Rare-event Simulation for Multidimensional Regularly Varying Random Walks

Jose Blanchet and Jingchen Liu

January, 2009

Abstract

We consider the problem of efficient estimation via simulation of first passage time probabilities for a multidimensional random walk with regularly varying increments. 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 process in connection to the 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 inequalities that our estimator is strongly efficient in the sense that the mean square error of our estimator can be made arbitrarily small by increasing the number of replications, uniformly as the probability of interest approaches zero. An important feature of our algorithm involves the interplay between large deviations for regularly varying processes and linear programming. When the target set is the union of half-spaces, our sampler, which can be described in terms of mixtures, can be shown to approximate in total variation the conditional distribution of the walk given that it hits the target set in finite time.

1 Introduction

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 first-passage time probabilities for multidimensional regularly varying random walks. 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 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. First one considers a parametric family of importance sampling distributions 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, which takes us to the second ingredient of the program. Recently developed techniques based on Lyapunov inequalities for Markov processes (see, for instance, Blanchet and Glynn (2007) or Blanchet, Liu and Glynn (2007a)) are applied, allowing us to bound the second moment of the estimator and conclude that the coefficient of variation of the estimator is bounded as the probability of interest decreases to zero. 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. In contrast to the work of Blanchet and Liu (2008) (CITE PAPER IN ADVANCES IN APP. PROB.), we concentrate on multidimensional problems, which give rise to additional complications derived from the interplay between the location of the target set and the drift of the walk (we discuss these issues in the statement of Assumption A) and B) in Section 4). In addition, finite termination of the algorithm becomes an important issue that requires special attention in the multidimensional case.

The mixture family of importance sampling distributions captures, at an intuitive level, the qualitative behavior induced by the zero-variance importance sampler (which is the conditional distribution of the walk given that it eventually reaches the target set of interest, see for instance Blanchet and Glynn (2007)). 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 and an increment that follows the nominal (original) distribution. The mixture probability is 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 which is obtained using heuristics based on a so-called fluid analysis – a standard technique in heavy-tailed approximations and described in Section 4, equation (6). The heuristics then are made rigorous going through the verification of the Lyapunov inequality, which involves tuning various 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. 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, asymptotic 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 bounds capture exactly the asymptotic rate of decay and sometimes can be shown to also capture the exact prefactor, as we explore in Section ???? for the case in which the target set is the union of half-spaces. In this case, as we shall also discuss in Section ????, the induced importance sampling distribution automatically gives a simple description of a Markov chain that approximates, in total variation, the conditional distribution of the process of interest (in this case a multidimensional random walk), given the rare event in question (in our current setting, that the walk hits a rare set eventually).

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 Important 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 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 this paper, W is a random walk with transition kernel,

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

where $f(\cdot)$ is the density function of the increments and $r(\cdot)$ shall be picked in the next section involving a mixture. 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 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

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

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

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

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, which we shall measure in terms of the mean squared error and the cost-per-replication under the proposed change-of-measure. First, we shall concentrate on the variance control. In particular, we shall select $r(\cdot)$ in order to achieve strong efficiency, that is,

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

The number of replications required to achieve a good relative precision for a strongly efficient estimator is insensitive to how small $u_b(w)$ is. So, 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 $b \geq 1$)

$$P(|Y_n(b) - u_b(w)| \geq \varepsilon u_b(w)) \leq \frac{\lambda}{\varepsilon^2 n}. \tag{1}$$

This implies, as we indicated, that in order to achieve relative precision ε with probability at least $1-\delta$, it suffices to generate $n = \lambda\varepsilon^{-2}\delta^{-1} = O(\varepsilon^{-2}\delta^{-1})$ replications uniformly as $u_b(w) \searrow 0$.

Now, clearly strong efficiency is not enough to characterize the complexity of an algorithm, which involves the expected time to hit A_b and the variate generation at each step. These issues shall be discussed in a future section.

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

Proposition 1 *Suppose that there exists a non-negative function* $g_b(w) : \mathcal{X} \to R^+ \cup \{0\}$*, such that*

$$E_w^Q r(w, W_1)^2 g_b(W_1) = E_w r(w, W_1) g_b(W_1) \leq g_b(w) \tag{2}$$

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 (2) is called Lyapunov inequality and $g_b(\cdot)$ is the corresponding Lyapunov function, which may not be unique. Ultimately, the design and performance analysis of the proposed estimator boils down to finding a solution $g_b(w)$ to the Lyapunov inequality. It is not difficult to see that if one chooses the zero-variance importance sampler $(r(w_0, w_1) = u_b(w_1)/u_b(w_0))$, then $u_b^2(w)$ is one Lyapunov function. Since we expect to select $r(\cdot)$ in order to mimic the behavior of the zero-variance importance sampler, it is natural to use a guess for $u_b^2(\cdot)$ as guidance for constructing $g_b(\cdot)$. Such approach is pursued in Section 4.

3 Problem setting and intuition

Let $(X_n : n \geq 1)$ be a sequence of independent and identically distributed (i.i.d.) regularly varying random vectors taking values in R^d. A random vector X_k is said to have a multivariate regularly varying distribution if there exists a Radon measure $\mu(\cdot)$ such that for any Borel set, A, that does not contain the origin we have

$$\lim_{b \longrightarrow \infty} \frac{P(X_n \in bA)}{P(\|X_n\|_2 > b)} = \mu(A)$$

as $b \nearrow \infty$. The X_n's have a relatively very small probability of jumping into sets for which $\mu(A) = 0$. Note that clearly, one can obtain $\mu(A) = \infty$ by including an appropriate ball containing the origin inside A. If $P(\|X_n\|_2 > b) = b^{-\alpha} L(b)$ for some $\alpha > 0$ and a slowly varying function $L(\cdot)$ (i.e. $L(tb)/L(b) \longrightarrow 1$ as $b \nearrow \infty$ for each $t > 0$), then we say that $\mu(\cdot)$ has (regularly varying) index α.

An alternative (equivalent) definition of a multivariate regularly varying distribution is that there exists a random variable Θ taking values on the surface of the unit sphere in d dimensions (denoted by $\mathbb{S}^{d-1}$) so that for each $x > 0$

$$\frac{P(\|X\|_2 > bx, X/\|X\|_2 \in \cdot)}{P(\|X\|_2 > b)} \longrightarrow x^{-\alpha} P(\Theta \in \cdot) \tag{3}$$

in the sense of vague convergence in $\mathbb{S}^{d-1}$ (see Kallenberg (1983)).

In order to illustrate the previous definition consider the following simple example.

Example 1: Suppose that X_1 follows a d-dimensional t-distribution with v degrees of freedom. In particular, X_1 has density

$$f_{X_1}(y) = \frac{c}{(1 + y^T y)^{(v+d)/2}}$$

for an appropriate constant $c \in (0, \infty)$. The associated limiting regularly varying measure for X_1 satisfies

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

Note that $P(\|X_1\|_2 > t) = ct^{-v}(1 + o(1))$ as $t \nearrow \infty$ for some $c > 0$. The fact that $\mu(\cdot)$ has a density with respect to the Lebesgue measure in $\mathbb{R}^d\backslash\{0\}$, implies that when a coordinate exhibits a large jump, the rest of the coordinates will also tend to be large. Such a feature is one of the reasons for which t-copulas are often applied when modeling extreme dependence, see for instance (Embrechts et al (2001)). In the context of (3) we have that Θ is uniformly distributed.

Coming back to our discussion on large deviations for multidimensional regularly varying random walks, let us assume that $EX_1 = \eta \neq 0$ and define $S_n = X_1 + ... + X_n$. Throughout the rest of the paper we shall use the notation $P_s(\cdot)$ for the probability measure on the path-space of the process $S = (S_n : n \geq 0)$ given that $S_0 = s$. Given a set A, we define $T_A = \inf\{n \geq 0 : S_n \in A\}$ and we write $bA = \{y : y = ba,\ a \in A\}$. Finally, we define

$$u_b(s) = P_s(T_{bA} < \infty).$$

Our goal is to develop a strongly efficient simulation estimator for $u_b(0)$ in a variety of settings where $u_b(0) \searrow 0$ as $b \nearrow \infty$.

Given that the random walk has drift η, it is not difficult to see geometrically that some conditions must be imposed to A in order to have a meaningful rare-event situation (i.e. $u_b(0) \searrow 0$ as $b \nearrow \infty$). One necessary condition is that the set A does not intersect the ray $\{t\eta : t > 0\}$. Otherwise, the Law of Large Numbers might eventually let the process hit the set bA. This condition, however, is not strong enough to rule out degenerate situations. For instance, suppose that A is a polyhedron, then, it could be the case that when one of the faces of the polyhedron is parallel to η in such a way that Central Limit Theorem-type fluctuations might eventually make the random walk hit the target set. In order to avoid this situation we impose two assumptions.

A) There exists linearly independent vectors $v_1^*, ..., v_m^* \in \mathbb{R}^d$, $\delta^* > 0$ such that $\|v_j^*\|_2 = 1$, $\eta^T v_j^* < -\delta^*$ and for all $z \in A$ we have $z^T v_j^* \geq \delta^*$ for at least one v_j^*.

Figure 1 depicts the situation described by A).

As indicated by the diagram, the vectors v_k^* point to the "direction" where the set A is located and such direction cannot be parallel to the drift of the process. Otherwise, the set A would be, as indicated previously, eventually attainable with probability one.

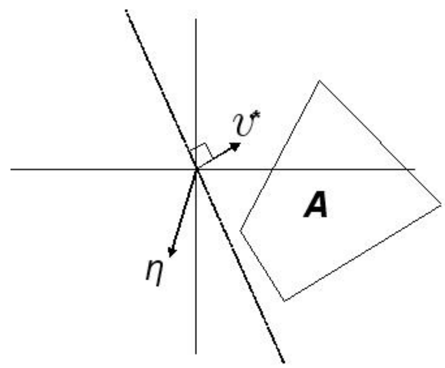


Figure 1: Diagram illustrating Assumption A or a two dimensional random walk

In order to avoid situations where the target set is “too thin”, for instance when A lives in a lower dimensional manifold such as a line, we impose the following assumption.

B) Assume that $\mu(\cdot)$ has regularly varying index $\alpha > 1$ and A contains an open subset, A°, such that $\mu(A^{\circ}) > 0$.

Assumptions A) and B) guarantee that 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. Moreover, just as the one dimensional case, 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.

Now, we are ready to provide some estimates on the rate of convergence to zero of $u_b(0)$. The following result follows as an easy consequence of. Hult et al (2005) and Hult and Lindskog (2005).

Proposition 2 *Under assumptions A) and B), there exists a constant $c > 0$ such that*

$$u_b(0) \geq cbP\left(\|X\|_2 > b\right)$$

as $b \nearrow \infty$.

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 $O\left(b^2 P\left(X_1 \in bA\right)^2\right)$ as $b \nearrow \infty$. This will imply both strong efficiency and also that $u_b(0) = O\left(bP\left(X_1 \in bA\right)\right)$ 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 an estimator.

4 Proposed Change-of-measure and Variance Analysis

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 the heavy-tailed case the rare-event is caused at sometime by a single large jump while all the increments prior to that time evolves approximately according to the nominal dynamics. This suggests a family of importance sampling distributions for the increments in the form of a mixture of two components: one that induces a large jump "in the direction of the target set" and another one that basically envolves according to the nominal/original dynamics. Let us explain more precisely in intuitive terms what we mean by "a direction to the target set". Suppose that A is a simply connected component as shown in Figure 1. Then, the vector v^* depicted in Figure 1 indicates a possible direction to the target set. A direction to be chosen is not unique. Any other vector that has positive inner product with each element of A could also serve as a possible direction. Moreover, if A has more than one connected component there maybe several directions that one can take simultaneously.

The vectors $v_1^*, ..., v_m^*$ will be used as possible directions. Our family of changes-of-measure is constructed so that with some probability, which maybe depend on the current state of the current position, the random walk reaches A in the immediate next step. More precisely, given the current position of the walk is s, and a constant $a \in (0,1)$, we define $C_i^a(s,b) = \{x : v_i^{*T} x > a\left(\delta^* b - v_i^{*T} s\right)\}$. The parameter $a \in (0,1)$ helps to over sample those sample paths that take several jumps to reach A. We propose sampling the next increment according to a mixture density of the form

$$q_{X|s}(x) = p(s) \frac{f_X(x) I(x \in \cup_{i=1}^m C_i^a(s,b))}{P(X \in \cup_{i=1}^m C_i^a(s,b))} + (1 - p(s)) f_X(x), \qquad (4)$$

where the mixture probability $p(s)$ is allowed to depend on the current state. Let us write

$$P(s) = P(X \in \cup_{i=1}^m C_i^a(s,b)).$$

Note that

$$\begin{aligned} q_{X|s}(x) &= (p(s) + (1-p(s)) P(s)) \frac{f_X(x) I(x \in \cup_{i=1}^m C_i^a(s,b))}{P(s)} \\ &\quad + (1-p(s)) f_X(x) I\left(x \in \cap_{i=1}^m \overline{C}_i^a(s,b)\right). \end{aligned}$$

We obtain that the likelihood ratio corresponding to (4) takes the form (using a notation consistent with the statement of Proposition 1)

$$\begin{aligned} \frac{f_X(x)}{q_{X|s}(x)} &\triangleq r(s, x+s) \qquad (5) \\ &= \frac{P(s)}{(p(s) + (1-p(s)) P(s))} I(x \in \cup_{i=1}^m C_i^a(s,b)) \\ &\quad + \frac{1}{(1-p(s))} I\left(x \in \cap_{i=1}^m \overline{C}_i^a(s,b)\right). \end{aligned}$$

We shall use the notation $E_s^q(\cdot)$ for the expectation operator induced by $r(\cdot)$ given that $S_0 = s$. Similarly we use $P_s^q(\cdot)$ for the probability measure associated to $E_s^q(\cdot)$. Finally, the corresponding (unbiased) estimator for $u_b(0)$ is

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

Once we have proposed a suitable parametric family of importance sampling distributions, we need to tune the mixture parameter, $p(s)$, 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) $(P_s^2(T_{bA} < \infty) : s \in \mathbb{R}^d)$. Therefore, a natural strategy is to obtain a heuristic guess for $P_s(T_{bA} < \infty)$ and use this guess to propose an explicit form for the Lyapunov function which then will be tested rigorously. A rough analysis, often called in the literature as "fluid analysis", suggests

$$\begin{aligned} P_s(T_{bA} < \infty) &\approx \int_0^\infty P(X + s + \eta t \in bA)\, dt \qquad (6) \\ &\le \sum_{i=1}^m \int_0^\infty P\left(v_i^{*T}(X + s + \eta t) \ge b\delta^*\right) dt \\ &= \sum_{i=1}^m \frac{1}{-v_i^{*T}\eta} G_i\left(b\delta^* - v_i^{*T} s\right), \end{aligned}$$

where

$$G_i\left(b\delta^* - v_i^{*T} s\right) = \int_{b\delta^* - v_i^{*T} s}^\infty P\left(v_i^{*T} X > u\right) du.$$

The idea behind the "fluid analysis" is that, prior to the occurrance of the big jump that reaches the target set, the process behaves according to the law of large numbers (or fluid dynamics). Then, a jump occurs at some time t which takes the process to the target set. The approximation is constructed adding up all possible times (over t), thereby arriving to expression (6).

So, we propose our Lyapunov function to be

$$g_b(s) = \min(c_g h_b(s)^2, 1), \qquad (7)$$

for some $c_g > 0$ and

$$h_b(s) = \sum_{i=1}^m G_i\left(b\delta^* - v_i^{*T} s\right).$$

The selection of c_g is performed in the verification argument of our Lyapunov function.

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

$$1 \ge E^q\left(\frac{g_b(s+X)}{g_b(s)} r(s, s+X)^2\right) = E\left(\frac{g_b(s+X)}{g_b(s)} r(s, s+X)\right) \qquad (8)$$

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

We shall establish bound (8) first on $\{s : g_b(s) < 1\}$. Note that if $g_b(s) = 1$, then selecting $p(s) = 0$ one immediately satisfies the Lyapunov inequality. If $g_b(s) < 1$ then

$$
\begin{aligned}
&E\left(\frac{g_b(s+X)}{g_b(s)} r(s, s+X)\right) \\
&= E\left(\frac{g_b(s+X)}{g_b(s)}; X \in \cup_{i=1}^m C_i^a(s,b)\right) \frac{P(s)}{(p(s) + (1-p(s)) P(s))} \\
&+ E\left(\frac{g_b(s+X)}{g_b(s)}; X \in \cap_{i=1}^m \overline{C}_i^\alpha(s,b)\right) \frac{1}{1-p(s)}.
\end{aligned}
$$

Let us define

$$
\begin{aligned}
J_1 &= E\left(\frac{g_b(s+X)}{g_b(s)}; X \in \cup_{i=1}^m C_i^a(s,b)\right) \frac{P(s)}{(p(s) + (1-p(s)) P(s))}, \\
J_2 &= E\left(\frac{g_b(s+X)}{g_b(s)}; X \in \cap_{i=1}^m \overline{C}_i^\alpha(s,b)\right) \frac{1}{1-p(s)}. \qquad (9)
\end{aligned}
$$

An immediate upper bound is obtained for J_1, namely,

$$
J_1 \leq \frac{P(s)^2}{c_g h_b(s)^2 p(s)}, \qquad (10)
$$

on the set that $g_b(s) < 1$. To handle J_2, the idea is to use a Taylor development with reminder, recall that if $f(\cdot)$ is absolutely continuous then

$$
\begin{aligned}
&f(s_1+\Delta_1,...,s_d+\Delta_d)-f(s_1,...,s_d)\\
&=\int_0^1 \frac{\partial f}{\partial s_1}\left(s_1+X^{(1)}u,s_2,...,s_d\right)X^{(1)}du\\
&+\int_0^1 \frac{\partial f}{\partial s_2}\left(s_1+X^{(1)},s_2+X^{(2)}u,s_3,...,s_d\right)X^{(2)}du\\
&...+\int_0^1 \frac{\partial f}{\partial s_d}\left(s_1+X^{(1)},...,s_{d-1}+X^{(d-1)},s_d+X^{(d)}u\right)X^{(2)}du.\\
&=X^{(1)}E\left(\frac{\partial f}{\partial s_1}\left(s_1+X^{(1)}U,s_2,...,s_d\right)\right)\\
&+X^{(2)}E\left(\frac{\partial f}{\partial s_2}\left(s_1+X^{(1)},s_2+X^{(2)}U,s_3,...,s_d\right)\right)\\
&...+X^{(d)}E\left(\frac{\partial f}{\partial s_2}\left(s_1+X^{(1)},s_2+X^{(2)},...,s_d+X^{(d)}U\right)\right),
\end{aligned}
$$

where $X=\left(X^{(1)},...,X^{(d)}\right)$ and U is a uniformly distributed random variable over $[0,1]$, independent of X. Applying this representation one at a time to each of the d components of $g_b(\cdot)$ we obtain (in probabilistic terms)

$$
g_b(s+X)=g_b(s)+E\left(\sum_{j=1}^{d}\frac{\partial g_b}{\partial s_j}(s+D_jX)X^{(j)}\right),
$$

where D_j denotes an appropriate diagonal matrix, in particular the $j-1$ first components in its main diagonal are ones, the j-th component in the main diagonal is U and the remaining components of the matrix are zeroes. The strategy then consists in using a bound of the form

$$
E\left(\sum_{j=1}^{d}\frac{\partial g_b}{\partial s_j}(s+D_jX)X^{(j)};X\in\cap_{i=1}^{m}\overline{C}_i^{\alpha}(s,b)\right)\leq\gamma_1\nabla g_b(s)\cdot\eta \tag{11}
$$

for some $\gamma_1>0$ and note that

$$
\frac{\nabla g_b(s)\cdot\eta}{g_b(s)}\leq-2\delta^*\frac{\sum_{i=1}^{m}P\left(v_i^{*T}X>b\delta^*-v_i^{*T}s\right)}{h_b(s)}.
$$

A key observation is that the right hand side of the previous inequality is a negative quantity which then can be used to develop inqualuity (8).

Before providing the necessary details behind our strategy, we first provide a useful lemma involving (11).

Lemma 1 *There exists $c_g > 0$ and $\gamma_1 > 0$ such that if $g_b(s) < 1$ then,*

$$E\left(\sum_{j=1}^{d}\frac{\partial_j g_b(s + D_jX)}{g_b(s)}X^{(j)}; X \in \cap_{i=1}^{m}\overline{C}_i^{\alpha}(s,b)\right) \leq \gamma_1\frac{\nabla g_b(s)\cdot\eta}{g_b(s)},$$

where $\partial_j g_b = \frac{\partial g_b}{\partial s_j}$.

Proof. If $g_b(s) < 1$, then we have that

$$\frac{\partial_j g_b(s + D_jX)}{g_b(s)}X^{(j)} = 2\frac{\sum_{i=1}^{m}P\left(v_i^{*T}X > b\delta^* - v_i^{*T}s - v_i^{*T}D_jX\right)v_i^{*(j)}X^{(j)}}{h_b(s)} \times\frac{h_b(s + D_jX)}{h_b(s)},$$

where $v_i^{*(j)}$ denotes the j-th component of the vector v_i^*. In addition, note that if $X \in \cap_{i=1}^{m}\overline{C}_i^{\alpha}(s,b)$, then we must have that

$$h_b(s + D_jX) \leq \sum_{i=1}^{m}G_i\left(\left(b\delta^* - v_i^{*T}s\right)(1-a)\right).$$

Because the G_i's are all regularly varying, by Karamata's theorem, we have that there exists a constant $\gamma_0 \in (0,\infty)$ such that

$$\frac{h_b(s + D_jX)}{h_b(s)} \leq \gamma_0.$$

Similarly, we have that if $X \in \cap_{i=1}^{m}\overline{C}_i^{\alpha}(s,b)$ is

$$\sum_{i=1}^{m}P\left(v_i^{*T}X > b\delta^* - v_i^{*T}s - v_i^{*T}D_jX\right) \leq \sum_{i=1}^{m}P\left(v_i^{*T}X > \left(b\delta^* - v_i^{*T}s\right)(1-a)\right).$$

Note that,

$$\frac{\partial_j g_b(s + D_jX)}{\partial_j g_b(s)}X^{(j)} \to X^{(j)}$$

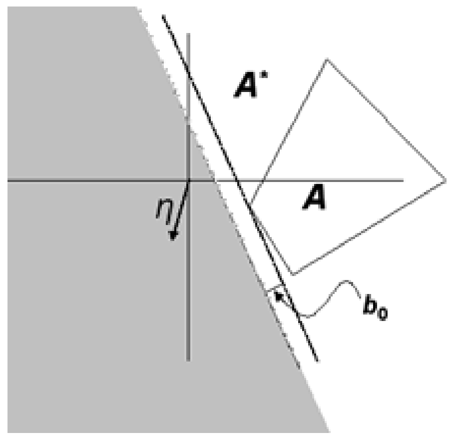


Figure 2: Two dimensional diagram illustrating the region where Lemma 1 is satisfied

as $b\delta^* - v_i^{*T}s \rightarrow \infty$. Also, the sequence is bounded by $X^{(j)}$ on the set $\cap_{i=1}^{m}\overline{C}_i^{\alpha}(s,b)$. Now, note that $g_b(s) < 1$ implies that, for all $1 \leq i \leq m$, $b\delta^* - v_i^{*T}s > G_i^{-1}(1/c_g^{1/2}) \nearrow \infty$ as $c_g \nearrow \infty$. As a consequence, if c_g is sufficiently large, and if $g_b(s) < 1$ the Dominated Convergence Theorem gives

$$E\left(\partial_j g(s + D_j X) X^{(j)}; X \in \cap_{i=1}^{m}\overline{C}_i^{\alpha}(s,b)\right) \leq \gamma_1 \partial_j g(s) E X^{(j)},$$

for some $\gamma_1 > 0$ and every j. This allows us to conclude the proof. ■

Remark 2 The gray area in Figure 2 shows the form of the region where $g_b(s) < 1$.

Lemma 1 combined with (10) and the Taylor development applied to (9) yields that if $g_b(s) < 1$,

$$J_1 + J_2 \leq \frac{1}{1 - p(s)}\left(1 - 2\gamma_1\delta^* \frac{\sum_{i=1}^{m} P\left(v_i^{*T}X > b\delta^* - v_i^{*T}s\right)}{h_b(s)}\right) + \frac{P(s)^2}{c_g h_b(s)^2 p(s)}.$$

Therefore, if we select

$$p(s) = \min\left(\theta\frac{P(s)}{h_b(s)}, 1/2\right) I(s; g_b(s) < 1), \tag{12}$$

for some $\theta > 0$. On the set that $g_b(s) < 1$, we have

$$J_1 + J_2 \leq 1 + 2\theta \frac{P\left(X \in \cup_{i=1}^m C_i^a(s,b)\right)}{h_b(s)} + \frac{P\left(X \in \cup_{i=1}^m C_i^a(s,b)\right)}{c_g \theta h_b(s)} - 2\gamma_1 \delta^* \frac{\sum_{i=1}^m P\left(v_i^{*T} X > b\delta^* - v_i^{*T} s\right)}{h_b(s)}. \tag{13}$$

Now, our task is to appropriately select θ 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 θ 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 *There exists $\gamma_2 \in (0,\infty)$ such that [γ_2 IS ONE???, ALSO, IS THAT SO IMPORTANT TO BE A LEMMA???]*

$$P\left(X \in \cup_{i=1}^m C_i^a(s,b)\right) \leq \gamma_2 \sum_{i=1}^m P\left(v_i^{*T} X > b\delta^* - v_i^{*T} s\right)$$

for all $s \in R^d$.

Proof. By the union bound, we obtain that

$$P\left(X \in \cup_{i=1}^m C_i^a(s,b)\right) \leq \sum_{i=1}^m P\left(v_i^{*T} X > a\left(b\delta^* - v_i^{*T} s\right)\right).$$

The conclusion of the lemma then follows directly using the definition of regular variation. ■

Applying the previous result to (13) we can construct the solution to a Lyapunov inequality that will allow us to control the variance of our estimator. We summarize this construction in our next result.

Proposition 3 *Given the family of importance samplers defined via (5), with $p(s)$ defined according to (12) and $g_b(s)$ satisfying (7). We have that one can compute θ and c_g such that the Lyapunov inequality (8) is satisfied throughout $s \in \mathbb{R}^d$. In particular, any pair (θ, c_g) such that $\theta \leq \gamma_1 \delta^*/(2\gamma_2)$ and $c_g \geq \gamma_2/\theta$ are valid choices.*

Proof. First, for s such that $g_b(s) < 1$, we have, combining the development behind (13) and Lemma 2,

$$E\left(\frac{g_b(s+X)}{g_b(s)} r(s, s+X)\right)$$
$$\leq 1 + \left(2\theta\gamma_2 + \frac{\gamma_2}{c_g\theta} - 2\gamma_1\delta^*\right) \sum_{i=1}^{m} P\left(v_i^{*T}X > b\delta^* - v_i^{*T}s\right).$$

So, for instance, if we select $\theta \leq \gamma_1\delta^*/(2\gamma_2)$ and $c_g \geq \gamma_2/\theta$ then the previous expression is guaranteed to be less than 1. On the other hand, if $g_b(s) = 1$, we clearly have (since $g_b \leq 1$) that selecting $p(s) = 0$ implies

$$E(g_b(s+X)\, r(s, s+X)) = Eg_b(s+X) \leq g_b(s) = 1$$

and the inequality then holds throughout $\mathbb{R}^d$. ■

We conclude with a summary of the proposed algorithm (for the generation of a single replication of Z_b). An issue that remains pending is the termination of the algorithm. We shall address this problem later sections.

Basic Algorithm

Set $b > 0$ and fix $a \in (0,1)$. Initialize $s = 0$, $REACH = 0$ and $Z = 1$. Assume that c_g and θ have been selected in order to satisfy (2) and set $g_b(w)$ according to (7)

STEP 1

While $REACH = 0$

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

Else set

$$p \longleftarrow \min\left(\theta \frac{P\left(X \in \cup_{i=1}^{m} C_i^a(s,b)\right)}{h_b(s)}, 1/2\right)$$

and sample X as follows. With probability p generate X with law $\mathcal{L}(\,X|\,X \in B_{s,bA})$, with probability $1-p$ sample X with law $\mathcal{L}(\,X|\,X \in B_{s,bA}^c)$. Then, update

$$Z_b \longleftarrow Z_b \cdot (p^{-1} P\left(X \in \cup_{i=1}^{m} C_i^a(s,b)\right) I\left(X \in \cup_{i=1}^{m} C_i^a(s,b), s \notin R_b\right)$$
$$+ (1-p)^{-1} P\left(X \in \cap_{i=1}^{m} \overline{C}_i^\alpha(s,b)\right) I\left(X \in \cap_{i=1}^{m} \overline{C}_i^\alpha(s,b), s \notin R_b\right)).$$

Endif

Update

$$s \longleftarrow s + X,$$

If $s \in bA$ then $REACH \longleftarrow 1$

Endif

Loop

STEP 2 RETURN Z_b.

The next result summarizes the variance properties of the estimator.

Theorem 1 *If Assumptions A and B are in force, then the estimator Z_b given by the previous algorithm has bounded relative error.*

Proof. The result follows by a straightforward application of the Lyapunov inequality and results of Hult and Lindskog (2005). In particular, we obtain that

$$E_s^q Z_b^2 = E_s Z_b \leq g_b(s).$$

The conclusion of the result follows directly from Proposition 2 and simple properties of regularly varying functions. ■

5 Path Generation, Normalizing Constants and Alternative Parameterizations

Several issues related to the implementation are the path generation under $E_s^q(\cdot)$ and the evaluation of $r(\cdot)$. These two issues are related.

It is usually possible to design an acceptance / rejection procedure to simulate increments without explicit knowledge of the normalizing constants, such as $P(X \in \cup_{i=1}^m C_i^a(s,b))$ involved in computing $r(\cdot)$. Moreover, such procedure can sometimes be designed, parametrically in s and b, so that the acceptance probability remains bounded away from zero uniformly over s and $b > 0$. This is explained in Blanchet and Glynn (2007) in one dimensional case but the ideas can be adapted to higher dimension distributions. Now, even if one can generate the paths in an appropriate way, there is problem of calculating $P(X \in \cup_{i=1}^m C_i^a(s,b))$, which is required to output the corresponding estimator. This calculation is the most interesting part in the higher dimensional implementation. [[[**THIS SEEMS TO PROPOSED AN OPEN PROBLEM WITHOUT SOLUTION.**]]]

An alternative approach is to note that there is a fair amount of flexibility in the selection of the mixture sampler in order to take advantage of particular problem structure to compute $P\left(X \in \cup_{i=1}^{m} C_i^a\left(s, b\right)\right)$. For instance, consider any parametric family of sets $\{B\left(s, b\right) : s \in \mathbb{R}^d,\ b > 0\}$ such that $\cup_{i=1}^{m} C_i^a\left(s, b\right) \subseteq B\left(s, b\right)$ and with the property that there exists a constant $\widetilde{\delta} > 0$ such that

$$\widetilde{\delta} P\left(X \in B\left(s, b\right)\right) \leq P\left(X \in \cup_{i=1}^{m} C_i^a\left(s, b\right)\right), \tag{14}$$

a property that is often easy to verify in the regularly varying setting. Define $\widetilde{P}\left(s\right) = P\left(X \in B\left(s, b\right)\right)$ and, given a mixture probability $\widetilde{p}\left(s\right)$ put

$$\frac{\widetilde{q}_{X|s}\left(x\right)}{f_X\left(x\right)} \triangleq \widetilde{r}\left(s, s+x\right) = \frac{\widetilde{P}\left(s\right)}{\widetilde{p}\left(s\right) + \left(1 - \widetilde{p}\left(s\right)\right)\widetilde{P}\left(s\right)} I\left(x \in B\left(s, b\right)\right) \tag{15}$$

$$+ \frac{1}{1 - \widetilde{p}\left(s\right)} I\left(x \in B\left(s, b\right)^c\right). \tag{16}$$

We shall use $\widetilde{E}_{s_0}\left(\cdot\right)$ to denote the expectation operator corresponding to path-measure (denoted by $\widetilde{P}_{s_0}\left(\cdot\right)$) generated by the state-dependent increment distribution $\widetilde{q}_{X|s}\left(\cdot\right)$ given that the initial position of the initial position of the process is $S_0 = s_0$.

One can follow the same steps described in Section 4 in order to verify the Lyapunov inequality. We claim that a parametric family of functions defined via (7) can be used to construct a Lyapunov function with a suitable selection of the parameter c_g. Indeed, we can define

$$E\left(\frac{g_b\left(s+X\right)}{g_b\left(s\right)}\widetilde{r}\left(s, s+X\right)\right) = \widetilde{J}_1 + \widetilde{J}_2,$$

where

$$\widetilde{J}_1 \leq E\left(\frac{g_b\left(s+X\right)}{g_b\left(s\right)}; X \in B\left(s, b\right)\right)\frac{P\left(X \in B\left(s, b\right)\right)}{\widetilde{p}\left(s\right)},$$

$$\widetilde{J}_2 = E\left(\frac{g_b\left(s+X\right)}{g_b\left(s\right)}; X \notin B\left(s, b\right)\right)\frac{P\left(X \notin B\left(s, b\right)\right)}{1 - \widetilde{p}\left(s\right)},$$

and

$$\widetilde{p}\left(s\right) = \min\left(\theta\frac{P\left(X \in B\left(s, b\right)\right)}{h_b\left(s\right)}, 1/2\right) I\left(s : g_b\left(s\right) < 1\right). \tag{17}$$

Assuming that $g_b(s) < 1$ we have that

$$\widetilde{J}_1 \le \frac{P(X \in B(s,b))^2}{c_g h_b(s)^2 \widetilde{p}(s)},$$

which corresponds to (10). Similarly, the analysis of $\widetilde{J}_2$ follows exactly the same steps as that of J_2 in Section 4, thereby obtaining the inequality

$$\begin{aligned}\widetilde{J}_1 + \widetilde{J}_2 &\le 1 + 2\theta \frac{P(X \in B(s,b))}{h_b(s)} \\ &\quad + \frac{P(X \in B(s,b))}{c_g \theta h_b(s)} - 2\gamma_1 \delta^* \frac{\sum_{i=1}^m P\left(v_i^{*T} X > b\delta^* - v_i^{*T} s\right)}{h_b(s)}. \qquad (18)\end{aligned}$$

We can conclude the following result entirely analogous to Proposition 3.

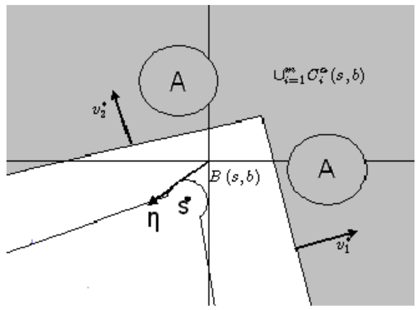
Proposition 4 *Given the family of importance samplers defined via (15), with $\widetilde{p}(s)$ defined according to (17). We have that one can compute θ and c_g such that $\widetilde{J}_1 + \widetilde{J}_2 \le 1$ for all $s \in \mathbb{R}^d$. In particular, any pair (θ, c_g) such that $\theta \le \gamma_1 \delta^* / (2\gamma_2 \widetilde{\delta})$ and $c_g \ge \gamma_2 \widetilde{\delta} / \theta$ are valid choices.*

Proof. First, for $s \notin \widetilde{R}_b$, we have, combining the development behind (18), Lemma 2 and the definition of $\widetilde{\delta}$ that

$$\begin{aligned}&E\left(\frac{g_b(s+X)}{g_b(s)} \widetilde{r}(s, s+X)\right) \\ &\le 1 + \left(\frac{2\theta\gamma_2}{\widetilde{\delta}} + \frac{\gamma_2}{c_g \theta \widetilde{\delta}} - 2\gamma_1 \delta^*\right) \sum_{i=1}^m P\left(v_i^{*T} X > b\delta^* - v_i^{*T} s\right).\end{aligned}$$

The rest of the analysis then follows just as in Proposition 3. ■

The previous analysis indicates that there is plenty of flexibility when choosing an appropriate family of importance sampling distributions in order to facilitate the path generation and calculation of constants such as $P(X \in B(s,b))$ required in the evaluation of the corresponding importance sampling estimator. For instance, consider the case of t-distributed increments discussed in Example 1. In such situation, evaluation of constants such as $P(X \in \cup_{i=1}^m C_i^a(s,b))$ is not entirely straightforward. However, it is not difficult to construct $B(s,b)$, suitably expressed in polar coordinates, that contains $\cup_{i=1}^{m} C_i^a(s,b)$ and satisfies (14); a diagram indicating the structure of $B(s,b)$ in a two dimensional case is given next.



The evaluation of $P(X \in B(s,b))$ and the corresponding path generation can then be easily done in Polar coordinates.

6 Introducing a Controlled Bias

The analysis of Section 4 allows to control the variance of the estimator given by the Basic Algorithm. However, nothing is said about the termination time of the algorithm. In fact, even in one dimensional case, the zero-variance change-of-measure might have a termination time with infinite mean. This occurs, for instance, when the increments have infinite variance. Therefore, we will develop a criterion that allows us to safely stop the path generation introducing a controlled bias in our procedure. We shall consider biased estimators for which the relative bias can be controlled in a suitable way. In particular, we shall construct a set $B_{b,\beta}$ depending on b and a suitable parameter β so that $T_{B_{b,\beta}} < \infty$ with probability one under the change-of-measure. Therefore,

$$P_0(T_{A_b} < \infty) = \upsilon_{b,\beta}(0) + P_0\left(T_{A_b} < \infty, T_{B_{b,\beta}} < T_{A_b}\right),$$

where $\upsilon_{b,\beta}(0) = P_0\left(T_{A_b} < T_{B_{b,\beta}}\right)$ and

$$\frac{P_0\left(T_{A_b} < \infty, T_{B_{b,\beta}} < T_{A_b}\right)}{P_0(T_{A_b} < \infty)} \leq \gamma(\beta) \tag{19}$$

for a function $\gamma(\beta)$ such that $\gamma(\beta) = o(\beta^{-p})$ for some $p > 1$ as $\beta \nearrow \infty$. It follows easily that with the aid of a strongly efficient estimator for $\upsilon_{b,\beta}(0)$

and a bound such as (19), one can efficiently estimate $P_0(T_{A_b} < \infty)$ with a controlled relative error. In doing so, one must:

- find a computable function $\gamma(\beta)$ to control the relative bias;
- provide a selection of $B_{b,\beta}$;
- provide bound for $E_s^q T_{B_{b,\beta}}$ to control the expected termination time.

Computing function $\gamma(\beta)$. Chebyshev's inequality combined with Proposition 1 yields that

$$\begin{aligned}
&P_0\left(T_{A_b} < \infty, T_{B_{b,\beta}} < T_{A_b}\right) \\
&= E_0\left(P\left(T_{A_b} < \infty \middle| W_{T_{B_{b,\beta}}}\right); T_{B_{b,\beta}} < T_{A_b}\right) \\
&\leq \varepsilon^{-1/2} E\left[g_b\left(W_{T_{B_{b,\beta}}}\right)^{1/2}; T_{B_{b,\beta}} < T_{A_b}\right] \\
&\leq \varepsilon^{-1/2} \sup_{w \in B_{b,\beta}} g_b(w)^{1/2}.
\end{aligned}$$

In the context of regularly varying functions, we can appropriately choose set $B_{b,\beta}$, such that,

$$\frac{P_0\left(T_{A_b} < \infty, T_{B_{b,\beta}} < T_{A_b}\right)}{P_0\left(T_{A_b} < \infty\right)} \leq \frac{\sup_{w \in B_{b,\beta}} g_b(w)^{1/2}}{v_{b,\beta}(0)} = c_0 \beta^{-p} \tag{20}$$

for some $p, c_0 \in (0, \infty)$, where $v_{b,\beta}(0)$ can be computed efficiently.

Selecting $B_{b,\beta}$. Our choice of $B_{b,\beta}$ takes advantage of the fact that, prior to the big jump, the process drifts according to its unconditional mean, η. We then construct a suitable set $B_{\beta,b}$ that intersects the fluid path, $\{\eta t : t > 0\}$. The parameter θ is selected so that $E_s^q X \approx \eta$ and we will be able to guarantee that the process hits either $B_{\beta,b}$ or A_b in an expected time of order $O(b)$. The selection of θ might come at a price of potentially increasing the value of c_g (as indicated in the constrains given in Proposition 3) but it does not have an effect in terms of complexity rates of the algorithm. The set $B_{\beta,b}$ is simply

$$B_{\beta,b} = \left\{t\eta / \|\eta\|_2 + \cap_{i=1}^m \overline{C}_i^1(0, b); t \geq \beta b\right\}.$$

The next diagram [[[**THE DIAGRAM IS MISSING**]]] illustrates the location of the set $B_{\beta,b}$.

Providing bound for $E_s^q T_{B_{b,\beta}}$. The following well known result (see Meyn and Tweedie (1993)) establishes the Lyapunov inequality required to control the behavior of $E_s^q T_{B_{\beta,b}}$.

Lemma 3 *Suppose that one can find a non-negative funciton $\varrho(s)$ and a constant $\rho > 0$ so that*

$$E\varrho(s+X)\, r(s, s+X)^{-1} \leq \varrho(s) - \rho,$$

for $s \notin B_{\beta,b}$. Then, $E_s^q T_{B_{\beta,b}} \leq \varrho(s)/\rho$ for $s \notin B_{\beta,b}$.

Given vectors $\{v_j : 0 \leq j \leq m\}$ and constants $c_0 > 0$ and $\{\chi_j : 0 \leq j \leq m\}$, all of which will be selected momentarily, define

$$\varrho(s) = c_0 \log\left(\sum_{i=0}^{m} \exp\left(\left(\chi_i + s^T v_i\right)/c_0\right)\right) \tag{21}$$

and put

$$w_j(s) = \frac{\exp\left(\left(\chi_j + s^T v_j\right)/c_0\right)}{\sum_{i=0}^{m} \exp\left(\left(\chi_i + s^T v_i\right)/c_0\right)}.$$

First we state the following result which summarizes useful properties of $\varrho(\cdot)$ and its derivatives.

Lemma 4

i) $\max_{i=0}^{m}\left(\chi_i + s^T v_i\right) \leq \varrho(s) \leq \max_{i=0}^{m}\left(\chi_i + s^T v_i\right) + c_0 \log(m+1)$
ii) $(\partial\varrho)(s) = \sum_{j=0}^{m} v_j w_j(s)$,
iii) $(\partial^2\varrho)(s) = \sum_{j=0}^{m} w_j(s)(1 - w_j(s))\, v_j v_j^T / c_0$.

Proof. Item i) is almost direct and part ii) follows from basic calculus. Part iii) is obtained by noting that

$$\begin{aligned}(\partial w_j)(s) &= w_j(s)\, \partial \log w_j(s) \\ &= w_j(s)\left(v_j^T/c_0 - w_j(s)\, v_j^T/c_0\right) = w_j(s)(1 - w(s)) v_j^T / c_0.\end{aligned}$$

Therefore,

$$\left(\partial^2 \rho\right)(s) = \sum_{j=0}^{m} w_j(s)(1 - w_j(s))\, v_j v_j^T / c_0$$

and the result follows. ■

Now suppose that $v_0 = 0$, $v_i = v_i^*$ for $1 \leq i \leq m$ and $\chi_i = 0$ for all i. The constant c_0 is selected of the form $c_0 = \kappa b$ for $\kappa \in (0, \infty)$ sufficiently large. The following result shows that $\varrho(\cdot)$ satisfies the Lyapunov inequality..

Proposition 5 *One can choose $\theta > 0$ small enough in (12) (or c_g large enough) so that the Lyapunov inequality in Lemma 3 holds for an appropriate selection of $\rho > 0$.*

Proof. Since $v_0 = 0 = \chi_0$ the non-negativity of ϱ is satisfied. Using Taylor's theorem with reminder as in Section 4 and Lemma 4 we obtain

$$
\begin{aligned}
&E\varrho(s+X)\, r^{-1}(s, s+X) \\
&= \varrho(s) + p(s)\, E\left(\sum_{j=1}^{m} w_j(s+DX)\, X^T v_j^* \middle| X \in \cup_{i=1}^{m} C_i^a(s,b) \right) \\
&+ (1-p(s))\, E\left(\sum_{j=1}^{m} w_j(s+DX)\, X^T v_j^* \right),
\end{aligned}
$$

where D is a diagonal matrix so that $D_{i,i} \in (0,1)$. Using basic regularly varying properties it follows that for each $\varepsilon \in (0, \delta^*/4)$ there exists $\theta > 0$ so that

$$
\begin{aligned}
&p(s)\, E\left(\sum_{j=1}^{m} w_j(s+DX)\, X^T v_j^* \middle| X \in \cup_{i=1}^{m} C_i^a(s,b) \right) \\
&\leq p(s) \left| E\left(\max_{j=1}^{m} X^T v_j^* \middle| X \in \cup_{i=1}^{m} C_i^a(s,b) \right) \right| < \varepsilon.
\end{aligned}
$$

Furthermore, by choosing χ_j large enough we can ensure, given that $EX^T v_j^* \leq -\delta^*$ that

$$
(1-p(s))E\left(\sum_{j=1}^{m} w_j(s+DX)\, X^T v_j^* \right) \leq -\delta^*/2,
$$

which yields

$$
E\varrho(s+X)\, r^{-1}(s, s+X) \leq \varrho(s) - \rho
$$

with $\rho = \delta^*/4$ and the result follows. ■

Given that a strongly efficient estimator, $Z_{\beta}(b)$, is available for $\upsilon_{b,\beta}(0)$, that is

$$\sup_{b\geq 1}\frac{EZ_{\beta}(b)^{2}}{\upsilon_{b,\beta}(0)^{2}}\leq c_{1}$$

for some $c_1 < \infty$, consider $Y_{n,\beta}(b)=\sum_{i=1}^{n}Z_{\beta,i}(b)/n$ where the $Z_{\beta,i}(b)$'s are independent copies of $Z_{\beta}(b)$. Then, given $\varepsilon > 0$ we obtain

$$\begin{aligned}
&P\left(\left|Y_{n,\beta}(b)-u_{b}(0)\right|\geq\varepsilon u_{b}(0)\right)\\
&\leq P\left(\left|Y_{n,\beta}(b)-\upsilon_{b,\beta}(0)\right|+\left|u_{b}(0)-\upsilon_{b,\beta}(0)\right|\geq\varepsilon u_{b}(0)\right)\\
&\leq P\left(\left|Y_{n,\beta}(b)-\upsilon_{b,\beta}(0)\right|\geq\left(\varepsilon-\gamma(\beta)\right)u_{b}(0)\right)\\
&\leq\frac{c_{1}}{\left(\varepsilon-\gamma(\beta)\right)^{2}n}\leq\frac{c_{1}}{\left(\varepsilon-c_{0}\beta^{-p}\right)^{2}n},
\end{aligned}$$

assuming $\beta > (\varepsilon/c_0)^{-1/p}$. Therefore, $n = O\left(\varepsilon^{-2-1/p}\delta^{-1}\right)$ replications suffice to obtain an estimator with ε-relative error and $(1-\delta)\cdot 100\%$ confidence.

7 Total Variation Approximation of the Zero-variance Sampler and Finite Termination time

We consider target sets that can be written as the union of half-spaces and our goal is to describe an algorithm that can be shown to approximate in total variation the conditional distribution of the random walk given that the target set is eventually reached (such conditional distribution coincides with the zero-variance importance sampling distribution). we shall construct an importance sampling estimator that achieves vanishing relative error and that reaches the target set in $O(b)$ step in expectation. As a corollary we shall obtain a sampler that approximates, in total variation, the conditional distribution of the random walk given the event $T_{bA} < \infty$.

We shall assume that $A = \cup_{i=1}^{m}\{y : y^{T}v_{i}^{*} \geq b\xi_{i}\}$ with $\xi_i \in (0,\infty)$ for $1 \leq i \leq m$. The next result introduces a linear transformation that will allows us to simplify the analysis of various large deviations approximations (such as (6)).

Lemma 5 *There exists an invertible matrix* $M \in \mathbb{R}^{d\times d}$ *so that*

$$\eta^T M^T v_i^* / \xi_i = -1$$

for $1 \leq i \leq m$.

Proof. Consider the matrix $V \in \mathbb{R}^{d\times m}$ with i-th column equal to v_i^*. By assumption A the matrix V is of rank m and there exists a matrix $G \in \mathbb{R}^{d\times d}$ such that $(GV)^T = [I_m \ 0]$, where I_m is the identity matrix of size $m \times m$ and 0 is the zero matrix of size $m \times (d-m)$. On the other hand, we can pick a rotation matrix so that $R\eta$ has only strictly negative components. We then put $M^T = R^T \Xi G$, where Ξ is a diagonal matrix so that $\Xi_{i,i} = \xi_i / (R\eta)_i$ for $1 \leq i \leq m$ and $\Xi_{i,i} = 1$ for $1 < i \leq d$.we then obtain (letting $\mathbf{e}_i = (0, ..., \underset{i\text{-th}}{1}, ..., 0)^T$)

$$\eta^T M^T v_i^* / \xi_i = \eta^T R^T \Xi G v_i^* \xi_i = (\Xi R \eta)^T e_i / \xi_i = -1$$

for all $1 \leq i \leq m$. Clearly the matrix M^T is invertible and therefore M is invertible. ■

Throughout the rest of the section we shall work with the linearly transformed increments, $X_i' = MX_i$, the transformed random walk $S_k' = X_1' + ... + X_k'$, and the set $A' = MA = \{z : z = My, y \in A\}$. We define

$$T_{bA'}' = \inf\{n \geq 0 : S_k'^T v_i' > b \text{ for some } 1 \leq i \leq m\},$$

where $v_i' = v_i^*/\xi_i$ and note that $u_b(0) = P_0(T_{bA'}' < \infty)$. We also that the following stronger from of assumption B holds:

B') Assume that $\mu(\cdot)$ has regularly varying index $\alpha > 2$ and that $\mu(B) > 0$ for every Borel set B with positive Lebesgue measure.

The condition that $\alpha > 2$ is required even in the one dimensional to show that the termination time of the zero-variance change-of-measure is of order $O(b)$ as $b \nearrow \infty$ (see CITE ASMUSSEN AND KLUPPELBERG'S PAPER).

The fluid approximation (6) takes the form

$$P_s\left(T'_{bA'}<\infty\right) \approx \int_0^{\infty} P\left(\max_{i=1}^{m}\left(X'^T v'_i + s^T v'_i + (M\eta)^T v'_i t\right) > b\right) dt$$
$$= \int_0^{\infty} P\left(\max_{i=1}^{m}\left(X'^T v'_i + s^T v'_i\right) > b + t\right) dt.$$

In turn, if we define

$$\varrho_b(s) = \varepsilon \log\left(\sum_{j=1}^{m} \exp\left(\left(b - s^T v'_j\right)/\varepsilon\right)\right)$$

for $\varepsilon > 0$, then property i) from Lemma 4 yields

$$= \int_0^{\infty} P\left(\max_{i=1}^{m}\left(X'^T v'_i + s^T v'_i\right) > b + t\right) dt.$$
$$\leq \int_0^{\infty} P\left(\max_{i=1}^{m} X'^T v'_i > \varrho_b(s) + t - \varepsilon \log m\right) dt$$
$$= \int_{\varrho_b(s)-\varepsilon \log m}^{\infty} P\left(\max_{i=1}^{m} X'^T v'_i > u\right) du.$$

We define

$$G(t) = \int_{t-\varepsilon \log m}^{\infty} P\left(\max_{i=1}^{m} X'^T v'_i > u\right) du,$$

then the upper bound for the fluid approximation given above can be written as $H_b(s) = G(\varrho_b(s))$. Note that the function $H_b(\cdot)$ is almost everywhere continuously differentiable.

We now describe our proposed change-of-measure. We set

$$\frac{q^*_{X|s}(x)}{f_X(x)} \triangleq r^*(s, s+x)$$
$$= \frac{P\left(\max_{i=1}^{m} X^T v'_i > a\varrho_b(s)\right) I\left(\max_{i=1}^{m} X^T v'_i > a\varrho_b(s)\right)}{p^*(s) + (1 - p^*(s)) P\left(\max X^T v'_i > a\varrho_b(s)\right)}$$
$$+ \frac{I\left(\max_{i=1}^{m} X^T v'_i \leq a\varrho_b(s)\right)}{1 - p^*(s)}.$$

and we use the notation $P^*_{s_0}$ for the probability measure on the path space associated to $q^*(\cdot)$ given that the initial state of the process is $S_0 = s_0$ (the corresponding expectation operator is denoted by $E^*_{s_0}$). We put

$$p^*(s) = \frac{\theta P\left(\max_{i=1}^m X^T v_i' > \varrho_b(s)\right)}{G(\varrho_b(s))}.$$

It follows easily that

$$p^*(s) \sim \frac{\theta(\alpha - 1)}{b - \max_{i=1}^m s^T v_i'}$$

as $b - \max_{i=1}^m s^T v_i' \nearrow \infty$. We propose as Lyapunov function

$$g_b^*(s) = c_1 G(\varrho_b(s) + c_2)^2$$

and we will pick c_1 as close to one as desired. First, note that Lemma 4 implies

$$\varrho_b(s + X) \geq \varrho_b(s) + \partial\varrho_b(s) X,$$

therefore

$$\begin{aligned} 0 &\leq G(\varrho_b(s+X) + c_2) \leq G(\varrho_b(s) + \partial\varrho_b(s) X + c_2) \\ &= G(\varrho_b(s) + c_2) \\ &- \int_0^1 P\left(\max_{i=1}^m X'^T v_i' > \varrho_b(s) + c_2 + \partial\varrho_b(s) Xu - \varepsilon \log m\right) du \partial\varrho_b(s) X. \end{aligned}$$

Put

$$\begin{aligned} J_1^* &= \frac{P\left(\max_{i=1}^m X'^T v_i' > a\varrho_b(s)\right)^2}{c_1 G(\varrho_b(s) + c_2)^2 (p^*(s) + (1 - p^*(s)) P(\max X'^T v_i' > a\varrho_b(s)))} \\ &\sim \frac{P\left(\max_{i=1}^m X'^T v_i' > a\varrho_b(s)\right)^2}{c_1 G(\varrho_b(s) + c_2)^2 p^*(s)} \sim \frac{P\left(\max_{i=1}^m X'^T v_i' > a\varrho_b(s)\right)}{c_1 \theta G(\varrho_b(s) + c_2)} \sim \frac{p^*(s)}{c_1 \theta^2}. \end{aligned}$$

Now set

$$
\begin{aligned}
&\left(1-p^{*}\left(s\right)\right) J_{2}^{*} \\
&= E\left(\frac{G\left(\varrho_{b}\left(s+X\right)+c_{2}\right)^{2}}{G\left(\varrho_{b}\left(s\right)+c_{2}\right)^{2}}; \max_{i=1}^{m} X^{T} v_{i}^{\prime} \leq a\varrho_{b}\left(s\right)\right) \\
&\sim \frac{1}{\left(1-p^{*}\left(s\right)\right)}\left(1-2\frac{P\left(\max_{i=1}^{m} X^{\prime T} v_{i}^{\prime} > \varrho_{b}\left(s\right)+c_{2}-\varepsilon\log m\right)}{G\left(\varrho_{b}\left(s\right)+c_{2}\right)}\right) \\
&\sim 1+p^{*}\left(s\right)-2\frac{p^{*}\left(s\right)}{\theta}.
\end{aligned}
$$

We then conclude that

$$
J_{1}^{*}+J_{2}^{*} \sim \frac{p^{*}\left(s\right)}{c_{1}\theta^{2}}+1+p^{*}\left(s\right)-2\frac{p^{*}\left(s\right)}{\theta}.
$$

Therefore, selecting $\theta = 1-\varepsilon$, $c_{1} = 1+2\varepsilon$ and c_{2} sufficiently large we conclude the following result.

Theorem 2 *For every $\varepsilon > 0$ there exists c_{2} such that $g_{b}^{*}\left(s\right) < 1$ for all s such that $\max_{i=1}^{m} s^{T} v_{i}^{\prime} < b$. Moreover, assuming $g_{b}^{*}\left(s\right) < 1$ we have that $J_{1}^{*}+J_{2}^{*} \leq 1$ and therefore*

$$
\begin{aligned}
(1-\varepsilon)P_{0}\left(T_{bA^{\prime}}^{\prime} < \infty\right)^{2} &\leq E_{0}^{*}\left(\textit{2nd Moment of Estimator}\right) \\
&\leq (1+2\varepsilon)G\left(0\right)^{2} \leq (1+3\varepsilon)P_{0}\left(T_{bA^{\prime}}^{\prime} < \infty\right)^{2}
\end{aligned}
$$

for all b sufficiently large.

The connection to the approximation in total variation distance is given by the following result taken from CITE L'ECUYER, BLANCHET, TUFFIN AND GLYNN (2009).

Lemma 6 *Given a probability measure P and Q such that $dQ = MdP$. Suppose that $E^{Q}\left(M^{2}\right) = EM \leq 1+\varepsilon$. Then,*

$$
d_{TV}\left(P,Q\right) \leq 2\varepsilon.
$$

As a corollary of the previous Lemma we have the following result.

Corollary 3 *The change-of-measure given in Theorem 2 approximates the conditional distribution of the random walk given that $T'_{bA'} < \infty$ in total variation.*

We now will establish that $E_0^* T'_{bA'} = O(b)$ as $b \nearrow \infty$. The strategy is first to show that it takes $O\left(b - \max_{i=1}^m s^T v_i'\right)$ steps to obtain a success jump (corresponding to $p^*(s)$). Then, we must show that, given the occurrence of a jump, the probability of hitting the target set bA' is uniformly bounded away from zero uniformly over the position of the walk at a time just prior to the jump. We start by arguing the later part, in other words,

$$\begin{aligned}
& P\left(\max_{i=1}^m (s+X')^T v_i' > b \middle| \max_{i=1}^m X'^T v_i' > a\left(b - \max_{i=1}^m s^T v_i'\right)\right) \\
&= \frac{P\left(\max_{i=1}^m (s+X')^T v_i' > b\right)}{P\left(\max X'^T v_i' > a\left(b - \max_{i=1}^m s^T v_i'\right)\right)} \geq \delta_0 > 0. \qquad (22)
\end{aligned}$$

To see this let i_1 such that $s^T v_{i_1}' = \max_{i=1}^m s^T v_i'$ then we have that

$$\begin{aligned}
& P\left(\max_{i=1}^m (s+X')^T v_i' > b \middle| \max_{i=1}^m X'^T v_i' > a\left(b - s^T v_{i_1}'\right)\right) \\
& \geq P\left(\max_{i=1}^m (s+X')^T v_i' > b \middle| X'^T v_{i_1} > a\left(b - s^T v_{i_1}'\right), \max_{i=1}^m X'^T v_i' = X'^T v_{i_1}\right) \\
& P\left(\max_{i=1}^m X'^T v_i' = X'^T v_{i_1} \middle| \max_{i=1}^m X'^T v_i' > a\left(b - s^T v_{i_1}'\right)\right).
\end{aligned}$$

That the previous quantity is bounded away from zero follows from conditions A and B' as follows:

$$\begin{aligned}
& P\left(\max_{i=1}^m X'^T v_i' = X'^T v_{i_1} \middle| \max_{i=1}^m X'^T v_i' > a\left(b - s^T v_{i_1}'\right)\right) \\
&= \frac{P\left(\max_{i=1}^m X'^T v_i' = X'^T v_{i_1}, X'^T v_{i_1} > a\left(b - s^T v_{i_1}'\right)\right)}{P\left(\max_{i=1}^m X'^T v_i' > a\left(b - s^T v_{i_1}'\right)\right)} \\
&\geq \frac{P\left(X'^T v_{i_1}' > a\left(b - s^T v_{i_1}'\right) \geq X'^T v_j' : j \neq i_1\right)}{P\left(\cup_{i=1}^m \{X'^T v_i' > a\left(b - s^T v_{i_1}'\right)\}\right)} \\
&\longrightarrow \frac{\mu\left(\{y : y^T M^T v_{i_1}' > a \geq y^T M^T v_j : j \neq i_1\}\right)}{\mu\left(\cup_{i=1}^m \{y : y^T M^T v_i' > a\}\right)} \qquad (23)
\end{aligned}$$

as $\left(b - s^T v_{i_1}'\right) \nearrow \infty$. It follows that the set $\{y : y^T M^T v_{i_1}' > a \geq y^T M^T v_j : j \neq i_1\}$ is a an unbounded open polytope. First note that the set $C = \{y :$

$a \geq y^T M^T v_j : j \neq i_1\}$ must be unbounded because of assumption A (since the ray $\{\eta t : t > 0\}$ must lie inside such set). Moreover, C must be a cylinder because is described by $m - 1 < d$ linearly independent vectors via the intersection of associated halfspaces. Finally, the halfspace $\{y : y^T M^T v'_{i_1} > a\}$ must cut C transversally because of linear independence (the contrary would imply that v'_{i_1} is a linear combination of the remaining v'_j's). The denominator in (23) is clearly bounded and therefore we conclude, by means of assumption B' that (23) follows.

Finally, we argue that it takes $O(b)$ steps to obtain a successful selection of the mixture component with probability given by p^*, regardless of the initial position of the chain. Let N_b be the number of steps required to obtain such succesful selection. We obtain that for each $k \in \mathbb{N}$

$$P_s^* (N_b > kb) = E_s \left(\prod_{j=0}^{kb} (1 - p^* (S_j)) \right).$$

Note that the probability measure in the right hand side is the nominal (original) probability; so, the process inside the expectation in the right hand side behaves like a standard random walk. We note that

$$\begin{aligned}
&E_s \left(\prod_{j=0}^{kb} (1 - p^* (S_j)) \right) \\
&\leq E_s \left(\prod_{j=0}^{k} (1 - p^* (S_{jb})) \right) \\
&= E_s \left(\prod_{j=0}^{k} (1 - p^* (S_{jb})); \|S_{jb} - \eta jb\|_2 \leq \varepsilon jb \right) \\
&+ E_s \left(\prod_{j=0}^{k} (1 - p^* (S_{jb})); \|S_{jb} - \eta jb\|_2 > \varepsilon jb \right).
\end{aligned}$$

It is not difficult to see, in view of assumption B', that the first term dominates the second one for large values of b (uniformly over s). Therefore, the previous expression can be bounded by

$$\kappa E_s\left(\prod_{j=0}^{k}(1-p^*(S_{jb}));\left\|S_{jb}-\eta jb\right\|_2\leq\varepsilon jb\right)$$

for an appropriate constant $\kappa_0 \in (0,\infty)$. Moreover, we have that for all s such that $b-\max_{i=1}^m s^T v_i' < \kappa_1 \in (0,\infty)$ for a constant κ_1 large enough,

$$\begin{aligned}
&E_s\left(\prod_{j=0}^{k}(1-p^*(S_{jb}));\left\|S_{jb}-\eta jb\right\|_2\leq\varepsilon jb\right)\\
&\leq\kappa_2\exp\left(-\sum_{j=0}^{k}\frac{\theta(1+\varepsilon)(\alpha-1)}{b-\max_{i=1}^m s^T v_i'+jb+\kappa_3\varepsilon b}\right)\\
&\leq\kappa_4\left(\frac{1}{bk/(b-\max_{i=1}^m s^T v_i'+b\kappa_3\varepsilon)+1}\right)^{\theta(1+\varepsilon)(\alpha-1)}\\
&=\kappa_4\left(\frac{1}{k/(1-\max_{i=1}^m s^T v_i'/b+\kappa_3\varepsilon)+1}\right)^{\theta(1+\varepsilon)(\alpha-1)},
\end{aligned}$$

where κ_2,κ_3 and κ_4 are finite constants. We conclude that

$$E_s^* N_b\leq\kappa_5(b-\max_{i=1}^m s^T v_i')$$

for some constant $\kappa_5\in(0,\infty)$ as long as $\theta(1+\varepsilon)(\alpha-1)>1$. The previous estimates combined allow to obtain the following result.

Theorem 4 *It takes $O(b)$ units of time to terminate the Basic Algorithm if one selects $\theta>1$ sufficiently close to one, under appropriate selections of c_1 and c_2.*

8 Numerical Experiment

We close this paper with some numerical experiments. Consider a random walk with in the increments following a multivariarte t distribution having

b	Estimate	SD	Covariates of Variation	Sample size
10	$2.67e-3$	$9.90e-5$	5.24	20000
20	$2.80e-4$	$8.42e-06$	4.26	20000
50	$2.61e-5$	$5.54e-07$	3.00	20000

Table 1: Simulations Results

density function

$$f(x) = \frac{c}{\left(1 + (x-\eta)^T (x-\eta)\right)^{(v+d)/2}}.$$

where $d = 2$, $v = 4$, and $\eta = (-1,-1)^\intercal$. We consider target sets of the form bA, where $A = \{(x,y)\,; x > 1 \text{ or } y > 1\}$. We implemented our algorithm to compute

$$P(T_{bA} < \infty)$$

for $b = 10$, 20, and 50. Note that Assumption A is satisfied with $v_1^* = (1,0)$ and $v_2^* = (0,1)$. For computational convenience, we choose $B(s,b) = \left\{(x_1,x_2)\,|\sqrt{x_1^2+x_2^2} > 0.9d(s)\,,\ x_1 > 0, \text{ or } x_2 > 0\right\}$, which correspond to choose $a = 0.9$. $d(s) = \min\left\{(b-x_1)^+, (b-x_2)^+\right\}$ is the distance from the current location to the target set. We implement the algorithm in (15) by choosing $\tilde{p}(s) = \frac{1}{3d(s)}$. To have a short termination time, we choose $B_{\beta,b} = \{(x_1,x_2)\,|x_1 < -\beta b, x_2 < -\beta b\}$, for $\beta = 100$. Therefore, we implement the algorithm as follows: simulate increments according to density function (15) until the radom walk reaches either bA or $B_{100,b}$. If it reaches bA first, output the accumulated likelihood ratio; otherwise, output zero.

Table 1 shows the simulation results. “Estimate”, “SD” are the estimates and estimated standard deviations based on 20,000 independent simulations. “Covariates of Variation” is the estimated standard deviation of one simulation divided by the estimate.

Acknowledgement 5 *This research was partially supported by NSF grant DMS 0595595*

References

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

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

[3] 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. QUESTA, 57, 99-113.

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

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

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

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

[8] 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.

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

[10] 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.

[11] Hult, H., and Lindskog, F. (2006) Heavy-tailed insurance portfolios: buffer capital and ruin probabilities. Technical Report No. 1441, School of ORIE, Cornell University.

[12] Kallenberg, O. (1983) *Random Measures.* 3rd Ed. Akademic Berlag, Berlin.

[13] Meyn, S. and Tweedie, R. (1993) *Markov Chains and Stochastic Stability.* http://decision.csl.uiuc.edu/~meyn/pages/book.html.

[14] 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.

[15] 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.