Asymptotics and Simulation of Heavy-Tailed Processes

Henrik Hult

Department of Mathematics
KTH Royal Institute of Technology
Stockholm, Sweden

Workshop on Heavy-tailed Distributions and Extreme Value Theory
ISI Kolkata
January 14-17, 2013

Outline

Simulation of Heavy-Tailed Processes

- Goal: improve on the computational efficiency of standard Monte Carlo.
- Design: The large deviations analysis lead to a heuristic way to design efficient algorithms.
- Efficiency Analysis: The large deviations analysis can be applied to theoretically quantify the computational performance of algorithms.

Outline

The Performance of Monte Carlo

- Let $X^1, \ldots, X^N$ be independent copies of X.
- The Monte Carlo estimator is

$$\widehat{p} = \frac{1}{N} \sum_{k=1}^{N} I\{X^k > b\}.$$

- Its standard deviation is

$$\text{std} = \frac{1}{\sqrt{N}} \sqrt{p(1-p)}$$

- To obtain a Relative Error (std/p) of 1% you need to take N such that

$$\frac{\frac{1}{\sqrt{N}}\sqrt{p(1-p)}}{p} \leq 0.01 \quad \Leftrightarrow N \geq 10^4 \frac{1-p}{p}$$

Outline

Introduction to Importance Sampling[1]

- The problem with Monte Carlo is that few samples hit the rare event.
- This problem can be fixed by sampling from a distribution which puts more mass on the rare event.
- Must compensate for not sampling from the original distribution.
- Key issue: How to select an appropriate sampling distribution?

[1] See e.g. [1] for an introduction.

Monte Carlo and Importance Sampling

The Basics for Computing $F(A) = \mathbb{P}\{X \in A\}$

Monte Carlo

- Sample $X^1, \dots, X^N$ from F.
- Empirical measure

$$\mathbf{F}_N(\cdot) = \frac{1}{N} \sum_{k=1}^{N} \delta_{X^k}(\cdot).$$

- Plug-in estimator

$$\widehat{p} = \mathbf{F}_N(A).$$

Importance Sampling

- Sample $\widetilde{X}^1, \dots, \widetilde{X}^N$ from $\widetilde{F}$.
- Weighted empirical measure

$$\widetilde{\mathbf{F}}^w_N(\cdot) = \frac{1}{N} \sum_{k=1}^{N} \frac{dF}{d\widetilde{F}}(\widetilde{X}^k)\delta_{\widetilde{X}^k}(\cdot).$$

- Plug-in estimator

$$\widehat{p} = \widetilde{\mathbf{F}}^w_N(A) = \frac{1}{N} \sum_{k=1}^{N} \frac{dF}{d\widetilde{F}}(\widetilde{X}^k) I\{\widetilde{X}^k \in A\}.$$

Monte Carlo and Importance Sampling

The Basics for Computing $F(A) = \mathbb{P}\{X \in A\}$

Monte Carlo

- Sample $X^1, \dots, X^N$ from F.
- Empirical measure

$$\mathbf{F}_N(\cdot) = \frac{1}{N}\sum_{k=1}^{N} \delta_{X^k}(\cdot).$$

- Plug-in estimator

$$\widehat{p} = \mathbf{F}_N(A).$$

Importance Sampling

- Sample $\widetilde{X}^1, \dots, \widetilde{X}^N$ from $\widetilde{F}$.
- Weighted empirical measure

$$\widetilde{\mathbf{F}}_N^w(\cdot) = \frac{1}{N}\sum_{k=1}^{N} \frac{dF}{d\widetilde{F}}(\widetilde{X}^k)\delta_{\widetilde{X}^k}(\cdot).$$

- Plug-in estimator

$$\widehat{p} = \widetilde{\mathbf{F}}_N^w(A) = \frac{1}{N}\sum_{k=1}^{N} \frac{dF}{d\widetilde{F}}(\widetilde{X}^k)I\{\widetilde{X}^k \in A\}.$$

Importance Sampling

- The importance sampling estimator is unbiased:

$$\widetilde{E}\left[\frac{dF}{d\widetilde{F}}(\widetilde{X})I\{\widetilde{X} \in A\}\right] = \int_A \frac{dF}{d\widetilde{F}}d\widetilde{F} = \int_A dF = F(A).$$

- Its variance is

$$\begin{aligned}\mathrm{Var}\left(\frac{dF}{d\widetilde{F}}(\widetilde{X})I\{\widetilde{X} \in A\}\right) &= \widetilde{E}\left[\left(\frac{dF}{d\widetilde{F}}\right)^2 I\{\widetilde{X} \in A\}\right] - F(A)^2 \\ &= \int_A \left(\frac{dF}{d\widetilde{F}}\right)^2 d\widetilde{F} - F(A)^2 \\ &= \int_A \frac{dF}{d\widetilde{F}} dF - F(A)^2 \\ &= E\left[\frac{dF}{d\widetilde{F}} I\{X \in A\}\right] - F(A)^2.\end{aligned}$$

Importance Sampling

- The importance sampling estimator is unbiased:

$$\widetilde{E}\left[\frac{dF}{d\widetilde{F}}(\widetilde{X})I\{\widetilde{X}\in A\}\right]=\int_A \frac{dF}{d\widetilde{F}}d\widetilde{F}=\int_A dF=F(A).$$

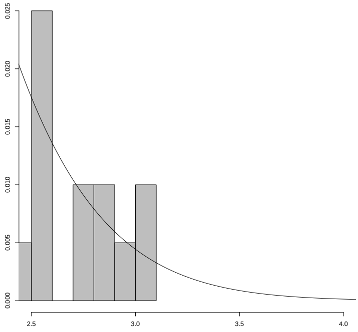
- Its variance is

$$\begin{aligned}\mathrm{Var}\Big(\frac{dF}{d\widetilde{F}}(\widetilde{X})I\{\widetilde{X}\in A\}\Big)&=\widetilde{E}\left[\Big(\frac{dF}{d\widetilde{F}}\Big)^2 I\{\widetilde{X}\in A\}\right]-F(A)^2\\&=\int_A\Big(\frac{dF}{d\widetilde{F}}\Big)^2 d\widetilde{F}-F(A)^2\\&=\int_A\frac{dF}{d\widetilde{F}}dF-F(A)^2\\&=E\left[\frac{dF}{d\widetilde{F}}I\{X\in A\}\right]-F(A)^2.\end{aligned}$$

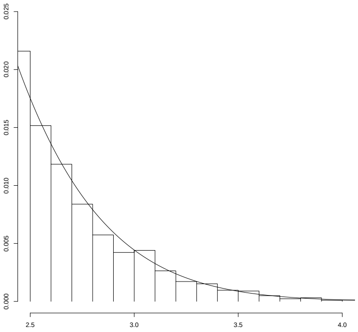
Illustration

Sampling the tail

Monte Carlo



Importance Sampling



Quantifying Efficiency

Variance Based Efficiency Criteria

- Basic idea: variance roughly the size of p^2.
- Embed in a sequence of problems such that $p_n = \mathbb{P}\{X_n \in A\} \to 0$.
- Logarithmic Efficiency: for some $\epsilon > 0$

$$\limsup_{n\to\infty} \frac{\mathrm{Var}(\widehat{p}_n)}{p_n^{2-\epsilon}} < \infty.$$

- Strong Efficiency/Bounded Relative Error:

$$\sup_n \frac{\mathrm{Var}(\widehat{p}_n)}{p_n^2} < \infty.$$

- Vanishing relative error:

$$\limsup_n \frac{\mathrm{Var}(\widehat{p}_n)}{p_n^2} = 0.$$

The Zero-Variance Change of Measure

- There is a best choice of sampling distribution, the zero-variance change of measure, given by

$$F_A(\cdot) = \mathbb{P}\{X_n \in \cdot \mid X_n \in A\}.$$

- For this choice

$$\frac{dF}{dF_A}(x) = \mathbb{P}\{X_n \in A\} I\{X_n \in A\} = p_n I\{X_n \in A\},$$

and hence

$$\mathrm{Var}(\widehat{p}_n) = E\left[p_n I\{X_n \in A\}\right] - F(A)^2 = 0.$$

Example Problem

A Random Walk with Heavy Tails

- Let $\{Z_k\}$ be iid non-negative and regularly varying (index α) with density f.
- Put $S_m = Z_1 + \cdots + Z_m$.
- Compute $\mathbb{P}\{S_m > n\}$.
- The zero-variance sampling distribution is

$$\mathbb{P}\{(Z_1, \ldots, Z_m) \in \cdot \mid S_m > n\}.$$

- By regular variation we know that

$$\mathbb{P}\{(Z_1, \ldots, Z_m) \in \cdot \mid S_m > n\} \to \mu(\cdot),$$

where μ is concentrated on the coordinate axis. But μ and F are singular!

Example Problem

A Random Walk with Heavy Tails

- Let $\{Z_k\}$ be iid non-negative and regularly varying (index α) with density f.
- Put $S_m = Z_1 + \cdots + Z_m$.
- Compute $\mathbb{P}\{S_m > n\}$.
- The zero-variance sampling distribution is

$$\mathbb{P}\{(Z_1, \dots, Z_m) \in \cdot \mid S_m > n\}.$$

- By regular variation we know that

$$\mathbb{P}\{(Z_1, \dots, Z_m) \in \cdot \mid S_m > n\} \to \mu(\cdot),$$

where μ is concentrated on the coordinate axis. But μ and F are singular!

Outline

The Conditional Mixture Algorithm[2]

- Suppose $S_{i-1} = s$. Sample Z_i as follows
  1. If $s \geq n$, sample Z_i from the original density f.
  2. If $s \leq n$, sample Z_i from the mixture

$$
\begin{aligned}
&p_i f(z) + (1 - p_i)\tilde{f}_i(y \mid s), \quad 1 \leq i \leq m - 1, \\
&\tilde{f}_m(y \mid s), \quad i = m.
\end{aligned}
$$

- Idea: take $\tilde{f}_i$ to produce large values of Z_i.

[2] This algorithm is due to [4]

The Conditional Mixture Algorithm

- We can take $\tilde{f}_i$'s to be the conditional distributions of the form

$$\tilde{f}_i(z \mid s) = \frac{f(z)I\{z > a(n-s)\}}{\mathbb{P}\{Z > a(n-s)\}}, \quad i \leq m-1,$$
$$\tilde{f}_m(z \mid s) = \frac{f(z)I\{z > n-s\}}{\mathbb{P}\{Z > n-s\}},$$

  where $a \in (0,1)$.
- Note: $\tilde{f}_i$ is the conditional distribution of Z given that Z is large.

Outline

The Performance of the Conditional Mixture Algorithm

- The normalized second moment is

$$\frac{E[\widehat{p}_n]}{p_n^2} = \frac{1}{p_n^2}\int \frac{dF}{d\widetilde{F}}(z_1,\dots,z_m)I\{s_m > n\}F(dz_1,\dots,dz_m)$$
$$\sim \frac{1}{m^2\mathbb{P}\{Z_1 > n\}^2}\int \frac{dF}{d\widetilde{F}}(z_1,\dots,z_m)I\{s_m > n\}F(dz_1,\dots,dz_m)$$
$$\sim \frac{1}{m^2}\int \frac{1}{\mathbb{P}\{Z_1 > n\}}\frac{dF}{d\widetilde{F}}(nz_1,\dots,nz_m)I\{s_m > 1\}\frac{F(ndz_1,\dots,ndz_m)}{\mathbb{P}\{Z_1 > n\}}.$$

The Performance of the Conditional Mixture Algorithm

- Weak convergence:

$$\frac{F(n\,\cdot)}{\mathbb{P}\{Z_1 > n\}} \to \sum_{k=1}^{m} \int I\{ze_k \in \cdot\}\mu_\alpha(dz)$$

- The normalized likelihood ratio is bounded:

$$\sup_n \frac{1}{\mathbb{P}\{Z_1 > n\}} \frac{dF}{d\widetilde{F}}(nz_1, \ldots, nz_m) I\{s_m > 1\} < \infty.$$

The Performance of the Conditional Mixture Algorithm

- The above enables us to show that the normalized second moment converges:

$$\lim_n \frac{E[\widehat{p}_n]}{p_n^2} = \frac{1}{m^2}\Big(\sum_{i=1}^{m-1} \frac{a^{-\alpha}}{1-p_i} \prod_{j=1}^{i-1} \frac{1}{p_j} + \prod_{j=1}^{m-1} \frac{1}{p_j}\Big).$$

- It is minimized at

$$p_i = \frac{(m-i-1)a^{-\alpha/2}+1}{(m-i)a^{-\alpha/2}+1},$$

with minimum

$$\frac{1}{m^2}\Big((m-1)a^{-\alpha/2}+1\Big)^2.$$

The Performance of the Conditional Mixture Algorithm

- The conditional mixture algorithm has (almost) vanishing relative error.
- Heavy-tailed heuristics indicate how to design the algorithm.
- Heavy-tailed asymptotics needed to prove efficiency of the algorithm.

MCMC for Rare Events[3]

- Let X be random variable with density f and compute $p = \mathbb{P}\{X \in A\}$, where A is a rare event.
- The zero-variance sampling distribution is

$$F_A(\cdot) = \mathbb{P}\{X \in \cdot \mid X \in A\}, \quad \frac{dF_A}{dx}(x) = \frac{f(x)I\{x \in A\}}{p}.$$

- It is possible to sample from F_A by constructing a Markov chain with stationary distribution F_A (e.g. Gibbs sampler or Metropolis-Hastings).
- Idea: sample from F_A and extract the normalizing constant p.

[3] This part is based on [5]

MCMC for Rare Events

- Let $X_0, \ldots, X_{T-1}$ be a sample of a Markov chain with stationary distribution F_A.
- Consider a non-negative function $v(x)$ with $\int_A v(x)dx = 1$. The sample mean

$$\widehat{q}_T = \frac{1}{T} \sum_{t=0}^{T-1} \frac{v(X_t) I\{X_t \in A\}}{f(X_t)},$$

is an estimator of

$$E_{F_A}\left[\frac{v(X) I\{X \in A\}}{f(X)}\right] = \int_A \frac{v(x)}{f(x)} \frac{f(x)}{p} dx = \frac{1}{p} \int_A v(x)dx = \frac{1}{p}.$$

- Take $\widehat{q}_T$ as the estimator of $1/p$.

MCMC for Rare Events

- The rare-event properties of $\widehat{q}_T$ are determined by the choice of v.
- The large sample properties of $\widehat{q}_T$ are determined by the ergodic properties of the Markov chain.

The Normalized Variance

- The rare-event efficiency (p small) is determined by the normalized variance:

$$
\begin{aligned}
p^2 \mathrm{Var}_{F_A}&\Big(\frac{v(X)}{f(X)} I\{X \in A\}\Big) \\
&= p^2 \Big(E_{F_A}\Big[\Big(\frac{v(X)}{f(X)} I\{X \in A\}\Big)^2\Big] - \frac{1}{p^2}\Big) \\
&= p^2 \Big(\int \frac{v^2(x)}{f^2(x)} \frac{f(x)}{p} dx - \frac{1}{p^2}\Big) \\
&= p \int_A \frac{v^2(x)}{f(x)} dx - 1.
\end{aligned}
$$

The Optimal Choice of v

- It is optimal to take v as $f(x)/p$. Indeed, then

$$p^2 \mathrm{Var}_{F_A}\left(\frac{v(X)}{f(X)} I\{X \in A\}\right) = p \int_A \frac{v^2(x)}{f(x)} dx - 1 = 0.$$

- Insight: take v as an approximation of the conditional density given the event.

Outline

Efficient Sampling for a Heavy-Tailed Random Walk

- Let $\{Z_k\}$ be iid with density f.
- Put $S_n = Z_1 + \cdots + Z_n$.
- Compute $\mathbb{P}\{S_n > a_n\}$.
- Heavy-tailed assumption:

$$\lim_{n\to\infty} \frac{\mathbb{P}\{S_n > a_n\}}{\mathbb{P}\{M_n > a_n\}} = 1.$$

(includes subexponential distributions, see [3]).

Gibbs Sampling of a Heavy-Tailed Random Walk

1. Start from a state with $Z_1 > a_n$.
2. Update the steps in a random order according to $j_1, \ldots, j_n$ (uniformly without replacement).
3. Update Z_{j_k} by sampling from

$$\mathbb{P}\{Z \in \cdot \mid Z + \sum_{i \neq j_k} Z_i > n\}.$$

- The vector $(Z_{t,1}, \ldots, Z_{t,n})'$ forms a uniformly ergodic Markov chain with stationary distribution

$$F_A(\cdot) = \mathbb{P}\{(Z_1, \ldots, Z_n)' \in \cdot \mid Z_1 + \ldots Z_n > a_n\}.$$

Rare-Event Efficiency

The Choice of v

- Take v to be the density of $\mathbb{P}\{(Z_1,\dots,Z_n)' \in \cdot \mid M_n > a_n\}$.
- Then

$$\frac{v(z_1,\dots,z_n)}{f(z_1,\dots,z_n)} = \frac{1}{\mathbb{P}\{M_n > a_n\}} I\{\vee_{i=1}^n z_i > a_n\}.$$

- Rare event efficiency:

$$\begin{aligned} p_n^2 \mathrm{Var}_{F_A}\Big(\frac{v(Z_1,\dots,Z_n)}{f(Z_1,\dots,Z_n)} I\{Z_1+\dots+Z_n > a_n\}\Big) &= \frac{\mathbb{P}\{S_n > a_n\}^2}{\mathbb{P}\{M_n > a_n\}^2}\mathbb{P}\{M_n > a_n \mid S_n > a_n\}\mathbb{P}\{M_n \le a_n \mid S_n > a_n\} \\ &= \frac{\mathbb{P}\{S_n > a_n\}}{\mathbb{P}\{M_n > a_n\}}\Big(1 - \frac{\mathbb{P}\{M_n > a_n\}}{\mathbb{P}\{S_n > a_n\}}\Big) \to 0. \end{aligned}$$

Rare-Event Efficiency

The Choice of v

- Take v to be the density of $\mathbb{P}\{(Z_1, \dots, Z_n)' \in \cdot \mid M_n > a_n\}$.
- Then

$$\frac{v(z_1, \dots, z_n)}{f(z_1, \dots, z_n)} = \frac{1}{\mathbb{P}\{M_n > a_n\}} I\{\vee_{i=1}^n z_i > a_n\}.$$

- Rare event efficiency:

$$\begin{aligned}
p_n^2 \mathrm{Var}_{F_A}&\Big(\frac{v(Z_1, \dots, Z_n)}{f(Z_1, \dots, Z_n)} I\{Z_1 + \dots + Z_n > a_n\}\Big) \\
&= \frac{\mathbb{P}\{S_n > a_n\}^2}{\mathbb{P}\{M_n > a_n\}^2} \mathbb{P}\{M_n > a_n \mid S_n > a_n\} \mathbb{P}\{M_n \le a_n \mid S_n > a_n\} \\
&= \frac{\mathbb{P}\{S_n > a_n\}}{\mathbb{P}\{M_n > a_n\}} \Big(1 - \frac{\mathbb{P}\{M_n > a_n\}}{\mathbb{P}\{S_n > a_n\}}\Big) \to 0.
\end{aligned}$$

Numerical illustration

$n = 5$, $\alpha = 2$, $a_n = an$.

$b = 25$, $T = 10^5$, $\alpha = 2$, $n = 5$, $a = 5$, $p_{\max} = 0.737\text{e-}2$			
	MCMC	IS	MC
Avg. est. Std. dev. Avg. time per batch(s)	1.050e-2 3e-5 12.8	1.048e-2 9e-5 12.7	1.053e-2 27e-5 1.4
$b = 25$, $T = 10^5$, $\alpha = 2$, $n = 5$, $a = 20$, $p_{\max} = 4.901\text{e-}4$			
	MCMC	IS	MC
Avg. est. Std. dev. Avg. time per batch(s)	5.340e-4 6e-7 14.4	5.343e-4 13e-7 13.9	5.380e-4 770e-7 1.5
$b = 20$, $T = 10^5$, $\alpha = 2$, $n = 5$, $a = 10^3$, $p_{\max} = 1.9992\text{e-}7$			
	MCMC	IS	
Avg. est. Std. dev. Avg. time per batch(s)	2.0024e-7 3e-11 15.9	2.0027e-7 20e-11 15.9	
$b = 20$, $T = 10^5$, $\alpha = 2$, $n = 5$, $a = 10^4$, $p_{\max} = 1.99992\text{e-}9$			
	MCMC	IS	
Avg. est. Std. dev. Avg. time per batch(s)	2.00025e-9 7e-14 15.9	2.00091e-9 215e-14 15.9	

Numerical Illustration

$n = 20$, $\alpha = 2$, $a_n = an$.

$b = 25$, $T = 10^5$, $\alpha = 2$, $n = 20$, $a = 20$, $p_{\max} = 1.2437\text{e-}4$			
	MCMC	IS	MC
Avg. est.	1.375e-4	1.374e-4	1.444e-4
Std. dev.	2e-7	3e-7	492e-7
Avg. time per batch(s)	52.8	50.0	2.0
$b = 25$, $T = 10^5$, $\alpha = 2$, $n = 20$, $a = 200$, $p_{\max} = 1.2494\text{e-}6$			
	MCMC	IS	MC
Avg. est.	1.2614e-6	1.2615e-6	1.2000e-6
Std. dev.	4e-10	12e-10	33,166e-10
Avg. time per batch(s)	49.4	48.4	1.9
$b = 20$, $T = 10^5$, $\alpha = 2$, $n = 20$, $a = 10^3$, $p_{\max} = 4.9995\text{e-}8$			
	MCMC	IS	
Avg. est.	5.0091e-8	5.0079e-8	
Std. dev.	7e-12	66e-12	
Avg. time per batch(s)	53.0	50.6	
$b = 20$, $T = 10^5$, $\alpha = 2$, $n = 20$, $a = 10^4$, $p_{\max} = 5.0000\text{e-}10$			
	MCMC	IS	
Avg. est.	5.0010e-10	5.0006e-10	
Std. dev.	2e-14	71e-14	
Avg. time per batch(s)	48.0	47.1	

Outline

Efficient Sampling for a Heavy-Tailed Random Sum

- Let $\{Z_k\}$ be iid with density f.
- Put $S_n = Z_1 + \cdots + Z_n$ and let $\{N_n\}$ be independent of $\{Z_k\}$
- Compute $\mathbb{P}\{S_{N_n} > a_n\}$.
- Heavy-tailed assumption:[4]

$$\lim_{n\to\infty} \frac{\mathbb{P}\{S_{N_n} > a_n\}}{\mathbb{P}\{M_{N_n} > a_n\}} = 1.$$

[4] See e.g. [6] for examples.

Constructing the Gibbs Sampler

- Suppose we are in state $(N_t, Z_{t,1}, \dots, Z_{t,N_t}) = (k_t, z_{t,1}, \dots, z_{t,k_t})$.
- Let $k_t^* = \min\{j : z_{t,1} + \cdots + z_{t,j} > a_n\}$.
- Update the number of steps N_{t+1} from the distribution

$$p(k_{t+1} \mid k_t^*) = \mathbb{P}\{N = k_{t+1} \mid N \geq k^{*t}\}$$

- If $k_{t+1} > k_t$, sample $Z_{t+1,k_t+1}, \dots, Z_{t+1,k_{t+1}}$ independently from F_Z.
- Proceed by updating all steps as before.
- Permute the steps.

Properties of the algorithm

- The Markov chain $\{(N_t, Z_{t,1}, \ldots, Z_{t,N_t})\}$ is uniformly ergodic with stationary distribution

$$F_A(\cdot) = \mathbb{P}\{(Z_1, \ldots, Z_n)' \in \cdot \mid Z_1 + \ldots Z_n > a_n\}.$$

- With v as the density of $\mathbb{P}\{(N, Z_1, \ldots, Z_N) \in \cdot \mid M_N > a_n\}$ we have

$$\begin{aligned} \frac{v(k, z_1, \ldots, z_k)}{f(k, z_1, \ldots, z_k)} &= \frac{1}{\mathbb{P}\{M_N > a_n\}} I\{\max z_1, \ldots, z_k > a_n\} \\ &= \frac{1}{1 - g_N(F_Z(a_n))} I\{\max z_1, \ldots, z_k > a_n\}. \end{aligned}$$

- The algorithm has vanishing normalized variance.

Numerical illustration

Geometric(ρ) number of steps: $\rho = 0.05$, $\alpha = 1$, $a_n = a/\rho$.

$b = 25$, $T = 10^5$, $\alpha = 1$, $\rho = 0.2$, $a = 10^2$, $p_{\max} = 0.990\text{e-}2$			
	MCMC	IS	MC
Avg. est. Std. dev. Avg. time per batch(s)	1.149e-2 4e-5 25.0	1.087e-2 6e-5 11.0	1.089e-2 35e-5 1.2
$b = 25$, $T = 10^5$, $\alpha = 1$, $\rho = 0.2$, $a = 10^3$, $p_{\max} = 0.999\text{e-}3$			
	MCMC	IS	MC
Avg. est. Std. dev. Avg. time per batch(s)	1.019e-3 1e-6 25.8	1.012e-3 3e-6 11.1	1.037e-3 76e-6 1.2
$b = 20$, $T = 10^6$, $\alpha = 1$, $\rho = 0.2$, $a = 5 \cdot 10^7$, $p_{\max} = 2.000000\text{e-}8$			
	MCMC	IS	
Avg. est. Std. dev. Avg. time per batch(s)	2.000003e-8 6e-14 385.3	1.999325e-8 1114e-14 139.9	
$b = 20$, $T = 10^6$, $\alpha = 1$, $\rho = 0.2$, $a = 5 \cdot 10^9$, $p_{\max} = 2.0000\text{e-}10$			
	MCMC	IS	
Avg. est. Std. dev. Avg. time per batch(s)	2.0000e-10 0 358.7	1.9998e-10 13e-14 130.9	

Numerical Illustration

Geometric(ρ) number of steps: $\rho = 0.05$, $\alpha = 1$, $a_n = a/\rho$.

$b = 25$, $T = 10^5$, $\alpha = 1$, $\rho = 0.05$, $a = 10^3$, $p_{\max} = 0.999\text{e-}3$			
	MCMC	IS	MC
Avg. est.	1.027e-3	1.017e-3	1.045e-3
Std. dev.	1e-6	4e-6	105e-6
Avg. time per batch(s)	61.5	44.8	1.3
$b = 25$, $T = 10^5$, $\alpha = 1$, $\rho = 0.05$, $a = 5 \cdot 10^5$, $p_{\max} = 1.9999\text{e-}6$			
	MCMC	IS	MC
Avg. est.	2.0002e-6	2.0005e-6	3.2000e-6
Std. dev.	1e-10	53e-10	55,678e-10
Avg. time per batch(s)	60.7	45.0	1.3

For Further Reading I

S. Asmussen and P.W. Glynn
Stochastic Simulation: Algorithms and Anal- ysis.
Springer-Verlag, New York, 2007.

S. Resnick
Heavy-Tail Phenomena: Probabilistic and Statistical Modeling.
Springer, New York, 2006.

D. Denisov, A.B. Dieker, and V. Shneer
Large deviations for random walks under subexponentiality: the big-jump domain.
Ann. Probab. 36(5), 1946-1991, 2008.

For Further Reading II

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

T. Gudmundsson and H. Hult
Markov chain Monte Carlo for computing rare-event probabilities for a heavy-tailed random walk.
http://arxiv.org/pdf/1211.2207v1.pdf

C. Klüppelberg and T. Mikosch
Large deviations for heavy-tailed random sums with applications to insurance and finance.
J. Appl. Probab. 37(2), 293-308, 1997.