Markov chain Monte Carlo for computing probabilities of rare events in a heavy-tailed random walk

Thorbjörn Gudmundsson and Henrik Hult

Department of Mathematics
KTH Stockholm

KTH, June 2012

Setup

- 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.

Setup

- 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.

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

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

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

Problem

Problem: compute $p = \mathbb{P}\left(\frac{S_m}{m} > a\right)$.

Problem

Problem: compute $p = \mathbb{P}\left(\frac{S_m}{m} > a\right)$.

- Sometimes no analytical solution known,
- Monte Carlo simulation approach computationally inefficient for small p.

Problem

Problem: compute $p = \mathbb{P}\left(\frac{S_m}{m} > a\right)$.

- Sometimes no analytical solution known,
- Monte Carlo simulation approach computationally inefficient for small p.
- Goal: construct an efficient estimator $\hat{p}$ in the sense that

$$\mathrm{RE}(\hat{p}) := \frac{\mathbb{V}ar(\hat{p})}{p^2}$$

is bounded or tends to zero as $p \to 0$.

Importance sampling

Goal: construct an efficient estimator $\hat{p}$.

Importance sampling

Goal: construct an efficient estimator $\hat{p}$.
The importance sampling approach (Dupuis et al 2007)

- Generate n copies of X. independently 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 continued

- Reduces to finding a suitable sampling distribution G.

Importance sampling continued

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

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

Seems difficult sampling directly from F_A since it requires knowledge of $\mathbb{P}(X \in A)$!

Idea

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

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

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.

Estimator

- Construct a Markov chain $(X_k)_{k\geq 1}$ via MCMC sampler, with the zero-variance distribution F_A as its invariant distribution.

Estimator

- 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

$$u((X_k)_{k\geq 1}) = \frac{1}{n}\sum_{k=1}^{n} \frac{v(X_k)\mathbb{I}\{X_k \in A\}}{f(X_k)}.$$

Estimator continued

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

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

Estimator continued

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

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

- Define $\hat{p} = \left(\frac{1}{n}\sum_{k=1}^{n}\frac{v(X_k)\mathbb{I}\{X_k \in A\}}{f(X_k)}\right)^{-1}$.

Design issues

Estimator $\hat{p} = \left(\frac{1}{n} \sum_{k=1}^{n} \frac{v(X_k)\mathbb{I}\{X_k \in A\}}{f(X_k)} \right)^{-1}$.

- Choice of v: controls the variance, set to ensure rare-event efficiency of the algorithm

Design issues

Estimator $\hat{p} = \left(\frac{1}{n} \sum_{k=1}^{n} \frac{v(X_k)\mathbb{I}\{X_k \in A\}}{f(X_k)} \right)^{-1}$.

- Choice of v: controls the variance, set to ensure rare-event efficiency of the algorithm
- Choice of the MCMC sampler: crucial to control the dependence of the Markov chain

Controlling the variance

Goal: ensure $\mathbb{V}ar(\hat{p})/p^2$ is bounded as $p \to 0$.

Controlling the variance

Goal: ensure $\mathbb{V}ar(\hat{p})/p^2$ is bounded as $p \to 0$.

- Taylor expansion of $g(Z)$ around $\mathbb{E}[Z]$ leads to

$$\begin{aligned} \mathbb{V}ar\big(g(Z)\big) &\approx \mathbb{V}ar\big(g(\mathbb{E}[Z]) + g'(\mathbb{E}[Z])(Z - \mathbb{E}[Z])\big) \\ &= \big(g'(\mathbb{E}[Z])\big)^2 \mathbb{V}ar(Z). \end{aligned}$$

Controlling the variance

Goal: ensure $\mathbb{V}ar(\hat{p})/p^2$ is bounded as $p \to 0$.

- Taylor expansion of $g(Z)$ around $\mathbb{E}[Z]$ leads to

$$\begin{aligned} \mathbb{V}ar\big(g(Z)\big) &\approx \mathbb{V}ar\big(g(\mathbb{E}[Z]) + g'(\mathbb{E}[Z])(Z - \mathbb{E}[Z])\big) \\ &= \big(g'(\mathbb{E}[Z])\big)^2 \mathbb{V}ar(Z). \end{aligned}$$

- Applied to $g(Z) = 1/Z$ and $Z = \frac{1}{n}\sum_{k=1}^{n} u(X_k)$ where

$$u(x) = \frac{v(x)\mathbb{I}\{x \in A\}}{f(x)},$$

then leads to

$$\mathbb{V}ar_{F_A}(\hat{p}) \approx p^4 \mathbb{V}ar\Big(\frac{1}{n}\sum_{k=1}^{n} u(X_k)\Big) \leq C \cdot p^4 \mathbb{V}ar\big(u(X)\big).$$

Controlling the variance continued

Proposition

If $p^2 \mathbb{V}ar_{F_A}(u(X)) \to 0$ as $p \to 0$ then $\hat{p}$ has vanishing relative error (is sufficient).

How do we choose v to fulfill this proposition?

Controlling the variance continued

- Consider the term

$$
\begin{aligned}
p^2 \mathbb{V}ar\big(u(X)\big) &= p^2\big(\mathbb{E}[u(X)^2] - \mathbb{E}[u(X)]^2\big) \\
&= p^2\Big(\int_A \frac{v^2(x)}{f^2(x)} \frac{f(x)}{p} dx - 1\Big) \\
&= p \int_A \frac{v^2(x)}{f(x)} dx - 1,
\end{aligned}
$$

Controlling the variance continued

- Consider the term

$$\begin{aligned} p^2 \mathbb{V}ar\big(u(X)\big) &= p^2 \big(\mathbb{E}[u(X)^2] - \mathbb{E}[u(X)]^2\big) \\ &= p^2 \Big(\int_A \frac{v^2(x)}{f^2(x)} \frac{f(x)}{p} dx - 1 \Big) \\ &= p \int_A \frac{v^2(x)}{f(x)} dx - 1, \end{aligned}$$

- choosing $v(x) = f_A(x) = f(x)\mathbb{I}\{x \in A\}/p$ implies

$$p^2 \mathbb{V}ar\big(u(X)\big) = p \int_A \frac{f^2(x)/p^2}{f(x)} dx - 1 = \frac{1}{p} \int_A f(x) dx - 1 = 0.$$

Controlling the variance continued

- Consider the term

$$
\begin{aligned}
p^2 \mathbb{V}ar(u(X)) &= p^2\big(\mathbb{E}[u(X)^2] - \mathbb{E}[u(X)]^2\big) \\
&= p^2\Big(\int_A \frac{v^2(x)}{f^2(x)} \frac{f(x)}{p} dx - 1\Big) \\
&= p \int_A \frac{v^2(x)}{f(x)} dx - 1,
\end{aligned}
$$

- choosing $v(x) = f_A(x) = f(x)\mathbb{I}\{x \in A\}/p$ implies

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

Choose v as an approximation of the zero-variance density!

Recipe

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

Recipe

- Sample $(X_k)_{k\geq 1}$ under F_A via MCMC
- Show $p^2 \mathbb{V}ar\big(u(X)\big) \to 0$ as $p \to 0$

Recipe

- Sample $(X_k)_{k\geq 1}$ under F_A via MCMC
- Show $p^2\mathbb{V}ar(u(X)) \to 0$ as $p \to 0$
- Show $(X_k)_{k\geq 1}$ is geometric ergodic

Setup

- Consider a random walk $S_m = Y_1 + \cdots + Y_m$ with non-negative steps Y's with known heavy-tailed distribution F_Y and objective of computing

$$p = \mathbb{P}\Big(\frac{S_m}{m} > a\Big),$$

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

Setup

- Consider a random walk $S_m = Y_1 + \cdots + Y_m$ with non-negative steps Y's with known heavy-tailed distribution F_Y and objective of computing
$$p = \mathbb{P}\Big(\frac{S_m}{m} > a\Big),$$
where a is much larger than $\mathbb{E}[Y]$.
- Construct $(\mathbf{Y}_k)_{k\geq 1}$ via MCMC with invariant density
$$f_A(\mathbf{y}) = \frac{f_{\mathbf{Y}}(\mathbf{y})\mathbb{I}\{y_1 + \cdots + y_m > am\}}{\mathbb{P}(S_m > am)}.$$

Setup

- Consider a random walk $S_m = Y_1 + \cdots + Y_m$ with non-negative steps Y's with known heavy-tailed distribution F_Y and objective of computing
$$p = \mathbb{P}\Big(\frac{S_m}{m} > a\Big),$$
where a is much larger than $\mathbb{E}[Y]$.
- Construct $(\mathbf{Y}_k)_{k\geq 1}$ via MCMC with invariant density
$$f_A(\mathbf{y}) = \frac{f_{\mathbf{Y}}(\mathbf{y})\mathbb{I}\{y_1 + \cdots + y_m > am\}}{\mathbb{P}(S_m > am)}.$$
- A typical such a random walk has a $m - 1$ number of "small" steps and one "large" step.

Gibbs sampler

Initial state $\mathbf{Y}_0 = (Y_{0,1}, \dots, Y_{0,m})$ such that $Y_{0,1} > am$ and $Y_{0,i} = 0$ for other indices. Given $\mathbf{Y}_k = (Y_{k,1}, \dots, Y_{k,m})$, $k = 0, 1, \dots$ the next state $\mathbf{Y}_{k+1}$ is sampled as follows

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

Gibbs sampler

Initial state $\mathbf{Y}_0 = (Y_{0,1}, \ldots, Y_{0,m})$ such that $Y_{0,1} > am$ and $Y_{0,i} = 0$ for other indices. Given $\mathbf{Y}_k = (Y_{k,1}, \ldots, Y_{k,m})$, $k = 0, 1, \ldots$ the next state $\mathbf{Y}_{k+1}$ is sampled as follows

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

Gibbs sampler

Initial state $\mathbf{Y}_0 = (Y_{0,1}, \dots, Y_{0,m})$ such that $Y_{0,1} > am$ and $Y_{0,i} = 0$ for other indices. Given $\mathbf{Y}_k = (Y_{k,1}, \dots, Y_{k,m})$, $k = 0, 1, \dots$ the next state $\mathbf{Y}_{k+1}$ is sampled as follows

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

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

Gibbs sampler

Initial state $\mathbf{Y}_0 = (Y_{0,1}, \ldots, Y_{0,m})$ such that $Y_{0,1} > am$ and $Y_{0,i} = 0$ for other indices. Given $\mathbf{Y}_k = (Y_{k,1}, \ldots, Y_{k,m})$, $k = 0, 1, \ldots$ the next state $\mathbf{Y}_{k+1}$ is sampled as follows

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

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

- Permutate the steps in $\mathbf{Y}_{k+1}$.

Gibbs sampler continued

Proposition

The Markov chain $(\mathbf{Y}_k)_{k\geq 1}$ constructed using the proposed Gibbs sampler has the conditional distribution F_A as its invariant distribution.

MCMC estimator

- The MCMC estimator $\hat{p} = \left(\frac{1}{n} \sum_{k=1}^{n} \frac{v(\mathbf{y}_k)\mathbb{I}\{S_m > am\}}{f(\mathbf{y}_k)} \right)^{-1}$. The steps are heavy-tailed in the sense that

$$\frac{\mathbb{P}(M_m > am)}{\mathbb{P}(S_m > am)} \to 1,$$

where $M_m = \max_i \{y_{k,i}\}$.

MCMC estimator

- The MCMC estimator $\hat{p} = \left(\frac{1}{n} \sum_{k=1}^{n} \frac{v(\mathbf{y}_k)\mathbb{I}\{S_m > am\}}{f(\mathbf{y}_k)} \right)^{-1}$. The steps are heavy-tailed in the sense that

$$\frac{\mathbb{P}(M_m > am)}{\mathbb{P}(S_m > am)} \to 1,$$

where $M_m = \max_i \{y_{k,i}\}$.

- Therefore seems smart to use

$$\mathbb{P}(\mathbf{Y} \in \cdot \mid M_m > am) \ \text{ as a proxy for } \ \mathbb{P}(\mathbf{Y} \in \cdot \mid S_m > am).$$

Propose

$$v(\mathbf{y}_k) = \frac{f(\mathbf{y}_k)\mathbb{I}\{M_m > am\}}{\mathbb{P}(M_m > am)}.$$

MCMC estimator continued

Choosing $v(\mathbf{y}) = \frac{f(\mathbf{y})\mathbb{I}\{M_m > am\}}{\mathbb{P}(M_m > am)}$ yields

$$u(\mathbf{y}) = \frac{v(\mathbf{y})\mathbb{I}\{S_m > am\}}{f(\mathbf{y})} = \frac{\mathbb{I}\{M_m > am\}}{\mathbb{P}(M_m > am)}.$$

MCMC estimator continued

Choosing $v(\mathbf{y}) = \frac{f(\mathbf{y})\mathbb{I}\{M_m > am\}}{\mathbb{P}(M_m > am)}$ yields

$$u(\mathbf{y}) = \frac{v(\mathbf{y})\mathbb{I}\{S_m > am\}}{f(\mathbf{y})} = \frac{\mathbb{I}\{M_m > am\}}{\mathbb{P}(M_m > am)}.$$

$$\hat{p} = \mathbb{P}(M_m > am)\Big(\frac{1}{n}\sum_{k=1}^{n} \mathbb{I}\{M_m(k) > am\}\Big)^{-1}$$

Efficiency

$$p^2 \mathbb{V}ar_{F_A}(u(\mathbf{Y})) = \frac{\mathbb{P}(S_m > am)^2}{\mathbb{P}(M_m > am)^2} \mathbb{V}ar_{F_A}(\mathbb{I}\{M_m > am\})$$

Efficiency

$$p^2 \mathbb{V}ar_{F_A}(u(\mathbf{Y})) = \frac{\mathbb{P}(S_m > am)^2}{\mathbb{P}(M_m > am)^2} \mathbb{V}ar_{F_A}(\mathbb{I}\{M_m > am\})$$

$$= \frac{\mathbb{P}(S_m > am)^2}{\mathbb{P}(M_m > am)^2} \Big(\mathbb{E}_{F_A}[\mathbb{I}\{M_m > am\}] - \mathbb{E}_{F_A}[\mathbb{I}\{M_m > am\}]^2 \Big)$$

Efficiency

$$p^2 \mathbb{V}ar_{F_A}(u(\mathbf{Y})) = \frac{\mathbb{P}(S_m > am)^2}{\mathbb{P}(M_m > am)^2} \mathbb{V}ar_{F_A}(\mathbb{I}\{M_m > am\})$$

$$= \frac{\mathbb{P}(S_m > am)^2}{\mathbb{P}(M_m > am)^2} \left(\mathbb{E}_{F_A}[\mathbb{I}\{M_m > am\}] - \mathbb{E}_{F_A}[\mathbb{I}\{M_m > am\}]^2 \right)$$

$$= \frac{\mathbb{P}(S_m > am)^2}{\mathbb{P}(M_m > am)^2} \left(\frac{\mathbb{P}(M_m > am)}{\mathbb{P}(S_m > am)} - \frac{\mathbb{P}(M_m > am)^2}{\mathbb{P}(S_m > am)^2} \right)$$

Efficiency

$$
\begin{aligned}
p^2 \mathbb{V}ar_{F_A}(u(\mathbf{Y})) &= \frac{\mathbb{P}(S_m > am)^2}{\mathbb{P}(M_m > am)^2} \mathbb{V}ar_{F_A}(\mathbb{I}\{M_m > am\}) \\
&= \frac{\mathbb{P}(S_m > am)^2}{\mathbb{P}(M_m > am)^2} \Big(\mathbb{E}_{F_A}[\mathbb{I}\{M_m > am\}] - \mathbb{E}_{F_A}[\mathbb{I}\{M_m > am\}]^2 \Big) \\
&= \frac{\mathbb{P}(S_m > am)^2}{\mathbb{P}(M_m > am)^2} \left(\frac{\mathbb{P}(M_m > am)}{\mathbb{P}(S_m > am)} - \frac{\mathbb{P}(M_m > am)^2}{\mathbb{P}(S_m > am)^2} \right) \\
&= \frac{\mathbb{P}(S_m > am)}{\mathbb{P}(M_m > am)} - 1 \to 0 \quad \text{as } p \to 0.
\end{aligned}
$$

Geometric ergodicity

- The design of the Gibbs sampler ensures that the Markov chain $(\mathbf{Y}_k)_{k\geq 1}$ is (uniformly) ergodic.

Geometric ergodicity

- The design of the Gibbs sampler ensures that the Markov chain $(\mathbf{Y}_k)_{k\geq 1}$ is (uniformly) ergodic.
- This guarantees that the chain mixes sufficiently and hence that $\mathbb{V}ar(\hat{p}) \to 0$ as $n \to \infty$ at same speed as $1/n$.

Geometric ergodicity

- The design of the Gibbs sampler ensures that the Markov chain $(\mathbf{Y}_k)_{k\geq 1}$ is (uniformly) ergodic.
- This guarantees that the chain mixes sufficiently and hence that $\mathbb{V}ar(\hat{p}) \to 0$ as $n \to \infty$ at same speed as $1/n$.
- The proof is technical ..

Concluding remarks

- $\hat{p}$ is an efficient estimator for heavy-tailed random walk for increasing (but fixed) number of steps.
- Extension to heavy-tailed random sum $\sum_{k=1}^{N} Y_k$ where N is stochastic.
- Other models such as recursion formulas, queues, ...

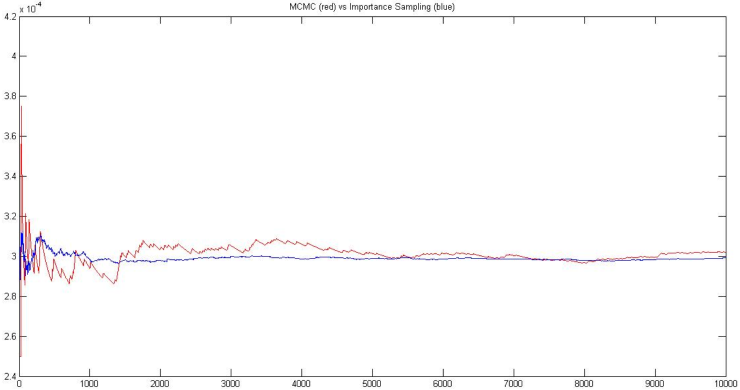
Assumptions

- The MCMC estimator $\hat{p}$ tested against importance sampling and standard Monte Carlo.
- Steps are Pareto(2) distributed.
- Number of batches: 25, simulations per batch: $10,000$.

Table

m	a	MCMC	IS	MC	
5	10	3.40e-3 (0.81e-4) [4.1]	2.91e-3 (1.77e-4) [3.4]	2.83e-3 (4.74e-4) [0.7]	Avg. est. (Std. dev.) [Avg. time (ms)]
10	20	3.34e-4 (5.83e-6)	3.02e-4 (2.02e-6)	2.68e-4 (162.58e-6)	Avg. est. (Std. dev.)

$10,000$ simulations for $m = 10$ and $a = 20$



10, 000 simulations for $m = 10$ and $a = 20$