..., d$ are available. In our case, $\theta = (N, X_1, ..., X_N)$ and the target distribution is $\pi(\theta) = \mathbb{P}(\theta | S_N > a_n)$. Then the algorithm is as follows,

1. Initialize the iteration counter of the chain $j = 1$ and set initial values $\theta^0 = (\theta_1^0, ..., \theta_d^0)$. 

2. Sample a new value $\theta^{(j)} = (\theta_1^{(j)}, ..., \theta_d^{(j)})$ from conditioning that given $\theta^{(j-1)}$ through successive generation of values.

   $$\theta_1^{(j)} \sim \pi(\theta_1 | \theta_2^{(j-1)}, ..., \theta_d^{(j-1)})$$

   $$\theta_2^{(j)} \sim \pi(\theta_1^{(j)}, \theta_3^{(j-1)}, ..., \theta_d^{(j-1)})$$

   $$\vdots$$

   $$\theta_d^{(j)} \sim \pi(\theta_1^{(j)}, ..., \theta_{d-1}^{(j)}).$$
3. Change the counter of the chain \( j \) to \( j + 1 \) and return to step 2 until convergence is reached.

When convergence is reached, the value \( \theta^{(j)} \) has the target distribution \( \pi \).

### 2.1.3 Metropolis Hastings

Metropolis Hastings algorithm does not assume that full conditional distribution is available, and the construction of algorithm is very flexible. The transition kernel has two steps, first sample from a proposal distribution \( q(\theta, \phi) \) given previous position \( \theta \), and then determine if accept the proposal. The algorithm can be described as follows:

1. Initialize the iteration counter of the chain \( j = 1 \) and set initial values \( \theta^0 = (\theta_1^0, ..., \theta_d^0) \).

2. Sample a new value \( \phi \) from the density \( q(\theta^{(j-1)}, \cdot) \).

3. Evaluate the acceptance probability of the candidate \( \alpha(\theta^{(j-1)}, \phi) \) defined by,

\[
\alpha(\theta, \phi) = \min \left\{ 1, \frac{\pi(\phi)q(\phi, \theta)}{\pi(\theta)q(\theta, \phi)} \right\}.
\]

4. Draw an independent uniform random variable \( u \sim U(0, 1) \). If the candidate is accepted,

\[
u < \alpha(\theta^{(j-1)}, \phi),
\]

then \( \theta^{(j)} = \phi \), else \( \theta^{(j)} = \theta^{(j-1)} \).

5. Change the counter of the chain \( j \) to \( j + 1 \) and return to step 2 until convergence is reached.

However, to choose an appropriate \( q \) is sometimes difficult. Construction of \( q \) affects on the convergence of MCMC. If \( q \) proposes the new value that has only a little difference from previous value, then the convergence becomes slow, but also if the proposal value has extremely big difference, then the acceptance probability becomes nearly zero. Therefore the choice of appropriate \( q \) is required.

### 2.1.4 Convergence of MCMC

For MCMC algorithms, there exists a serious problem: how long should be run the MCMC to have it convergent? As we mention the details in a later section, here the general view is shown.

Tools to determine the MCMC convergence is sometimes called MCMC diagnostics. There exist two types of diagnostics, one is to predetermined theoretically.
from the algorithm input but these methods are far from application, because they give only bounds, suggesting numbers of iterations that are several orders of magnitude beyond what could be considered as reasonable or feasible in practice. Moreover they include often complicated mathematics and have difficulty to compute even bounds for each MCMC setting.

The other one is quite useful in practice: to determine the convergence from MCMC output. [Cowles and Carlin 1996] is a review paper on these diagnostics, providing a review of 13 diagnostics algorithm. We will show two convergence diagnostics and apply them to our MCMC algorithm.

2.2 Heavy-tailed distributions

In statistical modelling, Gaussian models are widely used and they are easy to work with because a lot of analytical expressions are possible. For instance, Black-Scholes model by [Black and Scholes 1973] in finance, and Gaussian linear regression models. However, there are many situations where Gaussian distributions are inappropriate. When one focusses on the tail of distribution, Gaussian may underestimate the probability of extreme events, for example [McNeil and Frey 2000] shows that in the financial risk management domain. Therefore using the appropriate distribution is required for a corresponding purpose.

The usage of “heavy-tailed” depends on the area of interest, however some classes of heavy-tailed distribution are shown in this section.

2.2.1 Regularly varying distributions

A distribution $F$ is said to have regularly varying tail with index $\alpha$ when it satisfies

$$F(cx) \sim c^{-\alpha}F(x) \text{ as } x \to \infty,$$

(2.9)

for any constant $c > 0$. This class also includes some class of distributions such as Pareto, Burr, stable and log-gamma distributions. For these distribution with tail index $\alpha$, there exists finite moments for all $\gamma < \alpha$:

$$\int_{\Omega} x^\gamma dF(x) < \infty,$$

(2.10)

Regular variation property is often used in extreme value theory. There the maximum $M_n = \max(X_1, X_2, ..., X_n)$ is considered. One of the main interest is to know the distribution of $H = \lim_{n \to \infty} (M_n - d_n)/c_n$ with appropriate constants $c_n, d_n$. If there exists such $c_n, d_n$ and the distribution of $H$, then the distribution of $X$ is said to belong to the Maximum Domain of Attraction of extreme value distribution $F_H$. It is proved that there exist just the following three types of distributions for $F_H$, 

---

**Rare-event Simulation with MCMC**
• Fréchet:
\[ \Phi_\alpha(x) = \begin{cases} \exp\{-x^{-\alpha}\} & (\text{for } x > 0) \\ 0 & (\text{for } x \leq 0). \end{cases} \] (2.11)

• Weibull:
\[ \Psi_\alpha(x) = \begin{cases} 1 & (\text{for } x \geq 0) \\ \exp\{-(-x)^\alpha\} & (\text{for } x < 0). \end{cases} \] (2.12)

• Gumbel:
\[ \Lambda(x) = \exp\{-e^{-x}\} \ (\text{for } x \in \mathbb{R}). \] (2.13)

### 2.2.2 Subexponential distributions

A distribution \( F \) on \( \mathbb{R}^+ \) with a tail distribution function \( \bar{F}(x) = \mathbb{P}(x < X < \infty) \), is called subexponential when it satisfies
\[ F \ast F(x) \sim 2F(x) \quad \text{as} \quad x \to \infty. \] (2.14)

where \( \ast \) stands for the convolution. This class includes many kinds of distribution, i.e., Pareto, Cauchy, lognormal, Burr, and Weibull (with the shape parameter \( \beta < 1 \)).

Subexponential distributions are widely applied to express real phenomena. For instance, as we mentioned queueing theory, non-life insurance and also earthquake magnitudes, see Huijlet and Raynaud [1999]. Subexponentiality is sometimes defined by an equivalent proposition:
\[ \lim_{x \to \infty} \frac{\mathbb{P}(\max(X_1, X_2, \ldots, X_n) > x)}{\mathbb{P}(X_1 + X_2 + \cdots + X_n > x)} = 1, \quad n \geq 2. \] (2.15)

In this paper, we use this property to construct an estimator for small probabilities and also high quantiles.

### 2.3 Applications of the problem

In this section, applications of the following stochastic system are shown:
\[ S_N = \sum_{i=1}^{N} X_i, \] (2.16)

where \( X_i \)'s are iid random variables and \( N \) is an independent random variable of \( X_i \)'s.
2.3.1 Application for insurance risk

In the beginning of 20th century, a Swedish actuary, Filip Lundburg, founded the base of the stochastic insurance theory in [Lundberg 1903]. The basic insurance model is called the Cramér-Lundberg model. In the model, claim size $X_i$ and claim number $N$ are modeled as random variables. Originally, the model is:

$$S(t) = \sum_{i=1}^{N(t)} X_i,$$  \hspace{1cm} (2.17)

where $t$ represents time. The expectation is calculated as:

$$E[S(t)] = E\left[\sum_{i=1}^{N(t)} X_i \right] = E\left[ E\left[ \sum_{i=1}^{N(t)} X_i \mid N(t) \right] \right] = E[N(t)E[X_i]] = E[N(t)]E[X_i].$$  \hspace{1cm} (2.18)

Assume that the variance of $N(t)$ and $X_i$ are finite, then it holds:

$$\text{var}\left[ \sum_{i=1}^{N(t)} (X_i - E[X_i]) \mid N(t) \right] = \sum_{i=1}^{N(t)} \text{var}[X_i | N(t)] = N(t)\text{var}[X_i].$$  \hspace{1cm} (2.19)

Therefore

$$\text{var}[S(t)] = E[N(t)\text{var}[X_i] + \text{var}[E[X_i] | N(t)]]$$
$$= E[N(t)]\text{var}[X_i] + \text{var}[N(t)]E[X_i^2].$$  \hspace{1cm} (2.20)

We consider the time-fixed version of this system. Let us write $S(t) = S, N(t) = N$. We are interested in the distribution of $S$. The characteristic function, defined by $\phi_S(s) = E[\exp(isS)], s \in \mathbb{R}$, determines the distribution of $S$, meaning that the characteristic function and the distribution of any random variable have one-to-one correspondence. This fact is derived from the Lévy’s continuity theorem.

Suppose that the sequence of random variables $\{Y_i\}_{i=1}^{\infty}$ and the sequence of corresponding characteristic functions $\{\phi_i\}_{i=1}^{\infty}$. If the sequence of characteristic functions converges pointwise to some characteristic function $\phi$

$$\lim_{i \to \infty} \phi_i(s) = \phi(s), \ \forall s \in \mathbb{R},$$  \hspace{1cm} (2.21)

then

$$Y_i \xrightarrow{D} Y, \text{ as } i \to \infty,$$  \hspace{1cm} (2.22)

and $\phi(s)$ is the characteristic function of $Y$.

The characteristic function of $S$ is expressed as:

$$\phi_S(s) = E[E[\exp(is(X_1 + ... + X_N))] | N(t)]$$
$$= E[E[\exp(is(X_1))]^N] = E[\phi_{X_1}(s)^N]$$
$$= E[\exp(N \log \phi_{X_1}(s))] = m_N(\log \phi_{X_1}(s)).$$  \hspace{1cm} (2.23)
where \( m_N \) is the moment generating function of \( N \), defined by \( m_N(t) = \mathbb{E}[\exp(tS)] \), 
\( t \in (-t_0, t_0) \) for some positive \( t_0 \).
The important two examples, when \( N \) has Poisson and when \( N \) has Geometric distribution, are shown in the following.

For the Poisson case, \( N \sim \text{Poisson}(\lambda) \), the moment generating function is:
\[
m_N(t) = \exp(-\lambda(1-e^t)) \tag{2.24}
\]

Then we conclude that,
\[
\phi_S(s) = \exp(-\lambda(1-\phi_{X_i}(s))). \tag{2.25}
\]

For the Geometric case, \( N \sim \text{Geometric}(p) \), we have
\[
m_N(t) = \frac{p}{1 - (1-p)\exp(t)}. \tag{2.26}
\]

Therefore
\[
\phi_S(s) = \frac{p}{1 - (1-p)\phi_{X_i}(s)}. \tag{2.27}
\]

Moreover, assume that \( X_i \)'s are Exponentially distributed with the parameter \( \lambda \). The characteristic function of \( X_1 \) is:
\[
\phi_{X_i}(s) = \frac{\lambda}{\lambda - is}. \tag{2.28}
\]

The characteristic function for the total amount \( S \) becomes:
\[
\phi_S(s) = \begin{cases} 
\exp(-\lambda(1 - \frac{\lambda}{\lambda - is})) & \text{for Poisson case} \\
\frac{p}{1 - \frac{\lambda}{\lambda - is} \cdot (1-p)} & \text{for Geometric case}.
\end{cases} \tag{2.29}
\]

For large \( t \) and the light-tailed case such as \( X_i \)'s having exponential distribution, the Central Limit theorem is useful to approximate the probability:
\[
\sup_{x \in \mathbb{R}} \left\| \mathbb{P} \left( \frac{S(t) - \mathbb{E}[S(t)]}{\sqrt{\text{var}(S(t))}} \leq x \right) - \Phi(x) \right\| = \sup_{y \in \mathbb{R}} \left\| \mathbb{P}(S(t) \leq y) - \Phi \left( \frac{y - \mathbb{E}[S(t)]}{\sqrt{\text{var}(S(t))}} \right) \right\| \to 0, \text{ as } t \to \infty \tag{2.30}
\]

where \( \Phi \) is the distribution function of the standard normal, \( N(0,1) \). However, for the heavy-tailed case, this approximation is not satisfactory because the Central Limit theorem is assuming that the variance of \( X_i \)'s is finite, \( \text{var}[X_i] < \infty \). Therefore simulation-based approximation takes an important role in this field.
2.3.2 Application for queueing theory

In the context of queueing theory, $S$ is interpreted as a waiting time of queue. The case of M/G/1 queue, meaning that service times having General distribution, arrival process being Poisson and supposing a single-server, is one of the central issues in telecommunication systems and networks. [Sees Jr and Shortle 2002] studies the case when service times have Pareto distribution (M/P/1). They try to estimate quantiles of waiting times for different values of the Pareto parameter $\beta$. To know high-level quantiles helps a server to manage not to be down, and this is equivalent for insurance company to avoid bankruptcy in terms of mathematical model.

M/P/1 system is especially used for Internet modelling. [Gross et al. 2002] discuss the difficulties of simulating M/P/1 queues because small probabilities in the tail can cause extremely large errors.
Chapter 3

The estimator construction and sampler algorithm

In this chapter, the proposed method in Gudmundsson and Hult [2012] is presented.

3.1 Construction of the estimator for small probabilities

Throughout this paper, all stochastic variables are defined on a complete probability space \((\Omega, \mathcal{F}, P)\). Consider a sequence of iid non-negative random variables \(\{X_i, i \in \mathbb{N}\}\) with common cumulative distribution function (CDF) \(F\) and the density \(f\) with respect to the Lesbegues measure \(\mathcal{F}\). Set \(X = (X_1, X_2, ..., X_n)\). Notice that we first consider the case \(n\) being fixed. Our problem is to estimate

\[
p = \mathbb{P}(X \in A) = \int_A dF.
\]

(3.1)

Let \(F_A\) be the conditional distribution of \(X\) given \(X \in A\), and the density is given by

\[
\frac{dF_A}{dx} = \frac{f(x)I\{x \in A\}}{p}.
\]

(3.2)

Suppose that samples are from the target distribution \(F_A\), and consider a Markov chain \(\{X_t\}_{t=1}^T\). For any non-negative function \(v, \int_A v(x)dx = 1\) it follows.

\[
\mathbb{E}_{F_A} \left[ \frac{v(X)I\{X \in A\}}{f(X)} \right] = \int_\Omega \frac{v(x)I\{x \in A\}}{f(x)} dF_A(x)
\]

\[
= \int_A \frac{v(x)f(x)}{f(x)p} dx = \frac{1}{p} \int_A v(x)dx = \frac{1}{p}.
\]

(3.3)
Therefore an unbiased estimator $\hat{q}$ for $q = 1/p$ is calculated as
\[
\frac{1}{T} \sum_{t=1}^{T} \frac{v(X_t)I\{X_t \in A\}}{f(X_t)}.
\]
(Note that the estimation above is supposing stationarity of the Markov chain, having its invariant distribution $F_A$. It means also the burn-in period with enough length should be discarded.)

The function $v(x)$ determines a variance of the estimator. For the sake of efficient estimation, the variance is required to be small. Gudmundsson and Hult [2012] showed that when $X_i$’s are heavy-tailed in the sense that there exists a sequence $\{a_n\}$,
\[
\lim_{n \to \infty} \frac{\mathbb{P}(\max(X_1, X_2, ..., X_n) > a_n)}{\mathbb{P}(X_1 + X_2 + ... + X_n > a_n)} = \lim_{n \to \infty} \frac{\mathbb{P}(M_n > a_n)}{\mathbb{P}(S_n > a_n)} = 1,
\]
and choosing $v(\cdot) = \mathbb{P}(X \in \cdot | M_n > a_n)$, then the following estimator has vanishing normalized variance,
\[
\hat{q}^{(a)} = \frac{1}{T} \sum_{t=1}^{T} I\{M_n > a_n\}
\]
meaning that:
\[
\lim_{n \to \infty} (p)^2 \text{Var}_{F^{(a)}}(\hat{q}^{(a)}) = 0.
\]

3.2 Description of the MCMC algorithm

To sample from the conditional distribution $\mathbb{P}(X \in A) = \mathbb{P}(S_N > a_n)$ by MCMC, we use Gibbs sampler for the MCMC kernel. Description of the algorithm is as follows.

The algorithm

1. Sample $N_0$ from $\mathbb{P}(N \in \cdot)$, and set the initial state $X_0 = (X_{0,1}, ..., X_{N_0,1})$, so that satisfies $S_{0,N_0} = \sum_{i=1}^{N_0} X_{0,i} > a_n$.

2. Suppose that $X_t = (k_t, x_{t,1}, ..., x_{t,k_t})$ holding the condition $x_{t,1} + ... + x_{t,k_t} > a_n$ and let $k^*_t = \min\{j | x_{t,1} + ... + x_{t,j} > a_n\}$

\[Rare-event Simulation with MCMC\]
3. Iterate (a)-(c) until the enough length of Markov Chain is constructed.

(a) Sample the number of sum $N_{t+1}$ from the conditional distribution

$$p(k_{t+1}|k_t > k_t^*) = \frac{\mathbb{P}(N = k_{t+1})I\{k_{t+1} > k_t^*\}}{\mathbb{P}(k_{t+1} > k_t^*)}.$$  

If $N_{t+1} > N_t$, then sample $X_{t+1,k_t+1},...,X_{t+1,k_t+1}$ from $F_X$ independently and set $X_t' = (X_{t,1},...,X_{t,k_t},X_{t+1,k_t+1},...,X_{t+1,N_t+1})$

(b) Make an order of updating step \{j_1, j_2, ..., j_{N_t+1}\}, it is equivalent as a group of \{1, ..., N_t+1\}

(c) Update $X_t'$ for each $k = 1, ..., N_{t+1}$ as follows

i. Let $j = j_k$ and $X_{t,-j}' = (X_{t,1}',...,X_{t,j-1}',X_{t,j+1}',...,X_{t,N_t+1}')$.
   Sample $Z_j$ from the conditional distribution

   $$P(Z_j \in \cdot | X_{t,-j}') = \mathbb{P}(X \in \cdot | X + \sum_{k \neq j} X_{t,k}' > a_n).$$

ii. Set

   $$X_t' = (X_{t,1},...,X_{t,j-1},X_{t,j},X_{t+1,k_t+1},...,X_{t+1,N_t+1}),$$

   and return to i step.

(d) Draw an uniform random permutation $\pi$ of the numbers \{1, ..., N_{t+1}\} and put $X_{t+1} = (N_{t+1},X_{t+1,\pi(1)},...,X_{t+1,\pi(N_{t+1})}).$
Chapter 4

Assessing convergence

As we mentioned, to determine the length of the burn-in period is a difficult but also important matter. We try to determine burn-in period by using two methods, proposed by Gelman and Rubin [1992] and Geweke et al. [1991], in this chapter.

4.1 Gelman&Rubin’s method

The method in Gelman and Rubin [1992] is constructed as follows.

1. First simulate $m$ Markov chains, each chain has the length of $2n$ and each starting point $\{X^i_0| i = 1, ..., m\}$ should be overdispersed.

2. For any scalar function of interest $\theta$, calculate

$$\sqrt{R} = \sqrt{\left(\frac{n-1}{n} + \frac{m + 1}{mn} B \right) \frac{df + 1}{df + 3}} \quad (4.1)$$

where,

$$W = \frac{1}{m(n-1)} \sum_{i=1}^m \sum_{t=n+1}^{2n} (\theta^i_t - \bar{\theta}_i)^2, B = \frac{1}{m-1} \sum_{i=1}^m (\bar{\theta}_i - \bar{\theta}^\cdot)^2,$$

$$\bar{\theta}_i^\cdot = \frac{1}{2n} \sum_{t=n+1}^{2n} \theta^i_t, \bar{\theta}^\cdot = \frac{1}{m} \sum_{i=1}^m \bar{\theta}_i / m,$$

and $\theta^i_t$ is $t$ th observation from chain $i$.

$$df = \frac{2\hat{V}^2}{\text{Var}(\hat{V})}$$

$$\hat{V} = \left(\frac{n-1}{n} W + \frac{m + 1}{mn} B \right)$$
They insist that when $\sqrt{R}$ is close to 1, then we can conclude that each chain approaches the target distribution. For the purpose of determining the burn-in period which we shall use in following numerical experiments, we describe the plot of $\sqrt{R}$ with respect to iteration number in Figure 4.1. In our case, the statistics $\theta$ is chosen to be $S_N$. $\sqrt{R}$ converges to 1 and, 1000 is enough for the

![Figure 4.1: $\sqrt{R}$ with respect to the number of iteration steps where $X_i \sim \text{Pareto}(1), N \sim \text{Geometric}(0.2)$.](image)

iteration number, or may appear to be 500.

### 4.2 Geweke’s method

On the other hand, Geweke et al. [1991] insist the following.

1. Run a long Markov chain with the length $n$ and set $n_A, n_B$ satisfying $n_A + n_B < n$.

2. Calculate

$$W = \frac{\bar{\theta}_{n_A} - \bar{\theta}_{n_B}}{\sqrt{\hat{S}_G^A(0)/n_A + \hat{S}_G^B(0)/n_B}},$$

(4.2)

where $\hat{S}_G^A$ and $\hat{S}_G^B$ are spectral density estimates for $\{\theta_i | i = 1, ..., n_A\}$ and $\{\theta_j | j = n - n_B + 1, ..., n\}$ respectively, and

$$\bar{\theta}_{n_A} = \frac{\sum_{k=1}^{n_A} \theta_k}{n_A}, \quad \bar{\theta}_{n_B} = \frac{\sum_{k=n-n_B+1}^{n} \theta_k}{n_B}.$$

3. $W \to N(0, 1)$ as $n \to \infty$ if the sequence $\{\theta_i\}$ is stationary.
W is sometimes mentioned as z-score. We calculate it with respect to number of iteration, in figure 4.2. The upper and lower green lines represent the 99% line of $N(0, 1)$. It appears that 1000 is enough for the iteration number.

Figure 4.2: z-score with respect to the number of iteration steps where $X_i \sim \text{Pareto}(1), N \sim \text{Geometric}(0.2)$. 

Rare-event Simulation with MCMC
Chapter 5

Numerical experiments

In this chapter, we show the asymptotic behaviour of different distributions. Through this section, burn in period is set as 1000, and the number of sampling iteration is 5000. Each estimate is computed with different 25 chains. We show the two quantities, the ratio $p_{\text{max}}$ to $\hat{p}$ and the relative error:

$$\frac{p_{\text{max}}}{\hat{p}} = \frac{\mathbb{P}(M_n > a_n)}{\mathbb{P}(S_n > a_n)} = \frac{1 - g_N(F_X(a_n))}{\mathbb{P}(S_n > a_n)}, \quad RE = \frac{\text{Standard deviation}}{\text{Mean}}, \quad (5.1)$$

to see the tail behaviour.

First consider the case $X_i$ having Pareto distribution with density

$$f(x) = \frac{\beta}{(1 + x)^{\beta+1}},$$

and $N$ being Geometric with density $\mathbb{P}(N = k) = \rho(1 - \rho)^k$. We set parameters as $\beta = 1.0$ and $\rho = 0.2$ in Figure 5.1 and Figure 5.2. For the ease of notation, we write this setting as Pareto/Geometric and so on. Notice also that we change the threshold $a = a_n/E[N]$ instead of changing $a_n$ directly. We also consider different parameters for Pareto/Geometric, and calculate the probability $\hat{p}^* = \max\{\hat{p}\frac{p_{\text{max}}}{\hat{p}} > 0.99\}$ approximately, in Table 5.1. We notice that because our interest is to see the behaviour of tails, $\hat{p}^*$ is calculated approximately. It means the value is not as accurate as the proposed estimator denoted by $\hat{p}$.

<table>
<thead>
<tr>
<th>pareto/geometric</th>
<th>$\beta = 0.5$</th>
<th>$\beta = 1$</th>
<th>$\beta = 2.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho = 0.2$</td>
<td>9.10E-003</td>
<td>6.31E-004</td>
<td>5.79E-008</td>
</tr>
<tr>
<td>$\rho = 0.05$</td>
<td>9.50E-003</td>
<td>5.05E-004</td>
<td>7.25E-009</td>
</tr>
<tr>
<td>$\rho = 0.01$</td>
<td>1.05E-002</td>
<td>5.05E-004</td>
<td>6.49E-010</td>
</tr>
</tbody>
</table>
CHAPTER 5. NUMERICAL EXPERIMENTS

Figure 5.1: \( \hat{p} \) with respect to threshold, Pareto/Geometric, \( \beta = 1, \rho = 0.2 \).

Figure 5.2: Pareto/Geometric, \( \beta = 1, \rho = 0.2 \) (a) The ratio \( p_{\text{max}} \) to \( \hat{p} \) (b) The relative error of estimate, both are with respect to threshold \( a \).

Next, we consider the case \( X_i \) having log-normal distribution with density

\[
f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left\{ -\frac{(\ln x - \mu)^2}{2\sigma^2} \right\},
\]

and \( N \) being Geometric. We set parameters as \( \mu = 0, \sigma = 2, \rho = 0.2 \) in Figure 5.3 and Figure 5.4. In this case, we also consider different parameters, and calculate the probability \( \hat{p}^* \) approximately, in Table 5.2.
Figure 5.3: \( \hat{p} \) with respect to threshold. Lognormal/Geometric, \( \mu = 0, \sigma = 2, \rho = 0.2 \).

Figure 5.4: Lognormal/Geometric, \( \mu = 0, \sigma = 2, \rho = 0.2 \) (a) The ratio \( p_{\text{max}} \) to \( \hat{p} \) (b) The relative error of estimate, both are with respect to threshold \( a \).

Next, we consider the case \( X_i \) having Weibull distribution with density

\[
 f(x) = \frac{\beta}{\alpha} \left( \frac{x}{\alpha} \right)^{\beta-1} \exp \left\{ -\left( \frac{x}{\alpha} \right)^\beta \right\},
\]

and \( N \) being Geometric. We set parameters as \( \alpha = 1, \beta = 0.2, \rho = 0.2 \) in figure 5.5 and figure 5.6. In this case, we also consider different parameters, and calculate the probability \( \hat{p}^* \) approximately, in table 5.3.

Note that NA means the value could not be computed with the precision accuracy of Matlab, less than the order of \( 10^{-10} \).
Table 5.2: Approximated $\hat{p}^*$ with Lognormal/Geometric

<table>
<thead>
<tr>
<th>Lognormal/Geometric</th>
<th>$\sigma = 6$</th>
<th>$\sigma = 4$</th>
<th>$\sigma = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho = 0.2$</td>
<td>1.21E-002</td>
<td>1.80E-003</td>
<td>2.04E-006</td>
</tr>
<tr>
<td>$\rho = 0.05$</td>
<td>6.34E-003</td>
<td>9.15E-004</td>
<td>1.35E-007</td>
</tr>
<tr>
<td>$\rho = 0.01$</td>
<td>4.20E-003</td>
<td>4.59E-004</td>
<td>3.16E-009</td>
</tr>
</tbody>
</table>

Figure 5.5: $\hat{p}$ with respect to threshold. Weibull/Geometric, $\alpha = 1, \beta = 0.2, \rho = 0.2$. 

Next, we consider the case $X_i$ having Pareto distribution, and $N$ having Poisson distribution with density $P(N = k) = \lambda^k \exp(-\lambda) / k!$. We set parameters as $\beta = 1, \lambda = 5$ in figure 5.7 and figure 5.8.

Table 5.3: Approximated $\hat{p}^*$ with Weibull/Geometric

<table>
<thead>
<tr>
<th>Weibull/Geometric</th>
<th>$\beta = 0.1$</th>
<th>$\beta = 0.15$</th>
<th>$\beta = 0.2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho = 0.2$</td>
<td>5.55E-003</td>
<td>4.14E-004</td>
<td>1.33E-005</td>
</tr>
<tr>
<td>$\rho = 0.05$</td>
<td>1.78E-003</td>
<td>7.63E-005</td>
<td>3.98E-007</td>
</tr>
<tr>
<td>$\rho = 0.01$</td>
<td>5.96E-004</td>
<td>6.14E-006</td>
<td>NA</td>
</tr>
</tbody>
</table>
CHAPTER 5. NUMERICAL EXPERIMENTS

Figure 5.6: Weibull/Geometric, $\alpha = 1, \beta = 0.2, \rho = 0.2$ (a) The ratio $p_{max}$ to $\hat{p}$ (b) The relative error of estimate, both are with respect to threshold $\alpha$.

Figure 5.7: $\hat{p}$ with respect to threshold, Pareto/Geometric, $\beta = 1, \lambda = 5$.

Next, we consider the case $X_i$ having log-normal distribution, and $N$ having Poisson distribution. We set parameters as $\mu = 0, \sigma = 2, \lambda = 5$ figure 5.3 and figure 5.10.
Next, we consider the case $X_i$ having Weibull distribution, and $N$ having Poisson distribution. We set parameters as $\alpha = 1, \beta = 0.2, \lambda = 5$. For each distribution, accuracy of the estimates are almost same, or rather better than Pareto/Geometric case. It means the MCMC based sampling method works efficient in various settings. To use asymptotic approximation, $a = 10^4$ is at least needed in our setting.
Figure 5.10: Lognormal/Poisson, $\mu = 0, \sigma = 2, \lambda = 5$ (a) The ratio $p_{max}$ to $\hat{p}$ (b) The relative error of estimate, both are with respect to threshold $a$.

Figure 5.11: $\hat{p}$ with respect to threshold. Weibull/Poisson, $\alpha = 1, \beta = 0.2, \lambda = 5$.

Figure 5.12: Weibull/Poisson, $\alpha = 1, \beta = 0.2, \lambda = 5$ (a) The ratio $p_{max}$ to $\hat{p}$ (b) The relative error of estimate, both are with respect to threshold $a$. 
Chapter 6

Quantile estimation

In this chapter, the proposal approach to estimate quantiles and the numerical result are shown.

6.1 Procedure of estimation

In previous sections, the method to estimate probabilities has been shown. That method is supposing a given arbitrary threshold \( a_n \) and then compute the corresponding probability. Now our problem is opposite, to compute a percentile given a probability. For this solution, a trivial property is used:

\[ M_n = \max(X_1, ..., X_n) > S_n = X_1 + ... + X_n \]
\[ \Rightarrow P(M_n < x) > P(S_n < x) \]
\[ \Rightarrow G^{-1}(y) < H^{-1}(y), \quad (6.1) \]

where

\[ G(x) = P(M_n < x), \quad H(x) = P(S_n < x). \]

Let \( \{S_N^T\}_{j=1} \) be samples from the conditional distributions, \( P(S_N | S_N > b) \) and define the following given \( a_n > b \):

\[ p := P(S_N > a_n) = P(S_N > a_N | S_N > b)P(S_N > b). \quad (6.2) \]

We also define the empirical survival distribution,

\[ \hat{F}_T(x) = \frac{1}{T} \sum_{i=1}^{T} I\{S_j > x\}. \quad (6.3) \]
Then our motivation of quantile estimation method is derived from the following:

$$F_{SN}^{-1}(1-p) = \inf\{x : F_{SN}(x) \geq 1-p\}$$

$$= \inf\{x : F_{SN}(x) \leq p\}$$

$$\approx \inf\{x : F_T(x) \hat{p}_b \leq p\}$$

$$= \inf\{x : \hat{F}_T(x) \leq \frac{p}{\hat{p}_b}\}$$

$$= F_T^{-1}(1 - \frac{p}{\hat{p}_b}) = S_{N}^{T\hat{p}_b+1}, \quad (6.4)$$

where \( \hat{p}_b \) is the estimated value of \( \mathbb{P}(S_N > b) \), \([\cdot]\) denotes the floor function and \( S_{N}^{k} \) denotes the \( k \)th order statistics of \( \{S_N\} \).

However, how to select the pre-determined threshold \( b \) smaller than \( a_n \) is a problem, because we do not know \( a_n \). In addition, there are two hopeful requirements for selecting \( b \):

- \( b \) should be generated endogenously using the input \( p \). Otherwise the method becomes unstable and depends on user’s experience.
- \( b \) must be less than \( a_n \) but also needed to be close to \( a_n \), for the sake of efficiency of the estimation.

To solve these requirements, we use the property \([6.1]\). The procedure of estimation is as follows.

1. Set the interest small probability \( p \), and calculate

   $$b = G^{-1}(1-p) = F_X^{-1}(g^{-1}_X(1-p)) \cdot$$

2. Sample \( \{S_j| j = 1, ..., T\} \) from \( \mathbb{P}(|S_n > b) \) and calculate \( \hat{p}_b \), the estimate of \( \mathbb{P}(S_n > b) \) with using the MCMC described in Chapter 3.

3. Let \( \{\tilde{S}_j\} \) be the order statistics of \( \{S_j\} \) and take

   $$\hat{a}_n = \tilde{S}_{j^*} \quad \text{where, } j^* = [T \frac{p}{\hat{p}_b}] + 1$$

It is obvious that the way of selecting \( b \) satisfies that \( b \) is smaller than \( a_n \) and that \( b \) is generated endogenously. The property that \( b \) is close to \( a_n \), is shown by the subexponential property \((2.2.2)\) (reappeared):

$$\lim_{x \to \infty} \frac{\mathbb{P}(\max(X_1, X_2, ..., X_N) > x)}{\mathbb{P}(X_1 + X_2 + ... + X_N > x)} = \lim_{x \to \infty} \frac{\mathbb{P}(M_N > x)}{\mathbb{P}(S_N > x)} = 1.$$

This means that:

$$H^{-1}(y) \to G^{-1}(y), \quad \text{as } y \to 1. \quad (6.5)$$

Therefore this way of selecting \( b \) is efficient for estimating high-level quantiles.

---

*Rare-event Simulation with MCMC*
6.2 Numerical experiments

To see the accuracy of estimation, the relative error is shown in the following. We simulate 5000 number of iteration and 1000 number of burn-in period with 25 batches for calculating each estimates.
Table 6.1: Estimated quantiles and relative errors

<table>
<thead>
<tr>
<th>β = 1, ρ = 0.2</th>
<th>90%-quantile</th>
<th>95%-quantile</th>
<th>99%-quantile</th>
<th>99.9%-quantile</th>
<th>99.99%-quantile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pareto/Geometric</td>
<td>90%</td>
<td>95%</td>
<td>99%</td>
<td>99.9%</td>
<td>99.99%</td>
</tr>
<tr>
<td>Estimated quantile</td>
<td>74.0104590041</td>
<td>134.8339169221</td>
<td>553.7625066222</td>
<td>5072.8994369009</td>
<td>50100.7159356415</td>
</tr>
<tr>
<td>Relative error</td>
<td>0.0118412429</td>
<td>0.0096959224</td>
<td>0.0050031339</td>
<td>0.0016162748</td>
<td>0.0008733781</td>
</tr>
<tr>
<td>β = 1.5, ρ = 0.2</td>
<td>24.3031966069</td>
<td>36.585318664</td>
<td>83.2151448041</td>
<td>311.7404647179</td>
<td>1376.5918363372</td>
</tr>
<tr>
<td>Estimated quantile</td>
<td>0.0120766040</td>
<td>0.0126649585</td>
<td>0.0071424524</td>
<td>0.0034849575</td>
<td>0.0013393708</td>
</tr>
<tr>
<td>β = 2, ρ = 0.2</td>
<td>13.5903702139</td>
<td>19.1072095647</td>
<td>35.4203254331</td>
<td>82.247434762</td>
<td>233.326910947</td>
</tr>
<tr>
<td>Estimated quantile</td>
<td>0.0100780445</td>
<td>0.0155471971</td>
<td>0.0093854448</td>
<td>0.0056055904</td>
<td>0.0023464803</td>
</tr>
<tr>
<td>β = 1, ρ = 0.05</td>
<td>352.520629686</td>
<td>603.5021383505</td>
<td>2276.0904465504</td>
<td>20366.03299757</td>
<td>200416.37557328</td>
</tr>
<tr>
<td>Estimated quantile</td>
<td>0.0110692514</td>
<td>0.0118317257</td>
<td>0.0054489569</td>
<td>0.0023373697</td>
<td>0.0008168677</td>
</tr>
<tr>
<td>β = 1.5, ρ = 0.05</td>
<td>94.340598725</td>
<td>132.7156957951</td>
<td>258.3808683199</td>
<td>821.289723797</td>
<td>3501.6328598237</td>
</tr>
<tr>
<td>Estimated VaR</td>
<td>0.0124918796</td>
<td>0.0133734075</td>
<td>0.0086742283</td>
<td>0.004153514</td>
<td>0.0019351215</td>
</tr>
<tr>
<td>β = 2, ρ = 0.05</td>
<td>48.9253184649</td>
<td>65.842624951</td>
<td>109.6603339635</td>
<td>201.6334181388</td>
<td>490.6275220033</td>
</tr>
<tr>
<td>Estimated VaR</td>
<td>0.0110082533</td>
<td>0.0101320498</td>
<td>0.0187150096</td>
<td>0.0120333408</td>
<td>0.0028293819</td>
</tr>
</tbody>
</table>
6.3 Example with real date: Danish Fire Insurance 1980-1990

In this section, we illustrate a calculation of a quantile with the real data set, Danish fire insurance data. Its period is from January 1st, 1980 until December 31st, 1990, available at [http://www.macs.hw.ac.uk/~mcneil/data.html](http://www.macs.hw.ac.uk/~mcneil/data.html).

With the data, the model of the annual claim amount is estimated as:

\[
F_X(x) = k^\alpha (k + x)^{(-\alpha)}, N \sim \text{Poiss} (\lambda),
\]

and the parameters are:

\[
\alpha = 1.7, k = 2.2, \lambda = 197.
\]

Then the 0.005\% level VaR, which is the level recommended in the Solvency II framework, is estimated as 1743.53 million Danish Kronor, and Relative error = 0.004422. Compare with the annual expected value,

\[
E[S_N] = E[N]E[X_i] = \lambda \frac{k^\alpha}{\alpha - 1} = 619.143,
\]

the VaR is approximately 3 times of the expectation.
Chapter 7

Conclusion

We showed the efficiency of the MCMC based method and the tail behaviour for various parameters and distributions respectively. For the tail behaviour, we can conclude that the tail index is bigger than 1, then it is inadequate to use asymptotic estimator to estimate probabilities which is smaller than the order of $10^{-4}$. Considering the data in Section 6.3 one should not use the asymptotic estimator to estimate probability which is bigger than $10^{-7}$.

Our proposal method to estimate quantiles has two distinctive features. First, the estimates are remarkably accurate and as the quantile becomes higher, the relative errors becomes smaller. This property is significant because other methods such as standard Monte Carlo have opposite property. Second, the ease of use, the algorithm does not require complicated analytical calculation.
Acknowledgements

At first I would like to express my deepest gratitude for Professor Henrik Hult, my supervisor, to spend time for discussion and giving significant comments on my project.

At the same time I do thank to my parents for supporting my study in KTH, a precious opportunity.


