IMPORTANCE SAMPLING FOR STOCHASTIC RECURRENCE EQUATIONS WITH HEAVY TAILED INCREMENTS

Jose Blanchet

Henrik Hult

Columbia University
500 West 120th St
New York, NY 10027 USA

Royal Institute of Technology
SE-100 44 Stockholm
Sweden

Kevin Leder

University of Minnesota - Twin Cities
111 Church St. SE
Minneapolis, MN 55455

ABSTRACT

Importance sampling in the setting of heavy tailed random variables has generally focused on models with additive noise terms. In this work we extend this concept by considering importance sampling for the estimation of rare events in Markov chains of the form

$$X_{n+1} = A_{n+1}X_n + B_{n+1}, \quad X_0 = 0,$$

where the B_n's and A_n's are independent sequences of independent and identically distributed (i.i.d.) random variables and the B_n's are regularly varying and the A_n's are suitably light tailed relative to B_n. We focus on efficient estimation of the rare event probability $P(X_n > b)$ as $b \nearrow \infty$. In particular we present a strongly efficient importance sampling algorithm for estimating these probabilities, and present a numerical example showcasing the strong efficiency.

1 INTRODUCTION

The estimation of very small probabilities via Monte Carlo simulation is an important problem that has generated a very large amount of scientific literature over the past two decades (see, e.g., (Glasserman and Wang 1997, Glasserman and Kou. 1995, Dupuis and Wang 2007, Dupuis and Wang 2009). In the setting where the rare event is driven by random variables with heavy tails there has also been a significant amount of work done (Asmussen, Binswanger, and Hojgaard 2000, Asmussen and Binswanger 1997, Dupuis, Leder, and Wang 2006, Blanchet and Glynn 2008, Blanchet, Glynn, and Liu 2007). These works and others have studied the problem of rare even simulation for a large class of problems involving heavy tailed random variables. However, all of the work until now on the subject has focused on the setting of purely additive noise. Naturally, the ability to study rare events in models with additive and multiplicative noise is very important in many applications. Based on this we study rare event simulation for Markov chains of the form

$$X_{k+1} = A_{k+1}X_k + B_{k+1}, \quad X_0 = 0, \tag{1}$$

where A_k and B_k are independent i.i.d. sequences. We assume that B_k is regularly varying with index $\alpha > 0$ and that A_k satisfies

$$EA_k^{2\alpha+\varepsilon} < \infty, \tag{2}$$

for some $\varepsilon > 0$. In this paper we are focused solely on the efficient estimation of the probability $p_b = P(X_n > b)$ for b very large and n fixed, in a following work we will present simulation algorithms for estimating infinite horizon probabilities associated with the Markov chain. In this work we present an unbiased importance sampling estimator $\hat{p}_b$ of p_b that is strongly efficient as $b \to \infty$, i.e.,

$$\sup_{b < \infty} \frac{E[\hat{p}_b]}{p_b^2} < \infty.$$

2 RANDOM WALK MODEL

Before studying the Markov chain in (1) we consider rare event estimation for a simpler process. In particular, again let $\{B_k\}_{k=1}^n$ be an i.i.d. sequence of random variables with regularly varying tails, then we define a Markov chain Y_n as

$$Y_{k+1} = Y_k + c_{k+1}B_{k+1}, \quad Y_0 = 0,$$

for some positive constants $c_0, \ldots, c_n$. We denote the density of B_k by f, and its tail probability by $\bar{F}$. The goal of this section will be to develop strongly efficient estimators of the probability $r_b = P(Y_n > b)$. Of course in order to establish that an estimator is strongly efficient it is necessary to know the asymptotics of r_b. From the subexponential property of the increments we know that (see, e.g., Asmussen 2000)

$$r_b \sim \sum_{k=1}^n P(B_k > b/c_k). \tag{3}$$

Our method for establishing strong efficiency will be the construction of a Lyapunov function and then the verification of a Lyapunov inequality, for more details on this method see, e.g., (Blanchet, Glynn, and Liu 2007). A Lyapunov function will be a function of time and space, $V_b(j,x)$ that approximates $P(X_n > b | X_j = x)^2$, then using a verification argument it will be possible to control the second moment of our importance sampling estimator. The basic intuition for the construction of the Lyapunov function is to make it nearly proportional to the asymptotic approximation of $P(X_n > b | X_j = x)^2$.

Based on these ideas and (3) we propose the following Lyapunov function

$$V_b(j,x) = \min\left\{ d_j \left(\sum_{k=j+1}^n P(B > (b-x)/c_k) \right)^2, 1 \right\}, \tag{4}$$

for some constants d_j. Since $V_b(\cdot,\cdot)$ ultimately will serve as a bound for the second moment of our estimator, the d_j's must be chosen as small as possible but large enough so that $V_b(j,x) < 1$ implies $b - x > 0$.

2.1 Sampling Measure

The construction of the sampling measure will be based on the heuristic that rare events in sums of heavy tailed random variables are caused by one large jump. Therefore the increments of our random walk will be sampled using a mixture. The mixture samples the random variables B_k according to either their nominal distribution, or induces an extremely large value in the random variable. The method we will use to generate the large value in B_k will be to sample the random variables conditional on one of them taking a large value.

Our sampling density of B_{j+1} given $X_j = x$ is

$$
\begin{aligned}
g_{j+1}(y|x) = {} & f(y)I\{x \in \Gamma_{j+1}^c\} \\
& + \frac{p(j+1,x)f(y)}{P\{B > a(b-x)/c_{j+1}\}} I\{y > a(b-x)/c_{j+1}\}I\{x \in \Gamma_{j+1}\} \\
& + \frac{(1-p(j+1,x))f(y)}{P\{B \le a(b-x)/c_{j+1}\}} I\{y \le a(b-x)/c_{j+1}\}I\{x \in \Gamma_{j+1}\}.
\end{aligned}
$$

Here Γ_{j+1} is interpreted as the set of x-values where we perform importance sampling and we have that

$$\Gamma_j = \{x : V_b(j,x) < 1\}.$$

If $X_j = x$ is already close to b, then the event $X_n > b$ is not considered to be rare and importance sampling is not needed. Therefore B is sampled according to its original density f if $x \in \Gamma^c_{j+1}$. If $x \in \Gamma_{j+1}$ then we sample B according to the mixture distribution, which is the original distribution conditioned on not being too large (with probability $1 - p(j+1,x)$) and the original distribution conditioned on being large enough to nearly bring the sum near the threshold b (with probability $p(j+1,x)$). The phrase 'nearly' is present because of the parameter $a \in (0,1)$ that serves as a cushion.

2.2 Analysis of Sampling Measure

Our analysis of the sampling measure will be carried out via the Lyapunov function $V_b(j,x)$. In particular we will seek to establish the following inequalities

$$V_b(j,x) \geq E\left[\frac{f_{j+1|j}(X_{j+1}\mid X_j)}{g_{j+1|j}(X_{j+1}\mid X_j)}V_b(j+1,X_{j+1})I\{X_n > b\}\,\Big|\,X_j = x\right], \tag{5}$$

$$V_b(n,x) \geq I\{x > b\},$$

where $f_{j+1|j}$ is the transition kernel of Y_k under the original measure, and $g_{j+1|j}$ is the transition kernel of Y_k under the sampling measure. For the random walk model and sampling measures $\{g_j\}$ we can write this ratio as

$$\frac{f_{j+1|j}(X_{j+1}\mid x)}{g_{j+1|j}(X_{j+1}\mid x)} = \frac{f\left((X_{j+1}-x)/c_{j+1}\right)}{g_{j+1}\left((X_{j+1}-x)/c_{j+1}\mid x\right)}.$$

If we establish the Lyapunov inequalities, (5), then we can use a verification argument to establish the result

Lemma 1 Suppose there are constants β_j, $j = 0,\ldots,n-1$ such that $\beta_j \geq 1$ and for $j = 1,\ldots,n$

$$E\left[\frac{V_b(j+1,X_{j+1})}{V_b(j,x)}\frac{f_{j+1|j}(X_{j+1}\mid x)}{g_{j+1|j}(X_{j+1}\mid x)}\,\Big|\,X_j = x\right] \leq \beta_{j+1},$$

then,

$$E\left[\prod_{j=0}^{n-1}\frac{f_{j+1|j}(X_{j+1}\mid X_j)}{g_{j+1|j}(X_{j+1}\mid X_j)}I\{X_n > b\}\right] \leq \left(\prod_{j=0}^{n-1}\beta_{j+1}\right)V_b(0,0).$$

A proof of this result can be found in our forthcoming work (Blanchet, Hult, and Leder 2011).

In order for this lemma to be useful in establishing strong efficiency we need the property

$$\sup_{b<\infty}\frac{V_b(0,0)}{r_b^2} < \infty.$$

Looking at the definition of $V_b(j,x)$, and the asymptotics of r_b in (3) the previous result is obvious. Therefore if we establish the following result we will have established the strong efficiency of the importance sampling estimator based on $\{g_j\}_{j=1}^n$.

Theorem 2 There exists non-negative constants $\{\beta_j\}_{j=1}^n$ greater or equal to 1 such that for $j = 0,\ldots,n-1$ and $x \in \mathbb{R}$

$$E\left[\frac{V_b(j+1,X_{j+1})}{V_b(j,x)}\frac{f\left((X_{j+1}-x)/c_{j+1}\right)}{g_{j+1}\left((X_{j+1}-x)/c_{j+1}\mid x\right)}\,\Big|\,X_j = x\right] \leq \beta_{j+1}.$$

We will provide just a sketch of the ideas behind this proof, the interested reader is referred to our forthcoming work (Blanchet, Hult, and Leder 2011).

Proof. Assume that $V_b(j,x) < 1$ or $x \in \Gamma_{j+1}$. Then we can decompose the likelihood ratio as

$$J_1(j,x) = E\left[\frac{V_b(j+1,x+c_{j+1}B_{j+1})}{V_b(j,x)p(j+1,x)}I\{c_{j+1}B_{j+1} > a(b-x)\}\right]P\{c_{j+1}B_{j+1} > a(b-x)\}$$

$$J_2(j,x) = E\left[\frac{V_b(j+1,x+c_{j+1}B_{j+1})}{V_b(j,x)(1-p(j+1,x))}I\{c_{j+1}B_{j+1} \le a(b-x)\}\right]P\{c_{j+1}B_{j+1} \le a(b-x)\}.$$

The control of both terms reduces to bounding ratios of the form

$$\left(\frac{P(B > \gamma(b-x))}{P(B > b-x)}\right)^2, \tag{6}$$

for a positive constant γ (that will depend on the constant a), uniformly over all b and x such that $b-x$ is bounded above 0. Since we assume that B has regularly varying tails these ratios can be controlled by the use of Potter bounds, see Theorem 1.5.6 of (Bingham, Goldie, and Teugels 1987), i.e.,

$$\frac{P(B > \gamma(b-x))}{P(B > b-x)} \le A\max\{\gamma^{-\alpha+\delta}, \gamma^{-\alpha-\delta}\},$$

for constant $A > 1$ and $\delta > 0$. □

It follows that the constants $\{\beta_j\}$ from theorem 2 are controlled by the ratio in (6).

3 STOCHASTIC RECURRENCE EQUATION

Building on the analysis of the previous section, we are now able to study the Markov chain or stochastic recurrence equation (1). First from Breiman's theorem we know that

$$p_b = P\{X_n > b\} \sim P\{B_1 > b\}\sum_{k=0}^{n-1}E[(A_n\dots A_{n-k})^\alpha] \tag{7}$$

see, e.g., Lemma 2.2 of (Konstantinides and Mikosch 2005)

Observe that the state of the Markov chain at time step n can be written as

$$X_n = B_n + A_nB_{n-1} + \dots + A_n\dots A_2B_1.$$

Conditioning on $A_1 = a_1,\dots,A_n = a_n$ it follows that

$$X_{j+1} = X_j + c_{j+1}B_{j+1},$$
$$c_{j+1} = a_n\dots a_{j+2}. \tag{8}$$

Then, by first simulating $A_1,\dots,A_n$ from the original distribution the random walk algorithms can be implemented, conditional on $A_1,\dots,A_n$. With the constants c_k defined in terms of the values of $A_1,\dots,A_n$, we can use the Lyapunov function from the previous section (4). We then verify that this function is sufficiently small at the origin.

Lemma 3 The Lyapunov function $V_b(\cdot,\cdot)$ is sufficient for establishing bounded relative error for estimating p_b, that is,

$$\sup_{b<\infty}\frac{E[V_b(0,0)]}{p_b^2} < \infty.$$

Proof.   Referring to formula for $V_b(0,0)$, (4), we see that

$$E[V_b(0,0)] \leq d_0 E\left(\sum_{k=1}^{n} \bar{F}(b/C_k)\right)^2,$$

where $C_k = A_n \cdots A_{k+1}$, $k = 1,\ldots,n$. Using Potter bounds again we see that given $\delta > 0$ there exists $K(\delta) > 1$ such that

$$\frac{\bar{F}(b/C_k)}{\bar{F}(b)} \leq K(\delta)C_k^{\alpha+\delta},$$

where $\delta > 0$ is chosen so that $E[A^{2+\delta}] < \infty$. Therefore

$$E[V_b(0,0)] \leq d_0\bar{F}(b)^2 E\left(\sum_{k=1}^{n} K(\delta)C_k^{\alpha+\delta}\right)^2,$$

the lemma then follows by comparing this result with the tail probability asymptotics in (7).   □

Observe that in order for our algorithm to efficiently estimate $\hat{p}_b$ we needed the moment assumption (2).

We start by analyzing the algorithm based on the conditional distribution for the stochastic recurrence equation. Our algorithm for generating a sample of X_n is as follows.

- Set $\{d_j\}_j = 1^n$ according to (4)
- Sample $A_1,\ldots,A_n$ from their original distribution. Output is called $a_1,\ldots,a_n$.
- Set $c_{j+1} = a_n \cdots a_{j+2}$ for $j = 0,\ldots,n-1$.
- Choose $p_{j+1} \in (0,1)$ for $j = 0,\ldots,n-1$.
- Set $x = c_0$ and $j = 0$.
- Repeat the following until $j = n-1$.
    - If $V_b(j,x) \geq 1$,
        * Sample B_{j+1} from original distribution $B_{j+1} \rightarrow b_{j+1}$.
    - Otherwise
        * Draw U from uniform on $[0,1]$. If $U \leq p_{j+1}$ draw B_{j+1} from density

        $$\frac{f_B(y)}{P\{B > a(b-x)/c_{j+1}\}}I(y > a(b-x)/c_{j+1}),$$

        otherwise sample B_{j+1} from density

        $$\frac{f_B(y)}{P\{B \leq a(b-x)/c_{j+1}\}}I(y \leq a(b-x)/c_{j+1}).$$

        Store output as b_{j+1}.
    - Update $j \rightarrow j+1$ and $x \rightarrow x + c_{j+1}b_{j+1}$.

The vector of multiplicative terms will be denoted $\mathbf{A} = (A_1,\ldots,A_n)$, with this notation we can write the likelihood ratio from one step of the process X as

$$L_{j+1}^C(B_{j+1}|\mathbf{A},X_j = x) = I(x \notin \Gamma_{j+1}) + \frac{\bar{F}(a(b-x)/C_{j+1})}{p(j+1,x)}I(B_{j+1} > a(b-x)/C_{j+1})I(x \in \Gamma_{j+1})$$

$$+ \frac{1 - \bar{F}(a(b-x)/C_{j+1})}{1 - p(j+1,x)}I(B_{j+1} \leq a(b-x)/C_{j+1})I(x \in \Gamma_{j+1}),$$

for $j = 0, \ldots, n-1$, where $C_{j+1} = A_n \cdots A_{j+2}$. Based on this we can write the importance sampling estimator for p_b based on our algorithm as

$$\hat{p}_b = I(X_n > b) \prod_{j=0}^{n-1} L_{j+1}^C(B_{j+1} | \mathbf{A}, X_j). \tag{9}$$

We now present our theorem regarding the performance of the importance sampling measure

Theorem 4 The importance sampling estimator $\hat{p}_b$, defined in (9), estimates the probability p_b with bounded relative error, i.e.,

$$\sup_{b < \infty} \frac{E[\hat{p}_b]}{p_b^2} < \infty.$$

The proof largely follows the random walk example. We omit the details of this proof, the interested reader can consult (Blanchet, Hult, and Leder 2011).

3.1 Choice of Mixture Probabilities

In the statement of the algorithm we did not discuss the selection of the mixture probabilities. The reason for this is that the algorithm is strongly efficient for any choice of mixture probabilities $p_{j+1} \in (0,1)$. In our numerical simulations presented in Table 1 we choose the mixture probabilities for $j = 0, \ldots, n-1$ according to the formula $p_{j+1} = v_{j+1}/(1 + v_{j+1})$ where

$$v_{j+1} = \sqrt{d} \sum_{i=j+1}^{n} \left(\frac{ac_i}{c_{j+1}} \right)^{\alpha}. \tag{10}$$

In addition we numerically verified that the algorithm is relatively robust to the selection of the mixture probabilities. For details on the optimal selection of the mixture probabilities please see our forthcoming paper (Blanchet, Hult, and Leder 2011).

4 NUMERICAL EXAMPLES

In the table below we present numerical results from our algorithm. The bounded relative error property is clearly observed by looking at the performance of the algorithm for increasing values of b.

Table 1: In the table below we display the results for estimating the probabilities $\hat{p}_b$ for a variety of the values of b. The random variables B_i have symmetric two sided Pareto distributions with $\alpha = 2$, the A_i have exponential distribution with mean $1/2$, we set $n = 25$, and all results use 10^6 Monte Carlo replications. In addition, we set $a = 0.95$, $d_j = 0.26$, and chose p_{j+1} via (10).

b	Estimate	Std. Error	Rel. Error	Comp. Time
1e+09	1.00e-18	5.22e-21	0.0052	21.66
1e+12	1.01e-24	9.88e-27	0.0098	21.88
1e+15	9.98e-31	4.18e-33	0.0042	21.61
1e+18	9.97e-37	6.56e-39	0.0066	21.57
1e+21	1.01e-42	1.10e-44	0.011	21.56

REFERENCES

Asmussen, S. 2000. *Ruin probabilities*. London, UK: World Scientific.

Asmussen, S., and K. Binswanger. 1997. "Simulation of Ruin probabilities for Subexponential Claims". *ASTIN Bulletin* 27:297–318.

Asmussen, S., K. Binswanger, and B. Hojgaard. 2000. "Rare events simulation for heavy-tailed distributions". *Bernoulli* 6:303–322.

Bingham, N., C. Goldie, and J. Teugels. 1987. *Regular variation*. New York: Cambridge University Press.

Blanchet, J., and P. Glynn. 2008. "Efficient rare-event simulation for the maximum of a heavy-tailed random walk". *Ann. of Appl. Probab.* 18:1351–1378.

Blanchet, J., P. Glynn, and J. C. Liu. 2007. "Fluid heuristics, Lyapunov bounds and efficient importance sampling for a heavy-tailed G/G/1 queue". *QUESTA* 57:99–113.

Blanchet, J., H. Hult, and K. Leder. 2011. "Strongly efficient sampling strategies for affine regularly varying Markov chains". *In preparation*.

Dupuis, P., K. Leder, and H. Wang. 2006. "Importance sampling for sums of random variables with regularly varying tails". *ACM TOMACS* 17.

Dupuis, P., and H. Wang. 2007. "Subsolutions of an Isaacs equation and efficient schemes of importance sampling". *Mathematics of Operations Research* 32:723–757.

Dupuis, P., and H. Wang. 2009. "Importance Sampling for Jackson Networks". *Queueing Systems: Theory and Applications* 62:113–157.

Glasserman, P., and S. Kou. 1995. "Analysis of an importance sampling estimator for tandem queues". *ACM TOMACS* 5:22–42.

Glasserman, P., and Y. Wang. 1997. "Counter examples in importance sampling for large deviations probabilities". *Annals of Applied Probability* 7:731–746.

Konstantinides, D., and T. Mikosch. 2005. "Large Deviations and Ruin Probabilities for Solutions to Stochastic Recurrence Equations with heavy-tailed innovations". *Annals of Applied Probability* 33:1992–2035.

AUTHOR BIOGRAPHIES

JOSE BLANCHET is a faculty member of the IEOR Department at Columbia University. Jose holds a Ph.D. in Management Science and Engineering from Stanford University. Prior to joining Columbia he was a faculty member in the Statistics Department at Harvard University. Jose is a recipient of the 2009 Best Publication Award given by the INFORMS Applied Probability Society and of the 2010 Erlang Prize. He also received a PECASE award given by NSF in 2010. He worked as an analyst in Protego Financial Advisors, a leading investment bank in Mexico. He has research interests in applied probability and Monte Carlo methods. He serves in the editorial board of Advances in Applied Probability, Journal of Applied Probability, Mathematics of Operations Research and QUESTA. His email address is jose.blanchet@columbia.edu.

HENRIK HULT is Associate Professor in the Department of Mathematics at the Royal Institute of Technology in Stockholm. He received his PhD in Mathematical Statistics from the Royal Institute of Technology and has post doc experience from the University of Copenhagen and Cornell University. He also held an Assistant Professorship at Brown University. His research interests are in heavy-tailed stochastic processes, efficiency analysis of simulation algorithms, and mathematical finance. His e-mail address is hult@kth.se.

KEVIN LEDER is an assistant professor in the ISyE program at the Univeristy of Minnesota Twin-Cities. Prior to joining the University of Minnesota he was a post-doctoral researcher in the Department of Biostatistics at Harvard University, and a post-doctoral researcher in the IEOR department at Columbia University. He received his bachelors degree in applied mathematics from University of Colorado, and

his PhD from the Division of Applied Mathematics at Brown University. His research interests include stochastic modelling and simulation, with a particular interest in cancer modelling. His email address is kevin.leder@isye.umn.edu