Markov Chain Monte Carlo for rare-event simulation in heavy-tailed settings

Thorbjörn Gudmundsson
Department of Mathematics
KTH Stockholm

Licentiat seminarium, December 2013
Computational problem. For instance, the probability of default or the expected loss given ruin.
Computational problem. For instance, the probability of default or the expected loss given ruin.

Complex system: no analytical solution available
Computational problem. For instance, the probability of default or the expected loss given ruin.

Complex system: no analytical solution available

Simulation techniques

- Monte Carlo
- Conditional Monte Carlo
- Splitting methods
- Importance sampling
Computational problem. For instance, the probability of default or the expected loss given ruin.

Complex system: no analytical solution available

Simulation techniques
  i Monte Carlo
  ii Conditional Monte Carlo
  iii Splitting methods
  iv Importance sampling
  v Markov chain Monte Carlo (NEW)
Consider a random variable $X$ with known distribution $F$ and the objective of computing

$$p = \mathbb{P}(X \in A),$$

where $\{X \in A\}$ is thought as rare in the sense that $p$ is small. Event of ruin for instance.
Problem

Consider a random variable $X$ with known distribution $F$ and the objective of computing

$$p = \mathbb{P}(X \in A),$$

where $\{X \in A\}$ is thought as rare in the sense that $p$ is small. Event of ruin for instance.

**Example.** Random walk $S_n = Y_1 + \cdots + Y_n$ with non-negative steps $Y$’s with known heavy-tailed distribution $F_Y$ and objective of computing

$$p = \mathbb{P}\left(\frac{S_n}{n} > a\right),$$

where $a$ is much larger than $\mathbb{E}[Y]$. 
Stochastic Simulation

Want to compute $p = \mathbb{P}(X \in A)$. In absence of an analytical solution, stochastic simulation offers an alternative.
Want to compute $p = P(X \in A)$. In absence of an analytical solution, stochastic simulation offers an alternative. Monte Carlo: sample identically distributed and independent copies $X_1, \ldots, X_N$ and compute

$$\hat{p} = \frac{1}{N} \sum_{k=1}^{N} I\{X_k \in A\}.$$
Shortcomings of Monte Carlo

The relative error of the Monte Carlo estimator is unbounded as $p \to 0$:

$$\frac{\text{Var}(\hat{p})}{p^2} = \frac{1}{N \left( \frac{1}{p} - 1 \right)} \to \infty, \quad \text{as } p \to 0.$$ 

**Example.** Standard normal variable $X$, compute $p = \mathbb{P}(X > a)$ using $N = 10^6$ number of simulations

- $a = 1 : \hat{p} = 0.158, \quad \frac{\text{Stdev}(\hat{p})}{\hat{p}} = 0.002$
- $a = 3 : \hat{p} = 0.0014, \quad \frac{\text{Stdev}(\hat{p})}{\hat{p}} = 0.027$
- $a = 5 : \hat{p} = 0, \quad \frac{\text{Stdev}(\hat{p})}{\hat{p}} = \infty$
Solutions

- Conditional Monte Carlo (Asmussen)
- Splitting methods (Creou et al)
- Importance sampling (Sigmund, Dupuis, Blanchet)
## Importance sampling

Goal: construct an efficient estimator \( \hat{p} \) of \( p = \mathbb{P}(X \in A) \), in the sense that its relative error is bounded.

\[
\hat{p} = \frac{1}{N} \sum_{k=1}^{N} dF/dG(X_k) I\{X_k \in A\}
\]

\[
E_G[\hat{p}] = \int_A dF/dG(X) dG(X) = F(A) = p.
\]
## Importance sampling

Goal: construct an efficient estimator $\hat{p}$ of $p = P(X \in A)$, in the sense that its relative error is bounded.

The importance sampling approach (Dupuis et al 2007)

- Generate independent copies $X_1, \ldots, X_N$ from a sampling distribution $G$.
- Compute empirical estimate

$$\hat{p} = \frac{1}{N} \sum_{k=1}^{N} \frac{dF}{dG}(X_k) \mathbb{I}\{X_k \in A\}.$$
Importance sampling

Goal: construct an efficient estimator \( \hat{p} \) of \( p = \mathbb{P}(X \in A) \), in the sense that its relative error is bounded.

The importance sampling approach (Dupuis et al 2007)

- Generate independent copies \( X_1, \ldots, X_N \) from a sampling distribution \( G \).
- Compute empirical estimate

\[
\hat{p} = \frac{1}{N} \sum_{k=1}^{N} \frac{dF}{dG}(X_k) \mathbb{I}\{X_k \in A\}.
\]

\[
\mathbb{E}_G[\hat{p}] = \int_A \frac{dF}{dG}(X) dG(X) = F(A) = p.
\]
Importance sampling continued

Reduces to finding a suitable sampling distribution $G$. 

\[ \hat{p} = \frac{1}{N} \sum_{k=1}^{N} \frac{dF}{dF_A}(X_k) I\{X_k \in A\} = p, \] 

with zero variance!

Requires knowledge of $P(X \in A)$. ...
Importance sampling continued

Reduces to finding a suitable sampling distribution \( G \). The zero-variance distribution

\[
F_A(x) = \mathbb{P}(X \leq x|X \in A).
\]

If we can choose \( G = F_A \), then \( \frac{dF}{dF_A}(X) \mathbb{I}\{X \in A\} = p \), so

\[
\hat{p} = \frac{1}{N} \sum_{k=1}^{N} \frac{dF}{dF_A}(X_k) \mathbb{I}\{X_k \in A\} = p,
\]

with zero variance!
Importance sampling continued

Reduces to finding a suitable sampling distribution $G$. The zero-variance distribution

$$F_A(x) = \mathbb{P}(X \leq x | X \in A).$$

If we can choose $G = F_A$, then $\frac{dF}{dF_A}(X)\mathbb{1}\{X \in A\} = p$, so

$$\hat{p} = \frac{1}{N} \sum_{k=1}^{N} \frac{dF}{dF_A}(X_k)\mathbb{1}\{X_k \in A\} = p,$$

with zero variance!

Requires knowledge of $\mathbb{P}(X \in A)$ ...
The idea

Want: sample from $F_A(x) = \mathbb{P}(X \leq x | X \in A)$.
Assuming the existence of a density, it takes the form

$$f_A(x) = \frac{f(x) \mathbb{1}\{x \in A\}}{\mathbb{P}(X \in A)}.$$
The idea

Want: sample from $F_A(x) = \mathbb{P}(X \leq x | X \in A)$.
Assuming the existence of a density, it takes the form

$$f_A(x) = \frac{f(x) \mathbb{I}\{x \in A\}}{\mathbb{P}(X \in A)}.$$

The main idea is to construct a Markov chain $(X_k)_{k \geq 1}$ for which $f_A$ is the invariant density via MCMC. Then extract information about the normalising constant from the sample.
Construct a Markov chain $(X_k)_{k \geq 1}$ via MCMC sampler, with the zero-variance distribution $F_A$ as its invariant distribution.
Construct a Markov chain \((X_k)_{k \geq 1}\) via MCMC sampler, with the zero-variance distribution \(F_A\) as its invariant distribution.

For any \(v \geq 0\) such that \(\int_A v(x) \, dx = 1\), consider

\[
u\left((X_k)_{k \geq 1}\right) = \frac{1}{N} \sum_{k=1}^N \frac{v(X_k) \mathbb{I}\{X_k \in A\}}{f(X_k)}.
\]
Estimator continued

For $\int_A v(x)dx = 1$ it holds

$$\mathbb{E}_{F_A} \left[ \frac{1}{N} \sum_{k=1}^{N} \frac{v(X_k)\mathbb{I}\{X_k \in A\}}{f(X_k)} \right] = \int_A \frac{v(x)f(x)}{p} dx$$

$$= \frac{1}{p} \int_A v(x) dx$$

$$= \frac{1}{p}.$$
Estimator continued

- For $\int_A v(x)dx = 1$ it holds

$$E_{F_A}\left[\frac{1}{N} \sum_{k=1}^{N} \frac{v(X_k)I\{X_k \in A\}}{f(X_k)}\right] = \int_A \frac{v(x) f(x)}{p} dx$$

$$= \frac{1}{p} \int_A v(x) dx$$

$$= \frac{1}{p}.$$

- Define $\hat{q} = \frac{1}{N} \sum_{k=1}^{N} \frac{v(X_k)I\{X_k \in A\}}{f(X_k)}$ estimator of $1/p$. 
Design issues

Estimator $\hat{q} = \frac{1}{N} \sum_{k=1}^{N} \frac{v(X_k)I\{X_k \in A\}}{f(X_k)}$ of $1/p$.

- Choice of the MCMC sampler: crucial to control the dependence of the Markov chain, to ensure the large sample efficiency

$$\nabla \text{var}(\hat{q}) \to 0, \quad \text{as } N \to \infty.$$
Design issues

Estimator \( \hat{q} = \frac{1}{N} \sum_{k=1}^{N} \frac{v(X_k)I\{X_k \in A\}}{f(X_k)} \) of \( 1/p \).

- Choice of the MCMC sampler: crucial to control the dependence of the Markov chain, to ensure the large sample efficiency

\[ \nabla \text{ar}(\hat{q}) \to 0, \quad \text{as } N \to \infty. \]

- Choice of \( v \): controls the variance, set to ensure rare-event efficiency

\[ \frac{\text{Std}(\hat{q})}{1/p} = p \text{Std}(\hat{q}) \to 0, \quad \text{as } p \to 0. \]
Controlling the variance

Estimator \( \hat{q} = \frac{1}{N} \sum_{k=1}^{N} u(X_k) \), with \( u(X_k) = \frac{v(X_k) I\{X_k \in A\}}{f(X_k)} \).

Goal is to show \( p \text{ Std}(\hat{q}) \) tends to zero as \( p \to 0 \).
Controlling the variance

Estimator \( \hat{q} = \frac{1}{N} \sum_{k=1}^{N} u(X_k) \), with \( u(X_k) = \frac{v(X_k)I\{X_k \in A\}}{f(X_k)} \).

Goal is to show \( p \text{Std}(\hat{q}) \) tends to zero as \( p \to 0 \).

- Consider the term

\[
p^2 \text{Var}(u(X)) = p^2 (\mathbb{E}[u(X)^2] - \mathbb{E}[u(X)]^2) = p^2 \left( \int_{A} \frac{v^2(x) f(x)}{p} \, dx - 1 \right) = p \int_{A} \frac{v^2(x)}{f(x)} \, dx - 1.
\]
Choosing \( v(x) = f_A(x) = \frac{f(x) \mathbb{I}\{x \in A\}}{p} \) implies

\[
p^2 \text{Var}(u(X)) = p \int_A \frac{f^2(x)/p^2}{f(x)} \, dx - 1 = \frac{1}{p} \int_A f(x) \, dx - 1 = 0.
\]
Choosing $\nu(x) = f_A(x) = \frac{f(x)I\{x \in A\}}{p}$ implies

$$p^2 \text{Var}(u(X)) = p \int_A \frac{f^2(x)/p^2}{f(x)} dx - 1 = \frac{1}{p} \int_A f(x) dx - 1 = 0.$$ 

Choose $\nu$ as an approximation of the zero-variance density!
Recipe

- Sample $(X_k)_{k \geq 1}$ under $F_A$ via some MCMC sampler
Recipe

- Sample \((X_k)_{k \geq 1}\) under \(F_A\) via some MCMC sampler
- Show \(p^2 \text{Var}(u(X)) \rightarrow 0\) as \(p \rightarrow 0\)
### Recipe

- Sample \((X_k)_{k \geq 1}\) under \(F_A\) via some MCMC sampler
- Show \(p^2 \text{Var}(u(X)) \to 0\) as \(p \to 0\)
- Show \((X_k)_{k \geq 1}\) is geometric ergodic
Consider a random walk \( S_n = Y_1 + \cdots + Y_n \) with non-negative steps \( Y \)'s with known heavy-tailed distribution \( F_Y \) and objective of computing
\[
\mathbb{P}\left( \frac{S_n}{n} > a \right),
\]
where \( a \) is much larger than \( \mathbb{E}[Y] \).
Consider a random walk $S_n = Y_1 + \cdots + Y_n$ with non-negative steps $Y$'s with known heavy-tailed distribution $F_Y$ and objective of computing

$$p = \mathbb{P}\left(\frac{S_n}{n} > a\right),$$

where $a$ is much larger than $\mathbb{E}[Y]$.

Construct $(Y_k)_{k \geq 1}$ via MCMC with invariant density

$$f_A(y) = \frac{f_Y(y) \mathbb{I}\{y_1 + \cdots + y_n > an\}}{\mathbb{P}(S_n > an)}.$$
Consider a random walk \( S_n = Y_1 + \cdots + Y_n \) with non-negative steps \( Y \)'s with known heavy-tailed distribution \( F_Y \) and objective of computing 

\[
p = \mathbb{P} \left( \frac{S_n}{n} > a \right),
\]

where \( a \) is much larger than \( \mathbb{E}[Y] \).

Construct \((Y_k)_{k \geq 1}\) via MCMC with invariant density 

\[
f_A(y) = \frac{f_Y(y) \mathbb{I}\{y_1 + \cdots + y_n > an\}}{\mathbb{P}(S_n > an)}.
\]

A typical such a random walk has a \( n - 1 \) number of "small" steps and one "large" step.
Gibbs sampler

Initial state $Y_0 = (Y_{0,1}, \ldots, Y_{0,n})$ such that $Y_{0,1} > an$ and $Y_{0,i} = 0$ for other indices. Given $Y_k = (Y_{k,1}, \ldots, Y_{k,n})$, $k = 0, 1, \ldots$ the next state $Y_{k+1}$ is sampled as follows:

- Take a copy of the current state, let $Y_{k+1,i} = Y_{k,i}$,
Gibbs sampler

Initial state $Y_0 = (Y_{0,1}, \ldots, Y_{0,n})$ such that $Y_{0,1} > an$ and $Y_{0,i} = 0$ for other indices. Given $Y_k = (Y_{k,1}, \ldots, Y_{k,n})$, $k = 0, 1, \ldots$ the next state $Y_{k+1}$ is sampled as follows

- Take a copy of the current state, let $Y_{k+1,i} = Y_{k,i}$,
- Draw a random index $j \in \{1, \ldots, n\}$,
Gibbs sampler

Initial state $Y_0 = (Y_{0,1}, \ldots, Y_{0,n})$ such that $Y_{0,1} > an$ and $Y_{0,i} = 0$ for other indices. Given $Y_k = (Y_{k,1}, \ldots, Y_{k,n})$, $k = 0, 1, \ldots$ the next state $Y_{k+1}$ is sampled as follows:

- Take a copy of the current state, let $Y_{k+1,i} = Y_{k,i}$,
- Draw a random index $j \in \{1, \ldots, n\}$,
- Sample $Y_{k+1,j}$ from the conditional distribution of $Y$ given that the sum exceeds the threshold,

$$
P(Y_{k+1,j} \in \cdot) = P(Y \in \cdot \mid Y + \sum_{i \neq j} Y_{k,i} > an).$$
Gibbs sampler

Initial state $Y_0 = (Y_{0,1}, \ldots, Y_{0,n})$ such that $Y_{0,1} > an$ and $Y_{0,i} = 0$ for other indices. Given $Y_k = (Y_{k,1}, \ldots, Y_{k,n})$, $k = 0, 1, \ldots$ the next state $Y_{k+1}$ is sampled as follows:

- Take a copy of the current state, let $Y_{k+1,i} = Y_{k,i}$,
- Draw a random index $j \in \{1, \ldots, n\}$,
- Sample $Y_{k+1,j}$ from the conditional distribution of $Y$ given that the sum exceeds the threshold,

$$
P(Y_{k+1,j} \in \cdot) = P(Y \in \cdot \mid Y + \sum_{i \neq j} Y_{k,i} > an).$$

- Permute the steps in $Y_{k+1}$. 
Gibbs sampler continued

**Proposition**

*The Markov chain \((Y_k)_{k \geq 1}\) constructed using the proposed Gibbs sampler has the conditional distribution \(F_A\) as its invariant distribution.*
The MCMC estimator \( \hat{q} = \frac{1}{N} \sum_{k=1}^{N} \frac{v(y_k) I\{S_n > an\}}{f(y_k)} \). The steps are heavy-tailed in the sense that

\[
\frac{\mathbb{P}(M_n > an)}{\mathbb{P}(S_n > an)} \to 1,
\]

where \( M_n = \max_i \{y_{k,i}\} \).
MCMC estimator

- The MCMC estimator \( \hat{q} = \frac{1}{N} \sum_{k=1}^{N} \frac{\nu(y_k) I\{S_n > an\}}{f(y_k)} \). The steps are heavy-tailed in the sense that

\[
\frac{\mathbb{P}(M_n > an)}{\mathbb{P}(S_n > an)} \to 1,
\]

where \( M_n = \max_i \{y_{k,i}\} \).

- Therefore seems smart to use

\[
\mathbb{P}(Y \in \cdot \mid M_n > an) \text{ as a proxy for } \mathbb{P}(Y \in \cdot \mid S_n > an).
\]

Propose

\[
\nu(y_k) = \frac{f(y_k) I\{M_n > an\}}{\mathbb{P}(M_n > an)}.
\]
Choosing $v(y) = \frac{f(y) \mathbb{I}\{M_n > an\}}{\mathbb{P}(M_n > an)}$ yields $u(y) = \frac{v(y) \mathbb{I}\{S_n > an\}}{f(y)} = \frac{\mathbb{I}\{M_n > an\}}{\mathbb{P}(M_n > an)}$. 
Choosing \( v(y) = \frac{f(y)I\{M_n > an\}}{P(M_n > an)} \) yields

\[
u(y) = \frac{v(y)I\{S_n > an\}}{f(y)} = \frac{I\{M_n > an\}}{P(M_n > an)}.
\]

\[
\hat{q} = P(M_n > an)^{-1} \frac{1}{N} \sum_{k=1}^{N} I\{M_n(k) > an\}
\]
Since \( u(y) = \frac{\mathbb{I}\{M_n > an\}}{\mathbb{P}(M_n > an)} \), we have:

\[
\begin{align*}
p^2 \text{Var}_{F_A}(u(Y)) &= \frac{\mathbb{P}(S_n > an)^2}{\mathbb{P}(M_n > an)^2} \text{Var}_{F_A}(\mathbb{I}\{M_n > an\})
\end{align*}
\]
Efficiency

Since \( u(y) = \frac{\mathbb{I}\{M_n > an\}}{\mathbb{P}(M_n > an)} \), we have:

\[
p^2 \text{Var}_{F_{A}}(u(Y)) = \frac{\mathbb{P}(S_n > an)^2}{\mathbb{P}(M_n > an)^2} \text{Var}_{F_{A}}(\mathbb{I}\{M_n > an\})
\]

\[
= \frac{\mathbb{P}(S_n > an)^2}{\mathbb{P}(M_n > an)^2} \left( \mathbb{E}_{F_{A}}[\mathbb{I}\{M_n > an\}] - \mathbb{E}_{F_{A}}[\mathbb{I}\{M_n > an\}]^2 \right)
\]
Since \( u(y) = \frac{\mathbb{I}\{M_n > an\}}{\mathbb{P}(M_n > an)} \), we have:

\[
p^2 \mathbb{V} \text{ar}_{F_A}(u(Y)) = \frac{\mathbb{P}(S_n > an)^2}{\mathbb{P}(M_n > an)^2} \mathbb{V} \text{ar}_{F_A}(\mathbb{I}\{M_n > an\})
\]

\[
= \frac{\mathbb{P}(S_n > an)^2}{\mathbb{P}(M_n > an)^2} \left( \mathbb{E}_{F_A}[\mathbb{I}\{M_n > an\}] - \mathbb{E}_{F_A}[\mathbb{I}\{M_n > an\}]^2 \right)
\]

\[
= \frac{\mathbb{P}(S_n > an)^2}{\mathbb{P}(M_n > an)^2} \left( \frac{\mathbb{P}(M_n > an)}{\mathbb{P}(S_n > an)} - \frac{\mathbb{P}(M_n > an)^2}{\mathbb{P}(S_n > an)^2} \right)
\]
Since \( u(y) = \frac{\mathbb{I}\{M_n > an\}}{\mathbb{P}(M_n > an)} \), we have:

\[
p^2 \text{Var}_{F_A}(u(Y)) = \frac{\mathbb{P}(S_n > an)^2}{\mathbb{P}(M_n > an)^2} \text{Var}_{F_A}(\mathbb{I}\{M_n > an\})
\]

\[
= \frac{\mathbb{P}(S_n > an)^2}{\mathbb{P}(M_n > an)^2} \left( \mathbb{E}_{F_A}[\mathbb{I}\{M_n > an\}] - \mathbb{E}_{F_A}[\mathbb{I}\{M_n > an\}]^2 \right)
\]

\[
= \frac{\mathbb{P}(S_n > an)^2}{\mathbb{P}(M_n > an)^2} \left( \frac{\mathbb{P}(M_n > an)}{\mathbb{P}(S_n > an)} - \frac{\mathbb{P}(M_n > an)^2}{\mathbb{P}(S_n > an)^2} \right)
\]

\[
= \frac{\mathbb{P}(S_n > an)}{\mathbb{P}(M_n > an)} - 1 \to 0 \quad \text{as } p \to 0.
\]
Geometric ergodicity

- The design of the Gibbs sampler ensures that the Markov chain $(Y_k)_{k \geq 1}$ is (uniformly) ergodic.
- This guarantees that the chain mixes sufficiently and hence that $\nabla \text{var}(\hat{\rho}) \to 0$ as $N \to \infty$ at same speed as $1/N$. 
Numerical experiments

- The MCMC estimator $\hat{q}^{-1}$ of the probability $p$ tested against importance sampling and standard Monte Carlo.
- Steps are Pareto(2) distributed.
- Number of batches: 25, simulations per batch: 10,000.
<table>
<thead>
<tr>
<th>$n$</th>
<th>$a$</th>
<th>MCMC</th>
<th>IS</th>
<th>MC</th>
<th>Avg. est. (Std. dev.)</th>
<th>Avg. time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>10</td>
<td>3.40e-3</td>
<td>2.91e-3</td>
<td>2.83e-3</td>
<td>Avg. est.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.81e-4)</td>
<td>(1.77e-4)</td>
<td>(4.74e-4)</td>
<td>(0.7)</td>
<td>Avg. time (ms)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[4.1]</td>
<td>[3.4]</td>
<td>[0.7]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>20</td>
<td>3.34e-4</td>
<td>3.02e-4</td>
<td>2.68e-4</td>
<td>Avg. est.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(5.83e-6)</td>
<td>(2.02e-6)</td>
<td>(162.58e-6)</td>
<td>(Std. dev.)</td>
<td></td>
</tr>
</tbody>
</table>
10,000 simulations for $m = 10$ and $a = 20$
Consider a random walk $S_{N_n} = Y_1 + \cdots + Y_{N_n}$ with non-negative heavy-tailed steps $Y$, discrete random variable $N_n$ and the objective of computing

$$p = \mathbb{P}(S_{N_n} > a\mathbb{E}[N_n]),$$

where $a$ is much larger than $\mathbb{E}[Y]$. 

Setup
The challenge

How to design a Gibbs sampler to construct a Markov chain with the following invariant distribution

$$F_A(\cdot) = \mathbb{P}\left( (N, Y_1, \ldots, Y_N) \in \cdot \mid S_{N_n} > a_n \right).$$
The challenge

How to design a Gibbs sampler to construct a Markov chain with the following invariant distribution

\[ F_A(\cdot) = \mathbb{P}((N, Y_1, \ldots, Y_N) \in \cdot \mid S_{N_n} > a_n). \]

The trick was to sample \( N \) from \( \mathbb{P}(N = k \mid N \geq k^*) \) where
\[ k^* = \min\{k : Y_1 + \ldots + Y_k > a_n\}. \]
Numerical experiments

- The MCMC estimator $\hat{q}^{-1}$ of the probability $p$ tested against importance sampling and standard Monte Carlo.
- Steps are Pareto(1) distributed.
- Number of steps is Geometric(0.2) distributed
- Number of batches: 25, simulations per batch: 10,000.
## Numerical experiments

<table>
<thead>
<tr>
<th>$a$</th>
<th>MCMC</th>
<th>IS</th>
<th>MC</th>
<th>Avg. est. (Std. dev.) [Avg. time (ms)]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$5 \cdot 10^7$</td>
<td>2.000003e-8 (6e-14)</td>
<td>1.999325e-8 (1114e-14)</td>
<td></td>
<td>Avg. est. (Std. dev.)</td>
</tr>
</tbody>
</table>
Setup

Consider the following setup for the risk reserve $U_k$, for positive claim size $B$:

$$U_k = R_k(U_{k-1} - B_k), \quad \text{for } k \geq 1,$$

$$U_0 = u.$$
Setup

Consider the following setup for the risk reserve $U_k$, for positive claim size $B$:

\[ U_k = R_k(U_{k-1} - B_k), \quad \text{for } k \geq 1, \]
\[ U_0 = u. \]

Iteration gives: $U_n = R_n \cdots R_1 u - (R_n \cdots R_1 B_1 + \cdots + R_N B_n)$. 

---

Thorbjörn Gudmundsson
KTH

Markov Chain Monte Carlo for rare-event simulation in heavy-tailed settings
Consider the following setup for the risk reserve $U_k$, for positive claim size $B$:

$$U_k = R_k(U_{k-1} - B_k), \text{ for } k \geq 1,$$
$$U_0 = u.$$

Iteration gives:

$$U_n = R_n \cdots R_1 u - (R_n \cdots R_1 B_1 + \cdots + R_N B_n).$$

Writing $A_k = 1/R_k$ then

$$A_1 \cdots A_n U_n = u - W_n, \text{ where}$$
$$W_n = B_1 + A_1 B_2 + \cdots + A_1 \cdots A_{n-1} B_n.$$
Problem

Thus the event of ruin can be expressed as follows

\[ \{ \inf_k U_k < 0 \} = \{ \sup_k W_k > u \} \]
Problem

Thus the event of ruin can be expressed as follows

$$\{\inf_k U_k < 0\} = \{\sup_k W_k > u\}.$$  

Goal: Construct an MCMC estimator for computing

$$p = \mathbb{P} (\sup_k W_k > u).$$
Gibbs sampler

Construct a Markov chain \((A_t, B_t)_{t \geq 0}\) with the invariant distribution

\[
\mathbb{P}\left((A, B) \in \cdot \mid \sup_k W_k > u\right).
\]
Gibbs sampler

Construct a Markov chain \((A_t, B_t)_{t \geq 0}\) with the invariant distribution

\[
\mathbb{P}\left((A, B) \in \cdot \mid \sup_{k} W_k > u\right).
\]

Carried out by updating one of \((A_1, \ldots, A_n, B_1, \ldots, B_n)\) at a time, conditioned so that

\[
\max_{1 \leq k \leq n} W_k = \max_{1 \leq k \leq n} B_1 + A_1 B_2 + \cdots + A_1 \cdots A_{k-1} B_k > u.
\]
Assume that

- The claim size $B$ is Pareto($\alpha$) distributed
- The stochastic return $R$ fulfills $\mathbb{E}[R^{-\alpha-\epsilon}] < \infty$ for some $\epsilon > 0$
Efficiency

Assume that
- The claim size $B$ is Pareto($\alpha$) distributed
- The stochastic return $R$ fulfills $\mathbb{E}[R^{-\alpha - \epsilon}] < \infty$ for some $\epsilon > 0$

Then we have the asymptotic result

$$\frac{\mathbb{P}(\sup_{1 \leq k \leq n} W_k > u)}{\mathbb{P}(B > u) \sum_{k=0}^{n-1} \mathbb{E}[A^\alpha]^k} \to 1, \quad \text{as } n \to \infty.$$
Efficiency continued

Now \( W_n = B_1 + A_1 B_2 + \cdots + A_1 \cdots A_{n-1} B_n \).

Based on the existing asymptotic results we propose the following choice for \( V \)

\[
V(\cdot) = \mathbb{P}((A, B) \in \cdot \mid (A, B) \in R),
\]

where

\[
R = \{ B_1 > u \}.
\]
Efficiency continued

Now $W_n = B_1 + A_1 B_2 + \cdots + A_1 \cdots A_{n-1} B_n$.

Based on the existing asymptotic results we propose the following choice for $V$

$$V(\cdot) = \mathbb{P}((A, B) \in \cdot \mid (A, B) \in R),$$

where

$$R = \{B_1 > u\} \cup \{A_1 > a, B_2 > u/a\},$$
Efficiency continued

Now \( W_n = B_1 + A_1 B_2 + \cdots + A_1 \cdots A_{n-1} B_n \).

Based on the existing asymptotic results we propose the following choice for \( V \)

\[
V(\cdot) = \mathbb{P}\left((A, B) \in \cdot \mid (A, B) \in R\right),
\]

where

\[
R = \{B_1 > u\} \cup \{A_1 > a, B_2 > u/a\} \cup \ldots \cup \{A_1 > a, \ldots, A_{n-1} > a, B_n > u/a^{n-1}\}.
\]
10,000 simulations for $n = 10$ and $u = 10^5$
Conclusion

Established a framework for new and simple method within stochastic simulation: Markov chain Monte Carlo methodology.
Established a framework for new and simple method within stochastic simulation: Markov chain Monte Carlo methodology. Applied the framework and proved efficiency on four concrete examples:

- Random walk with heavy-tails
- Random sum with heavy-tails
- Solution to stochastic recurrent equations with heavy-tailed innovations
- Insurance model with risky investments and Pareto distributed claim size
Conclusion

Possibilities for future work:

- Extension to random walk with light-tails
- Perfect simulation / coupling form the past
- Solution to stochastic recurrent equations where the ruin event is controlled by the stochastic returns rather than the claim size
Thank you for your attention!