Markov Chain Monte Carlo for Risk Measures

Yuya Suzuki[1,2] and Thorbjörn Gudmundsson[3]

[1] *Double Degree Program between School of Engineering Science KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden*

[2] *School of Science for Open and Environmental Systems Keio University, Yokohama, Japan*

[3] *Department of Mathematics Royal Institute of Technology, SE-100 44 Stockholm, Sweden*

Keywords: Markov Chain Monte Carlo, Risk Measures,Heavy Tails, Rare-event Simulation.

Abstract: In this paper, 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 construct an efficient method to calculate tail-based risk measures such as quantiles and conditional expectation (expected shortfalls). When assuming extreme quantiles and heavy-tailed increments, using standard Monte Carlo method can be inefficient. In previous works, there exists an efficient method to sample rare-events (tail-events) using a Markov chain Monte Carlo (MCMC) with a given threshold. We apply the sampling method to estimate statistics based on tail-information, with a given rare-event probability. The performance is compared with other methods by some numerical results in the case increments follow Pareto distributions . We also show numerical results with Weibull, and Log-Normal distributions. Our proposed method is shown to be efficient especially in cases of extreme tails.

1 INTRODUCTION

Simulation methods with intent to estimate extreme tails of distributions are of practical importance in various areas such as telecommunication systems, risk management for insurance and operational risk, etc. For instance, to assess the risk of rare-events, it helps an insurance company to avoid bankruptcy, also helps a server to avoid being down.

Suppose 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 follows a Poisson distribution, then the setting becomes equivalent to the waiting time of an M/G/1 queue system. In the context of queuing theory, especially for the application with high service time X_i, especially the case X_i following a 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 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 $\hat{p} = \sum_{t=1}^{T} I_{\{S_n^t > a_n\}}/T$. However, this method is inefficient to compute small probabilities with high threshold a_n. Let us see the relative error of the estimate:

$$\frac{\mathrm{Std}(\hat{p})}{\hat{p}} = \sqrt{\frac{1-\hat{p}}{T\,\hat{p}}} \to \infty,$$

as $a_n \to \infty$. This shows the difficulty of computation when using the standard Monte Carlo.

There are alternatively proposed 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 $\widetilde{\mathbb{P}}$, see (Rubino et al., 2009; Blanchet and Liu, 2008).

However, we use the method based on MCMC 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. For the MCMC method, implementation itself does not require us to do complex computation.The other reason is that (Gudmundsson

Suzuki Y. and Gudmundsson T..
Markov Chain Monte Carlo for Risk Measures.
DOI: 10.5220/0005035303300338
In *Proceedings of the 4th International Conference on Simulation and Modeling Methodologies, Technologies and Applications* (SIMULTECH-2014), pages 330-338
ISBN: 978-989-758-038-3

and Hult, 2012) showed the MCMC based method is compatible or rather better than existing importance sampling methods. The MCMC estimator is constructed as non-biased, 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 $n = N$ has Geometric distribution.

MCMC, stands for Markov chain Monte Carlo, is initially proposed by (Metropolis et al., 1953) to simulate the energy levels of atoms, and generalized by (Hastings, 1970) . Suppose that we are interested in sampling from a distribution $\pi(x)$, but we can not sample from the distribution directly. MCMC is one of the solutions for the problem, by constructing a Markov chain whose stationary (equilibrium) distribution is $\pi(x)$. States of the Markov chain after sufficient steps of transition are used for sampling.

To estimate quantile of S_n is an important 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 queuing 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 proposed method is remarkably accurate.

The rest of this paper is organized as follows. In Section 2, the MCMC-based method to estimate tail-based statistics is provided. In Section 3, we assess the convergence of the MCMC.In Section 4, numerical results are shown. In Section 5, conclusions are given.

2 ESTIMATION METHOD

In this section, the proposed method in (Gudmundsson and Hult, 2012) for sampling rare-events and estimating small probabilities using the MCMC is presented.

2.1 Estimator Construction

Throughout this paper, all stochastic variables are defined on a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$. 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 $\mathbf{X} = (X_1, X_2, \ldots, X_n)$. Notice that we first consider the case n being fixed. Our problem is to estimate

$$p = \mathbb{P}(\mathbf{X} \in A) = \int_A dF. \tag{1}$$

Let F_A be the conditional distribution of $\mathbf{X}$ given $\mathbf{X} \in A$, and the density is given by

$$\frac{dF_A}{dx} = \frac{f(x)I(x \in A)}{p}. \tag{2}$$

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

$$\begin{aligned} \mathbb{E}_{F_A}\left[\frac{v(\mathbf{X})I\{\mathbf{X} \in A\}}{f(\mathbf{X})}\right] &= \int_\Omega \frac{v(x)I\{x \in A\}}{f(x)} dF_A(x) \\ &= \int_A \frac{v(x)f(x)}{f(x)p} dx \\ &= \frac{1}{p}\int_A v(x)dx = \frac{1}{p}. \end{aligned} \tag{3}$$

Therefore an unbiased estimator $\widehat{q_T}$ for $q = 1/p$ is calculated as

$$\frac{1}{T}\sum_{t=1}^{T} \frac{v(\mathbf{X}_t)I\{\mathbf{X}_t \in A\}}{f(\mathbf{X}_t)}. \tag{4}$$

Note that the estimation above is assuming stationarity of the Markov chain, having its invariant distribution F_A. It means also the burn-in period with sufficient 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 $\{a_n\}$,

$$\lim_{n\to\infty} \frac{\mathbb{P}(\max(X_1, X_2, \ldots, X_n) > a_n)}{\mathbb{P}(X_1 + X_2 + \cdots + X_n > a_n)} = \lim_{n\to\infty} \frac{\mathbb{P}(M_n > a_n)}{\mathbb{P}(S_n > a_n)} = 1, \tag{5}$$

and choosing $v(\cdot) = \mathbb{P}(\mathbf{X} \in \cdot | M_n > a_n)$, then the following estimator has vanishing normalized variance,

$$\widehat{q_T}^{(n)} = \frac{1}{T}\sum_{t=1}^{T} \frac{I\{M_n > a_n\}}{1 - F_X(a_n)^n}, \tag{6}$$

meaning that:

$$\lim_{n\to\infty} (p)^2 \mathrm{Var}_{F^{(n)}_{A_n}}(q_T^{(n)}) = 0. \tag{7}$$

Note that the condition (5) holds for large class of distributions, including subexponential class. For the case when $n = N$ is non-negative integer valued random variable, the estimator becomes

$$\widehat{q_T}^{(N)} = \frac{1}{T}\sum_{t=1}^{T} \frac{I\{M_N > a_n\}}{1 - g_N(F_X(a_n))}, \tag{8}$$

where g_N is the probability generating function of N, defined by $g_N(z) = \mathbb{E}[z^N]$.

2.2 Description of the MCMC Algorithm

To sample from the conditional distribution

$$\mathbb{P}(\mathbf{X} \in A) = \mathbb{P}(S_N = X_1 + \cdots + X_N > a_n), \quad (9)$$

by the MCMC. We use Gibbs sampler for the MCMC kernel. Description of the algorithm is as follows.

1. Sample N_0 from $\mathbb{P}(N \in \cdot)$, and set the initial state
$$X_0 = (X_{0,1}, \ldots, X_{N_0,1}).$$
so that satisfies
$$S_{0,N_0} = \sum_{i=1}^{N_0} X_{0,i} > a_n.$$
2. Suppose that
$$\mathbf{X_t} = (k_t, x_{t,1}, \ldots, x_{t,k_t}),$$
holding the condition
$$x_{t,1} + \ldots + x_{t,k_t} > a_n,$$
and let
$$k_t^* = \min\{j | x_{t,1} + \ldots + x_{t,j} > a_n\}.$$
3. Iterate (a)-(c) until the sufficient length of Markov Chain is constructed.
   (a) Sample the number of sum N_{t+1} from the conditional distribution
   $$p(k_{t+1}|k_{t+1} > k_t^*) = \frac{\mathbb{P}(N = k_{t+1}) I\{k_{t+1} > k_t^*\}}{P(k_{t+1} > k_t^*)}.$$
   If $N_{t+1} > N_t$, then sample
   $$X_{t+1,N_t+1}, \ldots, X_{t+1,N_{t+1}}$$
   from F_X independently and set
   $$\mathrm{X_t}' = (X_{t,1}, \ldots, X_{t,N_t}, X_{t+1,N_t+1}, \ldots, X_{t+1,N_{t+1}})$$
   (b) Make an order of updating step $\{j_1, j_2, \ldots, j_{N_{t+1}}\}$, it is equivalent as a group of $\{1, \ldots, N_{t+1}\}$
   (c) Update $\mathrm{X_t}'$ for each $k = 1, \ldots, N_{t+1}$ as follows
      i. Let $j = j_k$ and
      $$\mathrm{X}'_{\mathrm{t},-\mathrm{j}} = (X'_{t,1}, \ldots, X'_{t,j-1}, X'_{t,j+1}, \ldots, X'_{t,N_{t+1}}).$$
      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 $X_{t,j} = Z_j$ and
      $$\mathrm{X_t}' = (X_{t,1}, \ldots, X_{t,j-1}, X_{t,j}, X_{t+1,k_t+1}, \ldots, X_{t+1,N_{t+1}}),$$
      then return to i step.
   (d) Draw an uniform random permutation π of the numbers $\{1, \ldots, N_{t+1}\}$ and put
   $$\mathbf{X_{t+1}} = (N_{t+1}, X_{t+1,\pi(1)}, \ldots, X_{t+1,\pi(N_{t+1})}).$$

2.3 Procedure of the Estimation

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

$$\begin{aligned} &M_n = \max(X_1, \ldots, X_n) < S_n = X_1 + \ldots + X_n \\ &\Rightarrow \mathbb{P}(M_n < x) > \mathbb{P}(S_n < x) \\ &\Rightarrow G^{-1}(y) < H^{-1}(y), \end{aligned} \quad (10)$$

where

$$G(x) = \mathbb{P}(M_n < x), H(x) = \mathbb{P}(S_n < x).$$

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

$$p := \mathbb{P}(S_N > a_n) = \mathbb{P}(S_N > a_N | S_N > b)\mathbb{P}(S_N > b). \quad (11)$$

We also define the empirical survival distribution,

$$\bar{F_T}(x) = \frac{1}{T} \sum_{i=1}^{T} I\{S_j > x\}. \quad (12)$$

Then our motivation of quantile estimation method is derived from the following:

$$\begin{aligned} F_{S_N}^{-1}(1-p) &= \inf\{x : F_{S_N}(x) \geq 1-p\} \\ &= \inf\{x : \bar{F}_{S_N}(x) \leq p\} \\ &\approx \inf\{x : \bar{F_T}(x)\hat{p}_b \leq p\} \\ &= \inf\{x : \bar{F_T}(x) \leq \frac{p}{\hat{p}_b}\} \\ &= \bar{F}_T^{-1}(1 - \frac{p}{\hat{p}_b}) = S_N^{\lfloor T\frac{p}{\hat{p}_b}\rfloor+1,T} \end{aligned} \quad (13)$$

where $\hat{p}_b$ is the estimated value of $\mathbb{P}(S_N > b)$, $\lfloor \cdot \rfloor$ denotes the floor function and $S_N^{k,T}$ denotes the kth order statistics of $\{S_N^j\}$.

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 (10). The procedure of estimation is as follows.

1. Set the interest small probability p, and calculate
$$b = G^{-1}(1-p) = F_X^{-1}(g_N^{-1}(1-p)).$$
2. Sample $\{S_j | j = 1, \ldots, T\}$ from $\mathbb{P}(\cdot | S_n > b)$ and calculate $\hat{p_b}$, the estimate of $\mathbb{P}(S_n > b)$ by using MCMC described in the previous section.
3. Let $\{\tilde{S}_j\}$ be the order statistics of $\{S_j\}$ and take
$$\hat{a_n} = \tilde{S}_{j^*} \text{ where, } j^* = \lfloor T \frac{p}{\hat{p_b}} \rfloor + 1.$$

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 :

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

This means that:

$$H^{-1}(y) \to G^{-1}(y), \quad as\, y \to 1. \tag{14}$$

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

Extensively, we construct an estimator for conditional expectations, mentioned as expected shortfall:

$$ES_\alpha = \mathbb{E}[S_N | S_N > F_{S_N}^{-1}(1-p)] = \frac{1}{p} \int_{1-p}^{1} F_{S_N}^{-1}(u) du. \tag{15}$$

The estimator for this statistics is

$$\widehat{ES}_\alpha = \frac{1}{\#\{\tilde{S}_j > \tilde{S}_{j^*}\}} \sum_{\tilde{S}_j > \tilde{S}_{j^*}} \tilde{S}_j, \tag{16}$$

where $\#\{\tilde{S}_j > \tilde{S}_{j^*}\}$ represents the number of components $\{\tilde{S}_j\}$ that satisfies $\tilde{S}_j > \tilde{S}_{j^*}$.

3 ASSESSING CONVERGENCE

To determine the length of the burn-in period is a difficult but also important matter when using the MCMC. We try to determine burn-in period by using the method, proposed by (Gelman and Rubin, 1992) in this section.

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 $\{X_0^i | i = 1, \ldots, m\}$ should be over-dispersed.
2. For any scalar function of interest θ, calculate

$$\sqrt{R} = \sqrt{\left(\frac{n-1}{n} + \frac{m+1}{mn}\frac{B}{W}\right)\frac{df+1}{df+3}}, \tag{17}$$

where,

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

$$\bar{\theta_i} = \sum_{t=n+1}^{2n} \theta_i^t / n, \bar{\theta_\cdot} = \sum_{i=1}^{m} \bar{\theta_i} / m,$$

and θ_i^t is t th observation from chain i. Also

$$df = \frac{2\widehat{V}^2}{\widehat{\mathrm{Var}}(\widehat{V})}, \widehat{V} = \left(\frac{n-1}{n}W + \frac{m+1}{mn}B\right).$$

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 1. In our case, the statistics θ is chosen to be S_N. $\sqrt{R}$ converges to 1. 1000 is sufficient for the iteration number.

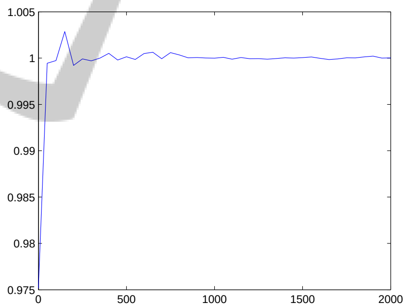

Figure 1: $\sqrt{R}$ with respect to the number of iteration steps.

4 NUMERICAL RESULTS

In this section, some numerical results are shown. To validate and compare the efficiency with other existing methods, we refer to (Hult and Svensson, 2009). They consider a problem of fixed sum, meaning the number of summands is not stochastic but constant. We first apply our method to the fixed-sum problem and show the performance, labeled as MCMC. MC represents standard Monte Carlo, DLW represents the conditional mixture algorithm (Dupuis et al., 2007) and SM represents the scaling mixture algorithm (Hult and Svensson, 2009). All figures of MC, DLW, and SM in (Hult and Svensson, 2009) are used as a reference. For each estimation, all statistics are calculated based on 10^4 samples and it is repeated 100 times to calculate the mean and standard deviation of the estimates. For the MCMC method, we set 1000 times of iteration for the burn-in period.

Table 1: Means (standard deviations). Simulating quantiles q such that $1-p = F_{S_n}(q)$ where $F_{S_n}(\cdot)$ is a cumulative distribution function of S_n, $S_n = \sum_{i=1}^n X_i$ and $\mathbb{P}(X_i > x) = (1+x)^{-2}$.

n	$1-p$	MCMC	SM	DLW	MC
10	1e-2	40.179 (0.130)	41.007 (0.246)	40.166 (0.459)	40.038 (1.78)
	1e-3	108.587 (0.197)	109.33 (0.847)	108.29 (1.081)	84.821 (47.23)
	1e-5	1008.1 (0.495)	1003.1 (18.5)	1007.5 (1.51)	609.42 (1594)
30	1e-2	84.585 (0.324)	85.841 (0.395)	84.681 (1.237)	84.362 (2.739)
	1e-3	202.557 (0.373)	203.56 (1.53)	202.29 (2.40)	171.16 (71.26)
	1e-5	1760.3 (0.903)	1753.7 (41.12)	1759.0 (1.487)	114.23 (443.5)

Table 2: Means (standard deviations). Simulating quantiles q such that $1-p = F_{S_n}(q)$ where $F_{S_n}(\cdot)$ is a cumulative distribution function of S_n, $S_n = \sum_{i=1}^n X_i$ and $\mathbb{P}(X_i > x) = (1+x)^{-3}$.

n	$1-p$	MCMC	SM	DLW	MC
10	1e-2	14.205 (0.069)	14.853 (0.090)	14.195 (0.154)	14.182 (0.305)
	1e-3	25.651 (0.062)	26.125 (0.171)	25.588 (0.412)	24.965 (2.212)
	1e-5	103.64 (0.091)	104.23 (0.799)	103.40 (0.553)	5.283 (16.03)
30	1e-2	29.845 (0.297)	31.054 (0.287)	29.943 (0.519)	29.949 (0.500)
	1e-3	46.016 (0.184)	46.725 (0.286)	46.277 (1.041)	44.608 (2.688)
	1e-5	158.057 (0.152)	158.460 (1.080)	157.620 (0.273)	13.847 (28.53)

From the table 1 and the table 2, the MCMC based method perform well comparing with other methods. For simulating conditional expectations in the table 3 and the table 4, we see MCMC based method can perform a little bit less than other methods in some parameter settings. However, when we see the extreme tail, our proposed method works the best.

Next we consider random sums, meaning the number of summands N is random. For numerical experiments, we calculate the case N is geometrically distributed:

$$\mathbb{P}(N = n) = \rho(1-\rho)^{(n-1)}, \qquad (18)$$

Table 3: Means (standard deviations). Simulating conditional expectations $\mathbb{E}(S_n|S_n > a_n)$ where $\mathbb{P}(S_n > a_n) = 1-p$, $S_n = \sum_{i=1}^n X_i$ and $\mathbb{P}(X_i > x) = (1+x)^{-2}$.

n	$1-p$	MCMC	SM	DLW	MC
10	1e-2	71.694 (1.44)	73.065 (1.06)	71.831 (1.22)	72.252 (8.75)
	1e-3	208.67 (4.16)	209.37 (3.60)	209.30 (4.99)	213.42 (65.8)
	1e-5	2005.4 (35.5)	2009.8 (37.1)	2009.3 (30.9)	4787.8 (23168)
30	1e-2	139.03 (3.74)	140.55 (2.22)	139.14 (3.09)	140.76 (17.34)
	1e-3	374.25 (6.45)	375.76 (5.00)	378.24 (11.49)	391.06 (96.36)
	1e-5	3496.7 (58.7)	3500.0 (65.2)	3496.9 (59.8)	745.7 (3671)

Table 4: Means (standard deviations). Simulating conditional expectations $\mathbb{E}(S_n|S_n > a_n)$ where $\mathbb{P}(S_n > a_n) = 1-p$, $S_n = \sum_{i=1}^n X_i$ and $\mathbb{P}(X_i > x) = (1+x)^{-3}$.

n	$1-p$	MCMC	SM	DLW	MC
10	1e-2	19.206 (0.218)	20.044 (0.167)	19.257 (0.395)	19.495 (0.905)
	1e-3	36.321 (0.282)	36.911 (0.327)	36.463 (0.776)	41.032 (5.53)
	1e-5	153.50 (0.901)	154.39 (1.326)	153.83 (2.705)	132.74 (491.7)
30	1e-2	36.922 (0.930)	38.603 (0.902)	37.200 (1.169)	37.744 (1.581)
	1e-3	61.125 (0.728)	62.013 (0.416)	62.066 (1.814)	69.369 (7.973)
	1e-5	230.22 (1.40)	230.27 (1.92)	230.00 (1.47)	225.14 (932)

with some different parameter settings. This time we simulate 5000 number of iteration and 1000 number of burn-in period with 25 batches to calculate each means and standard deviations of estimates. ES (expected shortfall) represents the conditional expectation. Note that we calculate relative errors instead of standard deviations in order to see how the accuracy changes. Table 7 shows that, when calculating quantiles, the more high quantiles we estimate, the more estimates become accurate. For expected short falls, relative errors do not change remarkably but the estimates are sufficiently accurate. For other settings, meaning X_i's follow Weibull and Log-normal distributions, see the Appendix.

Table 5: Means and relative errors. Simulating quantiles and expected shortfalls of $S_N = \sum_{i=1}^{N} X_i$ where $\mathbb{P}(X_i > x) = (1+x)^{-\beta}$, $\mathbb{P}(N = n) = \rho(1-\rho)^{(n-1)}$.

	$\beta = 3, \rho = 0.2$					
Pareto/Geometric	90%-quantile	95%-quantile	99%-quantile	99.9%-quantile	99.99%-quantile	99.999%-quantile
Estimated quantile	6.87441	9.33847	15.3579	25.8747	42.9086	84.0273
Relative error	0.01508	0.01176	0.01354	0.01151	0.004778	0.002684
	$\beta = 2, \rho = 0.2$					
Pareto/Geometric	90%-quantile	95%-quantile	99%-quantile	99.9%-quantile	99.99%-quantile	99.999%-quantile
Estimated quantile	13.61134	19.1129	35.4151	82.1348	233.466	715.999
Relative error	0.01223	0.01233	0.01058	0.006867	0.002183	0.0008949
	$\beta = 1.5, \rho = 0.2$					
Estimated quantile	24.3839	36.6780	82.9503	311.665	1375.97	6317.46
Relative error	0.01163	0.0096183	0.007067	0.002843	0.001107	0.0007270
	$\beta = 2, \rho = 0.05$					
Estimated quantile	48.8967	65.5929	109.858	202.007	490.835	1454.22
Relative error	0.01356	0.01625	0.01592	0.009312	0.0021054	0.001494
	$\beta = 1.5, \rho = 0.05$					
Estimated quantile	93.962	132.302	258.466	821.700	3500.29	15948.4
Relative error	0.01541	0.01283	0.01099	0.003944	0.001679	0.00050244
	$\beta = 3, \rho = 0.2$					
Pareto/Geometric	90%-quantile	95%-quantile	99%-quantile	99.9%-quantile	99.99%-quantile	99.999%-quantile
Estimated ES	10.3295	12.9366	19.8514	33.6573	60.7505	123.348
Relative error	0.01183	0.01346	0.01430	0.01447	0.01332	0.009787
	$\beta = 2, \rho = 0.2$					
Pareto/Geometric	90%-quantile	95%-quantile	99%-quantile	99.9%-quantile	99.99%-quantile	99.999%-quantile
Estimated ES	23.2326	31.1116	57.5991	150.307	457.184	1419.47
Relative error	0.02130	0.03127	0.03323	0.01582	0.02033	0.01836
	$\beta = 1.5, \rho = 0.2$					
Estimated ES	54.8321	83.7899	203.878	892.355	4006.29	18232.4
Relative error	0.06381	0.2062	0.07136	0.06701	0.06100	0.04304
	$\beta = 2, \rho = 0.05$					
Estimated ES	75.8177	95.6647	153.790	334.255	938.896	2867.00
Relative error	0.02580	0.03028	0.05590	0.02852	0.04177	0.02154
	$\beta = 1.5, \rho = 0.05$					
Estimated ES	179.384	249.374	572.885	2286.532	10194.5	46380.5
Relative error	0.05329	0.09100	0.1432	0.05691	0.07483	0.03651

5 CONCLUSIONS

The MCMC-based method for estimating tail-based statistics, such as quantiles and conditional expectations, was shown. Comparing with other methods, our method outperformed when estimating quantiles in the parameter settings.

Our proposed method to estimate quantiles has two distinctive features. First, the estimates are accurate. The level of quantile becomes higher, the relative errors becomes smaller. This property is significant because other methods such as standard Monte Carlo have the opposite property. Second, the ease of use, the algorithm does not require complicated analytical calculations. We also calculated the case X_i's follow Weibull and Log-normal distributions that results are shown in the Appendix. Relative errors do not change remarkably, or perform better for heavy and extreme tails.

REFERENCES

Blanchet, J. H. and Liu, J. (2008). State-dependent importance sampling for regularly varying random walks. *Advances in Applied Probability*, pages 1104–1128.

Dupuis, P., Leder, K., and Wang, H. (2007). Importance sampling for sums of random variables with regularly varying tails. *ACM T. Model Comput. S.*, 17(3).

Embrechts, P., Furrer, H., and Kaufmann, R. (2003). Quantifying regulatory capital for operational risk. *Derivatives Use, Trading & Regulation*, 9(3):217–233.

Fischer, M. J., Bevilacqua Masi, D. M., Gross, D., Shortle, J., and Brill, P. H. (2001). Using quantile estimates in simulating internet queues with pareto service times. In *Proceedings of the 33nd conference on Winter simulation*, pages 477–485. IEEE Computer Society.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, pages 457–472.

Gross, D., Shortle, J. F., Fischer, M. J., and Masi, D. M. B. (2002). Difficulties in simulating queues with pareto service. In *Simulation Conference, 2002. Proceedings of the Winter*, volume 1, pages 407–415. IEEE.

Gudmundsson, T. and Hult, H. (2012). Markov chain monte carlo for computing rare-event probabilities for a heavy-tailed random walk. *arXiv preprint*.

Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109.

Hult, H. and Svensson, J. (2009). Efficient calculation of risk measures by importance sampling–the heavy tailed case. *Preprint, KTH*.

McNeil, A. J., Frey, R., and Embrechts, P. (2010). *Quantitative risk management: concepts, techniques, and tools*. Princeton university press.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21:1087.

Rubino, G., Tuffin, B., et al. (2009). *Rare event simulation using Monte Carlo methods*. Wiley Online Library.

Sees Jr, J. C. and Shortle, J. F. (2002). Simulating m/g/1 queues with heavy-tailed service. In *Simulation Conference, 2002. Proceedings of the Winter*, volume 1, pages 433–438. IEEE.

APPENDIX

Table 6: Means and relative errors. Simulating quantiles and expected shortfalls of $S_N = \sum_{i=1}^{N} X_i$ where $\mathbb{P}(X_i > x) = \exp(x^{-\beta})$, $\mathbb{P}(N = n) = \rho(1-\rho)^{(n-1)}$.

	$\beta = 0.8, \rho = 0.2$					
Weibull/Geometric	90%-quantile	95%-quantile	99%-quantile	99.9%-quantile	99.99%-quantile	99.999%-quantile
Estimated quantile	15.5716	20.1657	31.2012	46.3961	62.3213	75.3965
Relative error	0.01313	0.01381	0.01707	0.02813	0.06063	0.08026
	$\beta = 0.6, \rho = 0.2$					
Weibull/Geometric	90%-quantile	95%-quantile	99%-quantile	99.9%-quantile	99.99%-quantile	99.999%-quantile
Estimated quantile	21.6363	29.11683	46.9427	73.2428	99.7388	125.817
Relative error	0.009764	0.01169	0.01553	0.02094	0.01930	0.03121
	$\beta = 0.4, \rho = 0.2$					
Weibull/Geometric	90%-quantile	95%-quantile	99%-quantile	99.9%-quantile	99.99%-quantile	99.999%-quantile
Estimated quantile	49.2906	73.5994	140.271	264.780	437.100	670.432
Relative error	0.009739	0.009350	0.01088	0.005121	0.003460	0.003296
	$\beta = 0.2, \rho = 0.2$					
Weibull/Geometric	90%-quantile	95%-quantile	99%-quantile	99.9%-quantile	99.99%-quantile	99.999%-quantile
Estimated quantile	1022.07	2344.92	9999.33	46018.6	149669.9	390358.1
Relative error	0.01131	0.007263	0.004710	0.001813	0.001379	0.0006157
	$\beta = 0.8, \rho = 0.2$					
Weibull/Geometric	90%-quantile	95%-quantile	99%-quantile	99.9%-quantile	99.99%-quantile	99.999%-quantile
Estimated ES	21.7579	26.4005	37.2303	52.4836	67.0727	77.8810
Relative error	0.01005	0.01249	0.01726	0.02707	0.05173	0.07376
	$\beta = 0.6, \rho = 0.2$					
Weibull/Geometric	90%-quantile	95%-quantile	99%-quantile	99.9%-quantile	99.99%-quantile	99.999%-quantile
Estimated ES	31.9828	39.4908	57.4457	83.9030	110.503	136.550
Relative error	0.008739	0.008870	0.01721	0.02071	0.01772	0.02988
	$\beta = 0.4, \rho = 0.2$					
Weibull/Geometric	90%-quantile	95%-quantile	99%-quantile	99.9%-quantile	99.99%-quantile	99.999%-quantile
Estimated ES	88.0119	116.721	193.811	338.062	536.740	802.705
Relative error	0.008471	0.01070	0.01110	0.0060	0.004726	0.003834
	$\beta = 0.2, \rho = 0.2$					
Weibull/Geometric	90%-quantile	95%-quantile	99%-quantile	99.9%-quantile	99.99%-quantile	99.999%-quantile
Estimated ES	5193.00	8860.84	25693.4	90051.7	252252.9	595906.0
Relative error	0.03858	0.02782	0.01659	0.01266	0.007925	0.007637

Table 7: Means and relative errors. Simulating quantiles and expected shortfalls of $S_N = \sum_{i=1}^{N} X_i$ where $\mathbb{P}(X_i > x) = \int_0^x \frac{1}{z\sigma\sqrt{2\pi}} \exp(-\frac{(\ln z)^2}{2\sigma^2}) dz$, $\mathbb{P}(N = n) = \rho(1-\rho)^{(n-1)}$.

	$\sigma = 1.5, \rho = 0.2$					
Log-normal/Geometric	90%-quantile	95%-quantile	99%-quantile	99.9%-quantile	99.99%-quantile	99.999%-quantile
Estimated quantile	43.0787	61.9818	116.712	244.512	511.005	1043.09
Relative error	0.01081	0.01143	0.01144	0.003841	0.003958	0.001575
	$\sigma = 2.0, \rho = 0.2$					
Log-normal/Geometric	90%-quantile	95%-quantile	99%-quantile	99.9%-quantile	99.99%-quantile	99.999%-quantile
Estimated quantile	94.6375	153.475	389.479	1271.15	3775.13	10203.1
Relative error	0.01229	0.007735	0.007387	0.003748	0.001743	0.0007950
	$\sigma = 2.5, \rho = 0.2$					
Log-normal/Geometric	90%-quantile	95%-quantile	99%-quantile	99.9%-quantile	99.99%-quantile	99.999%-quantile
Estimated quantile	225.102	426.242	1495.87	7191.37	29060.8	101773.9
Relative error	0.01195	0.006451	0.005317	0.002870	0.001553	0.00064034
	$\sigma = 3.0, \rho = 0.2$					
Log-normal/Geometric	90%-quantile	95%-quantile	99%-quantile	99.9%-quantile	99.99%-quantile	99.999%-quantile
Estimated quantile	563.363	1250.45	6032.53	41582.1	225487.2	1019519.9
Relative error	0.008669	0.008767	0.005775	0.002417	0.0008147	0.0004598
	$\sigma = 1.5, \rho = 0.2$					
Log-normal/Geometric	90%-quantile	95%-quantile	99%-quantile	99.9%-quantile	99.99%-quantile	99.999%-quantile
Estimated ES	75.0995	99.0854	172.281	357.807	736.980	1469.82
Relative error	0.01496	0.01084	0.01375	0.009311	0.006701	0.007204
	$\sigma = 2.0, \rho = 0.2$					
Log-normal/Geometric	90%-quantile	95%-quantile	99%-quantile	99.9%-quantile	99.99%-quantile	99.999%-quantile
Estimated ES	234.143	341.323	780.272	2344.29	6563.31	16763.9
Relative error	0.07241	0.02705	0.01559	0.01516	0.009385	0.01131
	$\sigma = 2.5, \rho = 0.2$					
Log-normal/Geometric	90%-quantile	95%-quantile	99%-quantile	99.9%-quantile	99.99%-quantile	99.999%-quantile
Estimated ES	889.343	1428.11	4176.62	17149.0	62259.2	197743.9
Relative error	0.08343	0.05737	0.02405	0.01527	0.02784	0.02122
	$\sigma = 3.0, \rho = 0.2$					
Log-normal/Geometric	90%-quantile	95%-quantile	99%-quantile	99.9%-quantile	99.99%-quantile	99.999%-quantile
Estimated ES	3928.26	7097.71	25372.2	131697.1	611342.3	2391876.4
Relative error	0.1106	0.07361	0.1296	0.06921	0.043466	0.02005