KTH VETENSKAP OCH KONST

ROYAL INSTITUTE OF TECHNOLOGY

MCMC for computing probabilities of rare events

Thorbjörn Gudmundsson & Henrik Hult

Royal Institute of Technology, Stockholm, Sweden

Setting

Consider a random walk

$$S_m = Y_1 + \cdots + Y_m,$$

where the Ys are i.i.d. with known distribution. The objective is to compute the probability

$$p_m = \mathbb{P}(S_m > am), \quad \text{for } m \text{ large and } a > \mathbb{E}[Y].$$

- Sometimes no analytical solution known.
- An alternative is stochastic simulation.

Monte Carlo

1. Generate n independent copies $S_m(1), \ldots, S_m(n)$.
2. Compute empirical estimate

$$\hat{p}_m = \frac{1}{n}\sum_{i=1}^{n} I\{S_m(i) > am\}.$$

Simple to implement, unbiased and consistent.

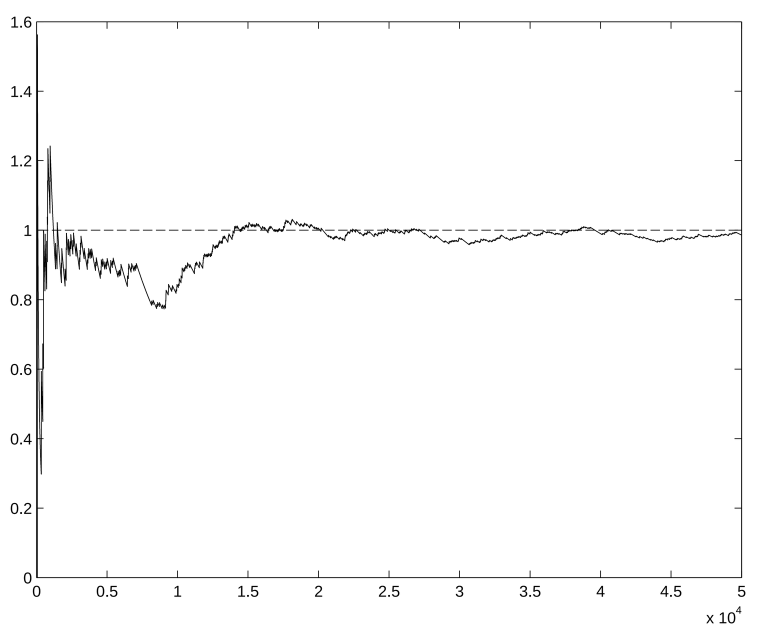

Figure 1. The Monte Carlo estimate compared against the true probability in the case when the Ys are Cauchy. The estimate is shown to converge to the true value with number of simulations, but with considerable variation.

What about efficiency?
We would like the relative error $\text{Std}(\hat{p}_m)/p_m$ to be bounded (or vanishing). For the Monte Carlo estimate

$$\frac{\text{Std}(\hat{p}_m)}{p_m} = \frac{1}{\sqrt{n}}\frac{\sqrt{p_m - p_m^2}}{p_m} \sim \frac{1}{\sqrt{np_m}} \to \infty$$

as $p_m \to 0$. For rare events Monte Carlo requires a large computational cost.

MC estimate is not efficient.

Importance sampling

Denote the original distribution of S_m by F and density by f.

1. Generate n independent copies $S_m(1), \ldots, S_m(n)$ from a sampling distribution G.
2. Compute empirical estimate

$$\hat{p}_m = \frac{1}{n}\sum_{i=1}^{n} \frac{dF}{dG} I\{S_m(i) > am\}.$$

Both unbiased and consistent.

Problem reduced to finding a suitable sampling distribution G - can be difficult.

MCMC algorithm

There exists a **best choice** for G that gives zero variance, the conditional distribution given by

$$\mathbb{P}(S_m \in \cdot \mid S_m > am).$$

Its density is given by

$$g(x) = \frac{f(x)I\{x > am\}}{\mathbb{P}(S_m > am)}.$$

An MCMC algorithm is a tool to sample a random variable despite only knowing its density up to a normalising constant. The density of S_m under G is precisely of that nature.

Sample n (dependent) copies $S_m(1), \ldots, S_m(n)$ via MCMC from the zero variance distribution G.

$$S_m(i) \sim g(\cdot) = \frac{f(\cdot)I\{\cdot > am\}}{p_m}.$$

How to extract the information about the normalising constant p_m from the sample?

MCMC estimator

$$\mathbb{E}[u(S_m)] = \int u(x)g(x)dx = \int_{x>am} u(x)\frac{f(x)}{p_m}dx.$$

Setting $u(x) = \frac{v(x)}{f(x)}I\{x > am\}$

$$\mathbb{E}[u(S_m)] = \frac{1}{p_m}\int_{x>am} v(x)dx.$$

So choosing v is such that $\int_{x>am} v(x)dx = 1$

$$\mathbb{E}[u(S_m)] = \frac{1}{p_m}.$$

Consistent MCMC estimator given by

$$\hat{p}_m = \Big(\frac{1}{n}\sum_{i=1}^{n} u(S_m(i))\Big)^{-1}, \qquad (1)$$

where u and v are given by the above.

How should one choose v to ensure efficiency such as in the Importance Sampling case?

Efficiency of the MCMC estimator

Consider the relative error of the estimator in (1). First order Taylor approximation gives

$$\frac{\mathbb{V}ar(\hat{p}_m)}{p_m^2} \approx \frac{p_m^2}{n}\mathbb{V}ar\big(u(S_m)\big) + \text{covariance term}.$$

The covariance term can be shown to be vanishing and have no significant impact on the convergence.

$$\begin{aligned}\frac{p_m^2}{n}\mathbb{V}ar\big(u(S_m)\big) &= \frac{p_m^2}{n}\Big(\mathbb{E}\big[u(S_m)^2\big] - \big(\mathbb{E}[u(S_m)]\big)^2\Big)\\ &= \frac{p_m^2}{n}\Big(\mathbb{E}\big[u(S_m)^2\big] - \frac{1}{p_m^2}\Big)\\ &= \frac{1}{n}\Big(p_m^2\int_{x>am}\frac{v(x)^2}{f(x)^2} - 1\Big)\end{aligned}$$

Choosing $v(x) = g(x) = \frac{f(x)I\{x>am\}}{p_m}$ gives zero variance.

Heuristics: v is chosen as an approximation of the zero variance density g

Main Result

Consider a random walk

$$S_m = Y_1 + \cdots + Y_m,$$

where the Ys are i.i.d. with heavy tails in the following sense

$$\frac{\mathbb{P}(S_m > am)}{\mathbb{P}(M_m > am)} \to 1 \quad \text{as } m \to \infty,$$

$M_m = \max\{Y_1, \ldots, Y_m\}$, e.g. Cauchy, regularly varying, subexponential. The objective is to compute the probability

$$p_m = \mathbb{P}(S_m > am), \quad \text{for } m \text{ large and } a > \mathbb{E}[Y].$$

The zero variance distribution is

$$\mathbb{P}(S_m \le x \mid S_m > am),$$

and because of the heavy-tail nature of the Ys, choose v as the density of

$$\mathbb{P}(S_m \le x \mid M_m > am).$$

This choice of v gives a consistent and efficient MCMC estimator:

$$\begin{aligned}\hat{p}_m &= \Big(\frac{1}{n}\sum_{i=1}^{n}\frac{v(S_m(i))}{f(S_m(i))}I\{S_m(i) > am\}\Big)^{-1}\\ &= p_{\max}\Big(\frac{1}{n}\sum_{i=1}^{n} I\{M_m(i) > am\}\Big)^{-1},\end{aligned}$$

where

$$p_{\max} = \mathbb{P}(M_m > am) = 1 - F_Y(am)^m,$$

is easily calculated.

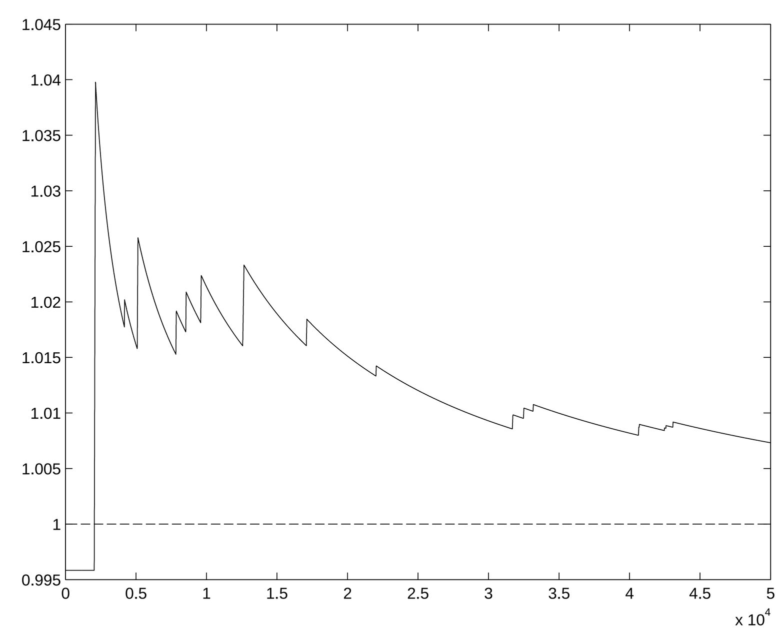

Figure 2. The Markov Chain Monte Carlo estimate compared against the true probability in the case when the Ys are Cauchy. The initial value is $p_{\max}$ and the estimate converges to the true value with number of simulations, with little variation.

References

- S. Asmussen and P. W. Glynn. *Stochastic Simulation*, volume 57 of *Stochastic Modelling and Applied Probability*. Springer, New York, 2007.
- P. Dupuis, K. Leder, and H. Wang. *Importance sampling for sums of random variables with regularly varying tails*. *ACM Transactions on Modeling and Computer Simulation*, 17(3), 2007.
- P. Doukhan. *Mixing, Properties and Examples*, volume 84 of *Lecture Notes in Statistics*, Springer-Verlag, 1994.