RARE-EVENT SIMULATION FOR A MULTIDIMENSIONAL RANDOM WALK WITH t DISTRIBUTED INCREMENTS

Jose H. Blanchet

Department of Statistics
Harvard University
Cambridge, MA 02138, U.S.A.

Jingchen Liu

Department of Statistics
Harvard University
Cambridge, MA 02138, U.S.A.

ABSTRACT

We consider the problem of efficient estimation of first passage time probabilities for a multidimensional random walk with t distributed increments, via simulation. In addition of being a natural generalization of the problem of computing ruin probabilities in insurance – in which the focus is a one dimensional random walk – this problem captures important features of large deviations for multidimensional heavy-tailed processes (such as the role played by the mean of the random walk in connection to the spatial location of the target set). We develop a state-dependent importance sampling estimator for this class of multidimensional problems. Then, we argue – using techniques based on Lyapunov type inequalities – that our estimator is strongly efficient.

1 INTRODUCTION

Estimation of large deviations probabilities for heavy-tailed processes is a known challenging problem. The focus in the literature has been dominated by the analysis of one dimensional random walks. Undoubtedly, however, a general theory of rare-event analysis for heavy-tailed processes should also deal with random walks in several dimensions. Indeed, one does not need to go far to see why is this the case – just note that a large class of queueing models and stochastic networks can be related to multidimensional random walks with constrained boundaries. In several dimensions, however, even a convenient definition of a heavy-tailed random variable, that is both rich enough to capture stylized features observed in applied models and that allows the development of a theory that parallels that of one dimensional random walks seems to be difficult to put forward. Recent advances in large deviations theory for regularly varying processes, see for instance Hult et al (2005), suggest that a promising development of general theory for large deviations in an important heavy-tailed environment, namely, that of regularly varying distributions. Nevertheless, despite

the new possibilities that this development has opened up, such theory does not seem to be naturally coupled with efficient rare-event simulation algorithms (as is the case of the well-developed large deviations theory for light-tailed systems, in which often exponential changes-of-measure suggest efficient importance sampling schemes).

Motivated by the need of exploring the connections between efficient importance sampling algorithms and the development of large deviations results for heavy-tailed systems in several dimensions, we study a stylized rare-event simulation problem for a multidimensional process. In particular, we consider a multidimensional random walk with t distributed increments and address the question of efficient estimation of first-passage time probabilities to reach a polyhedron, via simulation. This problem is the natural generalization of the classical problem of estimating the tail of the maximum of a one dimensional random walk – a quantity that yields the ruin probability of an insurance company that follows a renewal risk process. The one dimensional case has been substantially studied in the literature; see, for instance, the text of Asmussen (2003) and references therein for a detailed account problem and its connections to insurance and queueing. Hult and Lindskog (2006) argue that calculating first-passage time probabilities for multidimensional random walks correspond to computing ruin probabilities for insurance companies with several lines of business. In order to simplify the technical aspects of the discussion, in this paper we decided to deal with the case of t distributed increments and polyhedra. However, basically all of the ideas extend to multidimensional regularly varying increments and more general sets (that do not even need to be connected). Such extensions are given in Blanchet and Liu (2007).

In this paper, we illustrate a strategy that can be used both in the design of efficient rare-event simulation algorithms for heavy-tailed systems and also in the development of asymptotic upper bounds for large deviations. In order to achieve this, we propose using a parametric family of importance samplers based on mixtures – the precise form

of the mixture is given in Section 4. The mixture idea has also been used in the rare-event simulation literature for light-tailed systems (see, for instance, Sadowsky and Bucklew (1990) and, more recently, Glasserman and Juneja (2007)). In the heavy-tailed case, this idea was introduced by Dupuis, Leder and Wang (2006) for a geometric sum of one dimensional iid regularly varying variables. More recently, Blanchet, Glynn and Liu (2007a, 2007b) have exploited this idea more systematically and in more general environments in which rare events are caused by several heavy-tailed jumps in sequence (rather than just one). Now, a remarkable feature to be emphasized in the current work (and also in the cited work related to heavy-tailed case) is that the mixture parameters are state-dependent. As a consequence, since the suggested change-of-measure is state-dependent, the efficiency analysis of the algorithm is not completely direct. Fortunately, recently developed techniques based on Lyapunov inequalities for Markov processes (see, for instance, Blanchet and Glynn (2007) or Blanchet, Liu and Glynn (2007a)) come into play, allowing us to bound the second moment of the estimator and conclude the strong efficiency of our algorithm. As we shall see, solving a Lyapunov inequality involves finding a function (that we refer to as Lyapunov function) which satisfies a system of linear inequalities. The use of Lyapunov inequalities in the analysis of state-dependent rare-event simulation algorithms was introduced by Blanchet and Glynn (2007) in the context of a one dimensional first-passage time problem for heavy-tailed random walks (not necessarily regularly varying, but also including more general tails such as Weibull and lognormal among many other types). Blanchet and Glynn (2007) proposed the use of a well-known approximation for such first-passage time probability and develop the algorithm using such approximation directly in order to construct their importance sampling estimator. An important difference between Blanchet and Glynn's approach and our development here is that we do not make direct use of the approximations for the construction of our estimator, but instead propose a specific parametric family based on mixtures. In fact, only an asymptotic lower bound for the probability of interest is required because verifying strong efficiency in our procedure automatically yields the asymptotic upper bound.

Our strategy consists of proposing a family of changes-of-measure that captures, at an intuitive level, the qualitative behavior induced by the zero-variance importance sampler. Such changes-of-measure are parameterized by few constants. In particular, at each step, we consider a mixture between a large increment that makes the random walk hit the target set with relatively high probability and an increment that follows the nominal (original) distribution. The mixture probability must be chosen depending on the current position of the random walk. In order to properly choose the mixture probability, one needs to make sure that

the Lyapunov inequality is satisfied. Now, such inequality requires the choice of a convenient Lyapunov function (i.e. the solution to a system of linear equations), which is obtained using heuristics based on a so-called fluid analysis – a standard technique in heavy-tailed approximations. The heuristics then are made rigorous by going through the verification of the Lyapunov inequality, which involves tuning parameters such as the mixture probabilities. Once the inequality is rigorously verified, we are able to find an upper bound that controls the behavior of the second moment of the estimator and therefore conclude strong efficiency. Now, on the side of the asymptotics, the bound on the second moment of the importance sampling estimator can be translated, by means of Jensen's inequality, to an asymptotic upper bound on the first-passage time probability of interest. Since, upper bounds in large deviations analysis are often difficult to obtain, the techniques explained in Section 4 are of interest in asymptotic analysis in general.

The rest of the paper is organized as follows. In Section 2 we discuss basic concepts involving state-dependent importance sampling and efficiency in rare-event simulation. Section 3 describes the specific problem formulation and discussed basic results on large deviations. The analysis of the algorithm and numerical experiments are given in Section 4.

2 STATE-DEPENDENT IMPORTANCE SAMPLING AND STRONG EFFICIENCY

We shall design our estimator using state-dependent importance sampling (see, for instance, Glynn and Iglehart (1989) for more on importance sampling for Markov processes). Let $W = (W_n : n \geq 0)$ be a Markov chain, living in a space $\mathcal{X}$ endowed with a σ-field $\mathcal{F}_{\mathcal{X}}$, and with transition kernel $(K(x,A) : x \in \mathcal{X}, A \in \mathcal{F}_{\mathcal{X}})$. A state-dependent importance sampler is described by a probability transition kernel $K_q(\cdot)$ of the form

$$K_q(x,dy) = r(x,y)^{-1} K(x,dy),$$

where $r(\cdot)$ is normalized so that $K_q(\cdot)$ is a well defined Markov transition kernel. In the particular setting that is the focus of this paper, W is a random walk and

$$K(x,dy) = f(x-y)\,dy, \quad K_q(x,dy) = r(x,y)^{-1} f(x-y)\,dy,$$

where $f(\cdot)$ is the density of a multivariate t distribution and $r(\cdot)$ shall be picked in the next section involving a mixture. In particular, we will choose $r(\cdot)^{-1}$ to be the sum of two indicator functions. In addition, we use the symbols P_w^Q (E_w^Q) to denote the probability measure (expectation operator) induced by $K_q(\cdot)$ on the path space of W given

that $W_0 = w$. Similarly, for the probability induced by $K(\cdot)$ we use P_w (and E_w).

Throughout the rest of the paper we shall write T_{A_b} to denote the first-passage time of the underlying chain (in this case W) to the set A_b. More precisely, $T_{A_b} = \inf\{n > 0 : W_n \in A_b\}$ – the subscript b is the so-called rarity parameter which eventually will be sent to infinity. Consider the problem of estimating efficiently via simulation

$$u_b(w) = P_w\left(T_{A_b} < \infty\right),$$

where we assume that $u_b(w) \searrow 0$ as $b \nearrow \infty$. An unbiased estimator of $u_b(w)$ is given by

$$Z(b) = I\left(T_{A_b} < \infty\right) \prod_{i=0}^{T_{A_b}-1} r(W_i, W_{i+1}),$$

where W in the previous expression follows the law P_w^Q – which in particular yields, $E_w^Q Z(b) = u_b(w)$.

Ultimately, we are interested in selecting $r(\cdot)$ in order to achieve good complexity properties. In particular, we shall select $r(\cdot)$ in order to achieve strong efficiency, which means that

$$\sup_{b \geq 1} \frac{E_w^Q\left(Z(b)^2\right)}{u_b(w)^2} \leq \lambda$$

For some $\lambda \in (0, \infty)$. The number of replications required to achieve a good relative precision for a strongly efficient estimator is unsensitive to how small is $u_b(w)$. In particular, if we generate n iid copies $(Z_j(b) : 1 \leq j \leq n)$ of $Z(b)$ and consider

$$Y_n(b) = \frac{1}{n} \sum_{j=1}^{n} Z_j(b),$$

then, using Chebyshev's inequality we obtain (for all $b \geq 1$)

$$P\left(|Y_n(b) - u_b(w)| \geq \varepsilon u_b(w)\right) \leq \frac{\lambda}{\varepsilon^2 n}.$$

This implies, as we indicated, that in order to achieve relative precision ε with probability at least $1 - \delta$, we require $O\left(\varepsilon^{-2}\delta^{-1}\right)$ replications uniformly as $u_b(w) \searrow 0$. Now, strong efficiency is not enough to characterize the complexity of an algorithm. In particular, we need also to consider the computational cost per replication of the $Z(b)$. This is particularly important in our context because variate generation of multidimensional increments is not in general straightforward. As we shall see, in our current situation $r(\cdot)$ can be chosen so that the variate generation of the increments under the importance sampler can be done easily.

The following proposition, proved in Blanchet and Glynn (2007) provides means to evaluate the second moment of state-dependent importance sampling estimators.

Proposition 1   *Suppose that there exists a non-negative function $(g_b(w) : w \in \mathcal{X})$ such that*

$$E_w^Q r(w, W_1)^2 g_b(W_1) = E_w r(w, W_1) g_b(W_1) \leq g_b(w) \quad (1)$$

for $w \in A_b^c$ and $g_b(w) \geq \varepsilon$ for $w \in A_b$. Then,

$$E_w^Q\left(Z(b)^2\right) \leq \varepsilon^{-1} g_b(w)$$

if $w \in A_b^c$.

Remark 1 The bound (1) is called a Lyapunov inequality and $g_b(\cdot)$ is the corresponding Lyapunov function. Ultimately, the design and performance analysis of the proposed estimator boils down to finding a solution $(g_b(w) : w \in \mathcal{X})$ to the Lyapunov inequality. It is not difficult to see that if one chooses the zero-variance importance sampler, then $(u_b(w)^2 : w \in \mathcal{X})$ is a feasible Lyapunov function. Since we expect to select $r(\cdot)$ in order to mimic the behavior of the zero-variance importance sampler, then it is natural to use a guess for $u_b(\cdot)^2$ as guidance for constructing $g_b(\cdot)$. Such approach is pursued in Section 4.

3   PROBLEM SETTING AND INTUITION BEHIND LARGE DEVIATIONS IDEAS

Let us start by describing our problem setting in precise terms. Consider a d-dimensional random walk with increment density proportional to $(1 + (x - \widetilde{\eta})^T \Sigma (x - \widetilde{\eta}))^{-(v+d)/2}$, where $v > 2$, Σ is a symmetric positive definite matrix and $\widetilde{\eta}$ is a non-zero vector – such density is a general multivariate t density. We are interested in estimating via simulation the probability that the corresponding random walk hits a rare set say $\widetilde{A}$ that we parameterize as a convex polyhedron (i.e. the intersection of finitely many hyperplanes). Note that by applying G (with $GG^T = \Sigma$) both to the random walk and to the rare set, we can focus on standard t distributed random vectors. More precisely, it suffices to consider a sequence of iid d-dimensional random vectors $(X_k : k \geq 1)$ with density

$$f_{X_1}(x_1, ..., x_d) = \frac{\kappa_d}{\left(1 + \frac{1}{v}(x - \eta)^T (x - \eta)\right)^{(v+d)/2}},$$

where $\eta \neq 0$ and

$$\kappa_d = \frac{\Gamma((v+1)/2)}{\Gamma(v/2) v^{d/2} \pi^{d/2}}$$

Let $S_n = X_1 + ... + X_n$.

Our objective is to estimate via simulation.

$$u_b(0) = P_0(T_{bA} < \infty)$$

as $b \nearrow \infty$, where $A = \{x \in R^d : Cx \leq \beta\}$ for some matrix C and $bA = \{y : y = bx, \ x \in A\}$. In order to make sure that $u_b(0) \searrow 0$ as $b \nearrow \infty$, we shall impose some regularly conditions to the set A.

The conditions under which $u_b(0) \searrow 0$ as $b \nearrow \infty$ and the rate at which this occurs are delicate questions – specially in the context of general regularly varying multivariate random vectors. Sufficient conditions have been developed in the work of Hult et al (2005) and Hult and Lindskog (2006). We shall adopt those conditions in our development here. In order to describe our assumptions, we first need to compute the associated limiting regularly varying measure for the X_k's. Such measure, which we denote by $\mu(\cdot)$, is defined on the Borel sets that do not contain zero and it satisfies

$$\mu(A) = \lim_{t \longrightarrow \infty} t^{-\upsilon} P(X_1 \in tA) = \int_A \frac{\kappa_d}{(y^T y)^{(\upsilon+d)/2}} dy.$$

μ is not necessarily a probability measure and has the property that $\mu(tA) = t^{-\upsilon}\mu(A)$. Also, note that $P(\|X_1\|_2 > t) = ct^{-\upsilon}(1 + o(1))$ as $t \nearrow \infty$ for some $c > 0$. The measure $\mu(\cdot)$ plays an important role in the large deviations behavior of $S = (S_n : n \geq 0)$. For instance, if the increments of the random walk had d independent coordinates, the measure $\mu(\cdot)$ would have been concentrated along the axis, which would imply that extreme behavior is caused by individual jumps given by single coordinates (i.e. the possibility of simultaneous joint large jumps is relatively small). The fact that $\mu(\cdot)$ has a density with respect to the Lebesgue measure in $R^d \setminus \{0\}$, in particular implies that when a coordinate exhibits a large jump, then the rest of the coordinates will also tend to be large. Such feature is one of the reasons for which t-copulas are often applied when modeling extreme dependence, see for instance, in credit risk (see, Embrechts et al (2001)).

Now, given that the random walk has drift $EX_k = \eta$, it is not difficult to see geometrically that some conditions must be imposed to A in order to have a meaningful rare-event situation. In order to rule out non rare-event situations, we should assume at least that the set A does not intersect the ray $R_\eta = \{t\eta : t > 0\}$. However, little thought reveals that this condition is not strong enough to rule out degenerate situations. In particular, two situations can occur, in one hand, it could be the case that A lives in a lower dimensional manifold (like a line in R^d for $d > 1$), in which case the random walk will never A. The second situation arises, for instance, when one of the faces of the polyhedron is parallel to the η, in which case, Central Limit Theorem fluctuations would eventually make the random walk hit the target set.
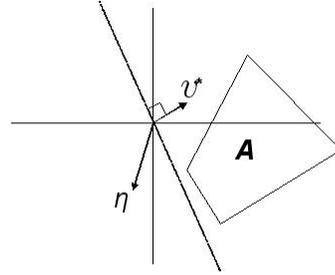
Figure 1: Diagram illustrating Assumption A for a two dimensional random walk.

In order to rule out these two cases we impose the following assumption.

A) Assume that $\mu(A) > 0$ and, in addition, suppose that there exists $v^* \in R^d$, $\delta > 0$ such that $\|v^*\|_2 = 1$, $\eta^T v^* < -\delta$ and $z^T v^* > \delta$ for all $z \in A$. See Figure 1.

Now, we are ready to provide some estimates on the rate of convergence to zero of $u_b(0)$. Since $\mu(A) > 0$ then the large deviations behavior of the system will be governed by a single large jump that makes the random walk eventually reach the target set. Let us borrow some of the intuition from the one dimensional case. In such case, we know that the random walk evolves according to its nominal (original) dynamics for $O(b)$ steps until a large jump occurs that causes the random walk to reach a large threshold b. See Hult et al (2005) for a large deviation result. Using this intuition, we propose the following lower bound for our multidimensional problem.

Proposition 2 *There exists a constant $c > 0$ such that*

$$u_b(0) \geq P_0(S_{\lfloor b \rfloor} \in bA) \geq cbP(\|X\|_2 > b)$$

as $b \nearrow \infty$

Proof. The first inequality from left to right is immediate. Now, it follows using regular variation and the definition of v^* that there exists $\beta_1 \in (0, 1)$ and $\beta_2 > 0$

$$P_0(S_{\lfloor b \rfloor} \in bA) \geq \beta_1 P_0(v^{*T} S_{\lfloor b \rfloor} > \beta_2 b).$$

Therefore, the result follows from properties of one dimensional regularly varying random variables (see, Rozovskii (1989)). ∎

The most interesting portion of the large deviation estimate for $u_b(0)$ involves developing an upper bound. The strategy that we shall pursue in the next section allows to obtain the desired upper bound by showing that the second moment of a suitable importance sampling estimator is of order $O\left(b^2 P(X_1 \in bA)^2\right)$ as $b \nearrow \infty$. This will imply both strong efficiency and also that $u_b(0) = O(bP(X_1 \in bA))$ as $b \nearrow \infty$. It should be noted that the constant multiplying the

previous "big-O" asymptotic for $u_b(0)$ can be computed explicitly (see Hult et al (2005)). However, such information is not necessary for showing strong efficiency of the estimator.

4   ALGORITHM AND ANALYSIS

As we discussed in the Introduction, our strategy involves the use of a parametric family of changes-of-measure that mimics the behavior of the zero-variance importance sampler – which corresponds to the conditional distribution of the random walk given that the rare-event occurs. Now, as we indicated before, in this case the rare-event is caused at some time by a single large jump while keeping – roughly speaking – all the increments prior to that time evolving according to the nominal dynamics. This suggests using as a family of importance sampling distributions for the increments a mixture of two components: one that induces a large jump that reaches the target set with relatively high probability and another one that basically involves the corresponding nominal dynamics. More precisely, given that the current position of the walk is s and that we have not reached the target set bA, we propose sampling the next increment according to a mixture density of the form

$$
\begin{aligned}
q_{X|s}(x) \quad = \quad & p(s)\frac{f_X(x)I\left(x \in B_{s,bA}\right)}{P\left(X \in B_{s,bA}\right)} \\
+ \quad & (1-p(s))\frac{f_X(x)I\left(x \in B_{s,bA}^c\right)}{P\left(X \in B_{s,bA}^c\right)},
\end{aligned}
\tag{2}
$$

where the mixture probability $p(s)$ is allowed to depend on the current state and the set $B_{s,bA}$ is closely related (in a precise way to be described later) to the translated set $bA - s$. Now, we also know from the literature on rare-event simulation for heavy-tailed systems that in order to control the variance of the estimator is important to recognize the contribution of such sample paths that achieve the rare-event due to more than one large jump. Therefore, to account for the contribution of such sample paths, we allow the set $B_{s,bA}$ to contain $bA - s$ in a suitable way. The construction of the set $B_{s,bA}$ pays attention not only to the contribution of such sample paths that take more than one large jump to reach the target set but also to the ability of the simulator to implement the path generation under the importance sampling distribution. As a consequence, we define the set $B_{s,bA}$ as follows. First let $\delta^* = \sup\{\delta > 0 : z^T v^* > \delta$ for all $z \in A\}$ (where v^* is defined in Assumption A) and put $A^* = \{x \in R^d : x^T v^* \geq \delta^*\}$. Then, pick $a \in (0,1)$ and finally

define

$$
B_{s,bA} = \left\{x \in R^d : \|x\|_2 > a\inf_{y \in A^*}\|by - s\|_2\,,\ x^T v^* \geq 0\right\}.
\tag{3}
$$

The likelihood ratio corresponding to (2) takes the form (using the notation of Proposition 1)

$$
\begin{aligned}
\frac{q_{X|s}(x)}{f_X(x)} \quad &\equiv \quad r(s, x+s) \tag{4}\\
&= \quad \frac{P\left(X_1 \in B_{s,bA}\right)}{p(s)}I\left(x \in B_{s,bA}\right)\\
&\quad + \frac{P\left(X_1 \in B_{a,s,bA}^c\right)}{1-p(s)}I\left(x \in B_{a,s,bA}^c\right).
\end{aligned}
$$

Finally, the corresponding (unbiased) estimator for $u_b(0)$ is

$$
Z_b = \prod_{k=0}^{T_{b\cdot A}-1} r(S_k, S_{k+1})I(T_{b\cdot A} < \infty)
$$

(that Z_b is unbiased is verified in Blanchet and Glynn (2007)). Once we have proposed a suitable parametric family of importance sampling distributions, we need to tune the mixture parameter in order to satisfy the Lyapunov inequality. For this matter, we also need to propose a parametric expression for the candidate Lyapunov function. As we discussed in the remark following Proposition 1, if the proposed importance sampler is close enough to the zero-variance change-of-measure, then we expect the Lyapunov inequality to be satisfied by a function that behaves like (or is an upper bound for) $(P_s^2(T_{bA} < \infty) : s \in R^d)$. Therefore, a natural strategy is to obtain a heuristic guess $P_s(T_{bA} < \infty)$ and use this guess to propose an explicit form for the Lyapunov function which then will be tested rigorously. Note that $P_s(T_{bA} < \infty) \leq P_s(T_{bA^*} < \infty)$, so trying to test a Lyapunov function that approximates $\left(P_s^2(T_{bA^*} < \infty) : s \in R^d\right)$ seems adequate for obtaining an upper bound. Now, a rough analysis suggests

$$
P_s(T_{bA^*} < \infty) \approx \int_0^\infty P(X_1 + s + \eta t \in bA^*)\,dt.
$$

This type of heuristic development is often called a "fluid analysis". The idea is to replace the dynamics of the random walk by the motion suggested using the Law of Large Numbers and also recognize that the rare event is caused by a single large jump. We then define

$$
h_b(s) \equiv \int_0^\infty P(X_1 + s + \eta t \in bA^*)\,dt.
$$

Note that

$$
\begin{aligned}
h_b(s) &= \int_0^\infty P\left(X_1^T v^* + s^T v^* + \eta^T v^* t \ge \delta^* b\right) dt \\
&= \frac{1}{-\eta^T v^*} \int_0^\infty P\left(X_1^T v^* \ge \delta^* b - s^T v^* - \eta^T v^* t\right) \\
&\qquad\qquad d(-\eta^T v^* t) \\
&= \frac{1}{-\eta^T v^*} \int_{\delta^* b - s^T v^*}^\infty P\left(X_1^T v^* \ge u\right) du \\
&= O\left(P(\|X\|_2 > b)\, b\right)
\end{aligned}
$$

as $b \nearrow \infty$. The last equality can be justified by that fact that $X_1^T v^*$ is distributed as t distribution with degree of freedom υ and $P(X_1^T v^* > u) \sim P(\|X_1\| > u)$. Our candidate for Lyapunov function then takes the form

$$
g_b(s) = \min\left(c_g \cdot h_b(s)^2, 1\right) \tag{5}
$$

for some constant $c_g > 0$ (to be determined over the course of the verification of the Lyapunov inequality).

Verifying the Lyapunov inequality involves checking, for all $s \in \mathbb{R}$,

$$
\begin{aligned}
1 &\ge E^q\left(\frac{g_b(s+X)}{g_b(s)} r(s, s+X)^2\right) \\
&= E\left(\frac{g_b(s+X)}{g_b(s)} r(s, s+X)\right) \tag{6} \\
&= \frac{1}{p(s)} E\left(\frac{g_b(s+X)}{g_b(s)}; X \in B_{s,bA}\right) \\
&\quad + \frac{1}{1-p(s)} E\left(\frac{g_b(s+X)}{g_b(s)}; X \in B_{s,bA}^c\right),
\end{aligned}
$$

(where $E^q(\cdot)$ represents the expectation induced by the density $q_{X|s}(\cdot)$). Let us define

$$
J_1 = E\left(\frac{g_b(s+X)}{g_b(s)}; X \in B_{s,bA}\right) \frac{P(X \in B_{s,bA})}{p(s)},
$$

$$
J_2 = E\left(\frac{g_b(s+X)}{g_b(s)}; X \in B_{s,bA}^c\right) \frac{P(X \in B_{s,bA}^c)}{1-p(s)}.
$$

Assume that $g_b(s) < 1$ (note that if $g_b(s) \ge 1$ we can just not apply importance sampling at all and the Lyapunov inequality will be automatically satisfied). An immediate upper bound is obtained for J_1, namely,

$$
J_1 \le E\left(\frac{1}{g_b(s)} \middle| X \in B_{s,bA}\right) \frac{P^2(X \in B_{s,bA})}{p(s)} \le \frac{P^2(X \in B_{s,bA})}{c_g h_b(s)^2 p(s)}. \tag{7}
$$
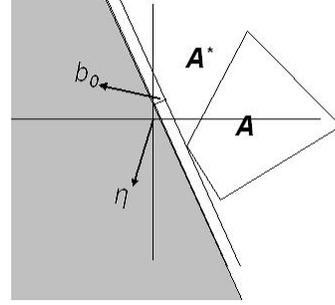
Figure 2: Two dimensional diagram illustrating the region where Lemma 1 is satisfied.

For J_2 we obtain

$$
J_2 \le \frac{1}{1-p(s)} E\left(\frac{g_b(s+X)}{g_b(s)}; X \in B_{s,bA}^c\right). \tag{8}
$$

The idea is then to take advantage of Taylor expansion. Since

$$
\begin{aligned}
\eta^T \nabla h_b(s) &= \frac{-1}{\eta^T v^*} P\left(X^T v^* > \delta b^* - s^T v^*\right) \eta^T v^* \\
&= -P\left(X^T v^* > \delta b^* - s^T v^*\right) = -P(X+s \in bA^*),
\end{aligned}
$$

then (if $g_b(s) < 1$)

$$
\frac{\eta^T \nabla g_b(s)}{g_b(s)} = -2 \frac{P(X+s \in bA^*)}{h_b(s)}.
$$

The previous expression can then be used to handle the contribution of the expectation in the right hand side of inequality (8). The complete details are given in Blanchet and Liu (2007), where the following result is proved.

Lemma 1 *Define $D_{bA^*}(z) = \{y : \inf_{y \in bA^*} \|s - y\|_2 \ge z\}$. Then, there exist computable constants $\gamma_1 \in (0, 1)$ and $b_0 > 0$ such that*

$$
\begin{aligned}
E\left(\frac{X^T \nabla g_b(s+UX)}{g_b(s)}; X \in B_{a,bA}\right) &\le \frac{\gamma_1}{2} \frac{\eta^T \nabla g_b(s)}{g_b(s)} \\
&= -\gamma_1 \frac{P(X+s \in bA^*)}{h_b(s)}
\end{aligned}
$$

for all $s \in D_{bA^*}(b_0)$

Remark 2 The gray area in Figure 2 shows the form of the region where the previous inequality holds. The constant c_g eventually is selected large enough so that $g(s) = 1$ on the complement of the gray area (so that finally, the Lyapunov inequality is satisfied throughout the whole space).

The previous lemma combined with (7) and

$$\frac{g_b(s+X)}{g_b(s)} = \frac{g_b(s) + \int_0^1 X^T \nabla g_b(s+uX) du}{g_b(s)}$$

yields

$$J_1 + J_2 \leq \frac{1}{1-p(s)} + \frac{P(X \in B_{s,bA})^2}{c_g h_b(s)^2 p(s)} - \gamma_1 \frac{P(X+s \in bA^*)}{h_b(s)}.$$

Therefore, if we select (for some $\rho_2 > 0$)

$$p(s) = \min\left(\rho_2 \frac{P(X \in B_{s,bA})}{h_b(s)}, 1/2\right), \qquad (9)$$

we conclude that

$$J_1 + J_2 \leq 1 + 2\rho_2 \frac{P(X \in B_{s,bA})}{h_b(s)}$$
$$+ \frac{P(X \in B_{s,bA})}{c_g \rho_2 h_b(s)} - \gamma_1 \frac{P(X+s \in bA^*)}{h_b(s)}. \qquad (10)$$

Now, our job is to appropriately select ρ_2 and c_g in order to make the right hand side of the previous inequality less than 1. The next technical result allows to conclude that ρ_2 and c_g can indeed be appropriately selected (the proof is an application of standard regularly varying properties, the details are given in Blanchet and Liu (2007)).

Lemma 2   *One can compute $\gamma_2 \in (0, \infty)$ such that*

$$P(X \in B_{s,bA}) \leq \gamma_2 P(X+s \in bA^*)$$

for all $s \in R^d$.

Applying the previous result to (10) we conclude that

$$J_1 + J_2 \leq 1 + 2\rho_2\gamma_2 \frac{P(X+s \in bA^*)}{h_b(s)}$$
$$+ \gamma_2 \frac{P(X+s \in bA^*)}{c_g \rho_2 h_b(s)} - \gamma_1 \frac{P(X+s \in bA^*)}{h_b(s)}. \quad (11)$$

Therefore, the Lyapunov inequality is satisfied if we select ρ_2 and c_g such that

$$2\rho_2\gamma_2 + \frac{\gamma_2}{c_g \rho_2} - \gamma_1 \leq 0. \qquad (12)$$

This inequality can be achieved by first selecting ρ_2 small (say $\rho_2 \leq \gamma_1/(4\gamma_2)$) and then c_g large enough (say $c_g \geq 4\gamma_2^2/\gamma_1$). As indicated in Remark 2, in addition to this constraint, c_g is selected large enough so that the region $\{s : g(s) = 1\}$ contains the white area displayed in Figure 2.

We conclude with a summary of the proposed algorithm (for the generation of a single replication of Z_b).

Algorithm

Set $b > 0$ and fix $a \in (0,1)$. Initialize $w = 0, REACH = 0$ and $Z = 1$. Assume that c_g and ρ_2 have been selected in order to satisfy (12) and set $g_b(w)$ according to (5).

STEP 1

- If $g_b(s) = 1$ then sample X according to the nominal distribution.

  Else set $p \longleftarrow \min\left(\rho_2 \frac{P(X \in B_{s,bA})}{h_b(s)}, 1/2\right)$ and sample X as follows. With probability p generate X with distribution $(X|X \in B_{s,bA})$, with probability $1 - p$ sample X with distribution $(X|X \in B_{s,bA}^c)$. Then, update

$$Z \longleftarrow Z \cdot [p^{-1} P(X \in B_{s,bA}) I(X \in B_{s,bA})$$
$$+ (1-p)^{-1} P(X \in B_{s,bA}^c) I(X \in B_{s,bA}^c)].$$

  Endif.
  Update $s \longleftarrow s + X$,
- If $w \in bA$ then $REACH \longleftarrow 1$. Endif.
- Repeat **STEP 1** until $REACH = 1$.

STEP 2 RETURN Z.

Before summarizing the efficiency properties of the estimator proposed by the previous algorithm, let us provide a graphical explanation of how the algorithm evolves. Figure 3 contains, in addition to the elements illustrated in Figure 1, the construction of the set A^*. In the diagram, we show the evolution of a typical path of the algorithm in fluid scale right before a big jump (coming from the mixture component with probability $p(s)$) occurs. The next step of the random walk will be selected according to a conditional density that is supported in the gray area. Note that it is possible for the increment to jump outside the set A, in which case, the path continues following its fluid dynamics (roughly a line with gradient η) until another jump occurs and so forth.

The next result summarizes the complexity of the algorithm.

Theorem 1   *If Assumption A is in force, then the previous algorithm takes on average $O(b)$ variates to terminate. In addition, the estimator, Z, obtained at Step 2 is strongly efficient.*

Proof. From inequality (11) and Proposition 1 we conclude

$$g_b(0) \geq E\left[\prod_{k=0}^{T_{b \cdot A}-1} r(S_k, S_{k+1})^2 I(T_{bA} < \infty)\right].$$
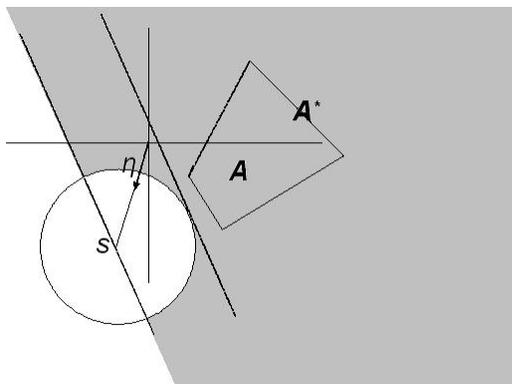
Figure 3: The evolution of a typical sample path generated by the algorithm, right before obtaining a large jump.

On the other hand, combining the lower bound obtained in Proposition 2 together with the fact that $g_b(0) = O(P(\|X\|_2 > b)^2 b^2)$ as $b \nearrow \infty$, we conclude that

$$\sup_{b \geq 1} \frac{E^q(Z^2)}{P(T_{b \cdot A} < \infty)^2} \leq \sup_{b \geq 1} \frac{g_b(0)}{cP(\|X\|_2 > b)^2 b^2} < \infty.$$

The fact that the expected number of increments required to reach bA is $O(b)$ follows again using a Lyapunov-type argument as in Blanchet and Glynn (2007). The details are given in Blanchet and Liu (2007). ∎

We close this section with some numerical experiments. We assume that $d = 2$, $\upsilon = 3$, $\Sigma = I$ and $\eta = (0, -1)^{\mathsf{T}}$. We consider target sets of the form bA, where $A = \{(x, y); x > 1, y > 1\}$ and $b = 10, 20$, and 50. Note that Assumption A is satisfied with $v^* = -\eta$. We select $A^* = \{(x, y); y \leq 1\}$ (i.e. $\delta^* = 1$), $\rho_2 = 1/3$, $c_g = 12$ and $a = 0.9$ (recall that a was defined in (3)). In order to verify the validity of our implementation we used crude Monte Carlo to estimate the probability for the case $b = 10$ based on 200,000 replications and obtained an estimate equal to $1.74e - 03$ with standard deviation $9.32e - 05$. The implementation of our algorithm yields an estimated coefficient of variation that remains very stable as b increases (around 6). In Table 1, "Est" is the average of 20000 estimates; "SD" is the standard deviation of a single run.

Table 1: Simulation results

b	Est	SD	Sample size
10	$1.83e - 03$	$1.31e - 02$	20000
20	$4.28e - 04$	$2.54e - 03$	20000
50	$8.80e - 05$	$4.74e - 04$	20000

REFERENCES

Asmussen, S. (2003) *Applied Probability and Queues*. Springer-Verlag. New York.

Blanchet, J. and Glynn, P. (2007) Efficient rare event simulation for the maximum of heavy-tailed random walks. Submitted.

Blanchet, J., Glynn, P., and Liu, J. C. (2007a) Fluid heuristics, Lyapunov bounds and efficient importance sampling for a heavy-tailed G/G/1 queue. Submitted.

Blanchet, J., Glynn, P., and Liu, J. C. (2007b) Efficient rare event simulation for multiserver queues. Preprint.

Blanchet, J. and Liu, J. C. (2007) First-passage time probabilities for multidimensional heavy-tailed random walks: Algorithms and Asymptotics. Preprint.

Dupuis, P., Leder, K., and Wang, H. (2006) Notes on importance sampling for random variables with regularly varying tails. Preprint.

Embrechts, P., Lindskog, F., and McNeal, A. (2001) Modeling dependence with copulas and applications to risk management. Technical Report, Risklab, ETH Zurich.

Glasserman, P., and Juneja, S. (2007) Uniformly Efficient Importance Sampling for the Tail Distribution of Sums of Random Variables. To appear in Math. of O.R.

Glynn, P., and Iglehart, D. (1989) Importance sampling for stochastic simulations. Management Science, vol.35, No.11, p.1367-1392.

Hult, H., Lindskog, F., Mikosch, T., and Samorodnitsky, G. (2005) Functional large deviations for multivariate regularly varying random walks. Ann. Appl. Probab. No. 15, p. 2651-2680.

Hult, H., and Lindskog, F. (2006) Heavy-tailed insurance portfolios: buffer capital and ruin probabilities. Tech. Rep. No. 1441, School of ORIE, Cornell University.

Rozovskii, L. V. (1989) Probabilities of Large Deviations of Sums of Independent Random Variables with Common Distribution Function in the Domain of Attraction of the Normal Law, Theory Probab. Appl., 34, pp. 625–644.

Sadowsky, J. S. and Bucklew, J. A. (1990) On large deviations theory and asymptotically efficient Monte Carlo estimation. IEEE Transactions on Information Theory, Vol. 36, No 3, p. 579 - 588.

AUTHOR BIOGRAPHIES

JOSE H. BLANCHET is Assistant Professor of Statistics at Harvard University. Jose holds a M.Sc. in Engineering-Economic Systems and Operations Research and a Ph.D. in Management Science and Engineering, both from Stanford University. He also holds two B.Sc. degrees: one in Actuarial Science and another one in Applied Mathematics from ITAM (Mexico). Jose worked for two years as an analyst in Protego Financial Advisors, a leading investment bank in Mexico. He has research interests in applied probability, computational finance, performance engineering, queueing theory, risk management, rare-event analysis, statistical inference, stochastic modelling, and simulation.

JINGCHEN LIU is currently a fifth year Ph.D. student in the Department of Statistics at Harvard University. Jingchen Liu graduated with BS in Mathematics from Beijing University, China.