

# EFFICIENT IMPORTANCE SAMPLING IN RUIN PROBLEMS FOR MULTIDIMENSIONAL REGULARLY VARYING RANDOM WALKS

JOSE BLANCHET\* AND  
JINGCHEN LIU,\*\* *Columbia University*

## Abstract

We consider the problem of efficient estimation via simulation of first passage time probabilities for a multidimensional random walk with heavy-tailed increments. In addition to being a natural generalization to the problem of computing ruin probabilities in insurance—in which the focus is the maximum of a one-dimensional random walk with negative drift—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, using techniques based on Lyapunov inequalities, we argue that our estimator is strongly efficient in the sense that the *relative* mean squared error of our estimator can be made arbitrarily small by increasing the number of replications, uniformly as the probability of interest approaches 0.

*Keywords:* State dependent; importance sampling; regularly varying; random walk; heavy tail; strong efficiency; rare-event simulation; mixture family; multidimension

2010 Mathematics Subject Classification: Primary 60G50; 68W40

Secondary 60G70

## 1. Introduction

We consider a class of efficient rare-event simulation estimators for first passage time probabilities of multidimensional regularly varying random walks. Our setting is a natural generalization of the classical problem of estimating the tail distribution of the maximum of a one-dimensional random walk with negative drift—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 the references therein for a detailed account of the problem and its connections to insurance and queueing. Hult and Lindskog (2006) noted that calculating first passage time probabilities for multidimensional random walks corresponds to computing ruin probabilities for insurance companies with several lines of business.

Let  $S = (S_n : n \geq 1)$  be a regularly varying random walk in  $\mathbb{R}^d$  with drift  $\eta \in \mathbb{R}^d$  so that  $E S_n = n\eta$  (see Section 3 for the definition of multivariate regular variation). Our focus is on studying efficient importance sampling methodology for estimating the probability,  $u_b$ , that  $S$

---

Received 8 July 2009; revision received 28 January 2010.

\* Postal address: Department of Industrial Engineering and Operations Research, Columbia University, S. W. Mudd Building, 500 West 120th Street, New York, NY 10027-6699, USA. Email address: jose.blanchet@columbia.edu

\*\* Postal address: Department of Statistics, Columbia University, 1255 Amsterdam Avenue, Room 1030, New York, NY 10027, USA. Email address: jcliu@stat.columbia.edu

eventually hits a target set  $bA$  ( $b > 0$  is a scale parameter which eventually we will send to  $\infty$ ). We consider the case in which  $\eta$  points to the interior of an open cone that does not intersect the set  $A$ , which in turns is assumed to satisfy mild conditions to avoid degenerate situations (for instance,  $A$  does not lie in a lower-dimensional manifold; see Section 3).

In this paper we study 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 we consider a parametric family of importance sampling distributions based on mixtures—the precise form of which 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 (2008). In the light-tailed environment, mixtures are often used for nonconvex rare events to protect the behavior of the likelihood ratio due to rogue sample paths that deviate from the most likely large deviations path.

In the heavy-tailed case, mixtures arise as a natural way to mimic the behavior of the zero variance change of measure. The use of mixtures for importance sampling in heavy-tailed problems was introduced in Dupuis *et al.* (2006) for a geometric sum of one-dimensional independent and identically distributed (i.i.d.) regularly varying variables. More recently, Blanchet *et al.* (2007) and Blanchet and Liu (2008) exploited this idea more systematically and in environments in which rare events are caused by several heavy-tailed jumps in sequence, rather than just one. A feature to be emphasized in the current work and also in the cited work related to the heavy-tailed case is that the mixture parameters are state dependent. As a consequence, the efficiency analysis of the algorithm is not direct, which takes us to the second ingredient of our strategy. Recently developed techniques based on Lyapunov inequalities for Markov processes (see, for instance, Blanchet and Glynn (2008) or Blanchet *et al.* (2007)) 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 0. As we will see, solving a Lyapunov inequality involves finding a function (that we refer to as a 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 recently introduced in Blanchet and Glynn (2008) 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). Blanchet and Glynn (2008) proposed the use of a well-known approximation for such first passage time probabilities and developed 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 take advantage of 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) and Blanchet *et al.* (2007) 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 assumptions (A) and (B) in Section 3). In addition, contrary to the finite-time horizon problems treated in Blanchet and Liu (2008), the finite termination of the algorithm becomes an important issue that requires special attention in our current setting.

We provide the first rigorous analysis of estimators that can be shown to be strongly efficient for a class of multidimensional heavy-tailed processes. The estimators that we analyze here have been empirically explored in Blanchet and Liu (2007). Our contributions are summarized as follows.

1. Assuming mild conditions on the target set  $bA$  (refer to assumptions (A) and (B) in Section 3), we provide a state-dependent importance sampling estimator for  $u_b$  whose second moment can be shown to be of order  $O(u_b^2)$  as  $b \nearrow \infty$  (see Theorem 1). It is important to mention that the conditions that we impose on the set  $A$  are weaker than those known for the development of exact asymptotics (see Remark 2).
2. We provide new techniques to deal with the problem of finite termination time in first passage time simulation algorithms of heavy-tailed processes. Thus, we introduce a new class of simulation estimators whose running time can be controlled at an expense of a computable relative bias which can be reduced arbitrarily at the expense of a relatively small computational cost (see Section 6).
3. We introduce new techniques for asymptotic upper bounds of rare-event probabilities in multivariate heavy-tailed processes based on Lyapunov inequalities, which lie at the center of our technical arguments. Serving as an upper bound of the second moment of our importance sampling estimator, the Lyapunov function is also an upper bound of the rare-event probability squared, by Jensen's inequality.

It is worth pointing out that the specification of our importance sampling scheme depends only on the existence of a suitable open cone that does not contain  $A$ . Once the sampler is specified, we can use the same change of measure for any target set which is inside the set  $A$ . In contrast, in the light-tailed setting (see, for instance, Collamore (2002)), the design of efficient importance sampling schemes requires the evaluation of certain roots that are highly dependent on the geometry of  $A$ .

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 (2008)). Such changes of measure are parameterized by a few constants. In particular, at each step, we consider a mixture of a large increment that basically 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, we need to make sure that the Lyapunov inequality is satisfied. Now, such an inequality requires the choice of a convenient Lyapunov function which is obtained using heuristics based on a fluid-type analysis—a standard technique in heavy-tailed approximations which we briefly describe in Section 4 (see (5)). The heuristics are then made rigorous by 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.

It is important to emphasize that the implementation of the algorithm, which is discussed in Section 5, involves a rare-event estimation problem which corresponds to evaluating the probability that a large jump occurs in the next transition. This issue is not of concern in the

one-dimensional case given that large jump probabilities for a single increment can be easily computed. In the multidimensional case this computation is not immediate. Nevertheless, the evaluation of such large jump probabilities involves a static problem which is simpler than the original first passage time problem and efficient algorithms can be designed on a case-by-case basis. In addition, as we discuss in Section 5, we have built our algorithm in a flexible way so that we can take advantage of the shape of the increment distribution (for instance, by considering a transformation in polar coordinates) in order to efficiently compute large jump probabilities within a single transition of the random walk. When the increment distributions do not have enough structure, however, the issue of path generation leaves room for further investigation.

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. In Section 3 we describe the specific problem formulation and basic results on large deviations. The basic ideas behind our development and algorithm are described in Section 4. We deal with implementation issues (such as path generation) and related considerations in Section 5. A technique to control the bias of importance sampling estimators to a prescribed relative accuracy is discussed in Section 6.

### 2. State-dependent importance sampling and efficiency

We will 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  $\sigma$ -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,

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

where  $f(\cdot)$  is the density function of the increments and  $r(\cdot)$  will be chosen in the next section involving a mixture. In addition, we use the symbols  $P_w^Q$  and  $E_w^Q$  to denote the probability measure and 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  $E_w$  for the expectation).

Throughout the rest of the paper, we will write  $T_{A_b}$  to denote the first passage time of the underlying chain to the set  $A_b$ . More precisely,  $T_{A_b} = \inf\{n > 0 : W_n \in A_b\}$ . We use  $\{A_b : b \geq 0\}$  as generic notation for a family of rare sets. The subscript  $b$  is the so-called rarity parameter which will eventually be sent to  $\infty$ . Later on, we will focus on the family of rare sets  $A_b = \{bx : x \in A\}$ . 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

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

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 statistical efficiency properties, which we will measure in terms of the mean squared error. First, we will concentrate on the variance control. In particular, we will select  $r(\cdot)$  in order to achieve strong efficiency, that is,

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

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  $n$  i.i.d. 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  $b \geq 1$ )

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

This implies, as we indicated, that in order to achieve relative precision  $\varepsilon$  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  $A_b$  under the change of measure and the variate generation at each step. These issues will be discussed in future sections.

The following proposition provides the means to bound the second moment of state-dependent importance sampling estimators. See Blanchet and Glynn (2008).

**Proposition 1.** *Suppose that there exists a nonnegative function  $g_b(w): \mathcal{X} \rightarrow \mathbb{R}^+ \cup \{0\}$  such that*

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

for  $w \in \bar{A}_b$  and  $g_b(w) \geq \varepsilon$  for  $w \in A_b$ . Then

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

if  $w \in \bar{A}_b$ . Throughout the paper, we use  $\bar{A}$  to denote the complement of set  $A$ .

**Remark 1.** The bound in (1) is called a 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 we choose 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 a guidance for constructing  $g_b(\cdot)$ . Such an approach is pursued in Section 4.

### 3. Problem setting and intuition

Let  $(X_n: n \geq 1)$  be a sequence of i.i.d. regularly varying random vectors taking values in  $\mathbb{R}^d$ . We define multivariate regularly varying distribution as follows.

**Definition 1.** A random vector  $X$  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

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

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, we 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.  $L(tb)/L(b) \rightarrow 1$  as  $b \nearrow \infty$  for each  $t > 0$ ), then we say that  $\mu(\cdot)$  has (regularly varying) index  $\alpha$ .

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

$$\frac{P(\|X\|_2 > bx, X/\|X\|_2 \in \cdot)}{P(\|X\|_2 > b)} \rightarrow x^{-\alpha} P(\Theta \in \cdot) \tag{2}$$

in the sense of vague convergence in  $\mathbb{S}^{d-1}$  (see Kallenberg (1983, p. 32)).

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  $\nu$  degrees of freedom. In particular,  $X_1$  has density

$$f_{X_1}(y) = \frac{\kappa_d}{(1 + y^\top y)^{(\nu+d)/2}}$$

for an appropriate constant  $\kappa_d \in (0, \infty)$ . It follows that  $P(\|X_1\|_2 > t) = c_d t^{-\nu}(1 + o(1))$  as  $t \nearrow \infty$  for some  $c_d > 0$  and the associated limiting regularly varying measure for  $X_1$  satisfies

$$\mu(A) = \lim_{t \rightarrow \infty} \frac{P(X_1 \in tA)}{P(\|X_1\|_2 > t)} = \int_A \frac{\kappa_d/c_d}{(y^\top y)^{(\nu+d)/2}} dy.$$

The fact that  $\mu(\cdot)$  has a density with respect to the Lebesgue measure in  $\mathbb{R}^d \setminus \{0\}$  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.* (2003). In the context of (2),  $\Theta$  is uniformly distributed.

Returning to our discussion on large deviations for multidimensional regularly varying random walks, let us assume that  $E X_1 = \eta \neq 0$  and define  $S_n = X_1 + \dots + X_n$ . Throughout the rest of the paper, we will 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  $bA = \{y : y = ba, a \in A\}$ . The family of sets  $bA$  indexed by  $b$  corresponds to  $A_b$  in Section 2. 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  $b \nearrow \infty$ .

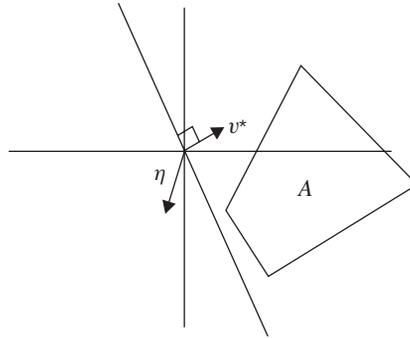


FIGURE 1: Diagram illustrating assumption (A) for a two-dimensional random walk.

Given that the random walk has drift  $\eta$ , it is not difficult to see geometrically that some conditions must be imposed on  $A$  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 happen that when one of the faces of the polyhedron is parallel to  $\eta$ , 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  $v_1^*, \dots, v_m^* \in \mathbb{R}^d$ ,  $\delta^* > 0$  such that  $\|v_j^*\|_2 = 1$  and  $\eta^\top v_j^* < -\delta^*$  and, for all  $z \in A$ , we have  $z^\top v_j^* \geq \delta^*$  for at least one  $v_j^*$ .

Figure 1 depicts the situation described by (A).

As indicated in Figure 1, the vectors  $v_k^*$  point in the ‘direction’ where the set  $A$  is located; this direction cannot be parallel to the drift of the process, otherwise the set  $A$  would eventually be attainable with probability 1.

In order to avoid situations where the target set is ‘too thin’, for instance, when  $A$  lives in a lower-dimensional manifold, we impose the following assumption.

- (B) Assume that  $\mu(\cdot)$  has regularly varying index  $\alpha > 1$  and that  $A$  contains an open subset,  $A^\circ$ , 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 in 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 0 of  $u_b(0)$ . The following result follows as an easy consequence of Hult *et al.* (2005, Theorem 3.1).

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

$$u_b(0) \geq cb \mathbb{P}(\|X\|_2 > b)$$

for all  $b \geq 1$ .

**Remark 2.** Note that we only require an asymptotic lower bound of the form  $cb \mathbb{P}(\|X\|_2 > b)$ , which is guaranteed under our assumptions.

The most interesting portion of the large deviations approximation for  $u_b(0)$  involves developing an upper bound. The strategy that we will pursue in the next section allows us to obtain the desired upper bound by showing that the second moment of a suitable importance sampling estimator is of order  $O(b^2 \mathbb{P}(X_1 \in bA)^2)$  as  $b \nearrow \infty$ . This will imply strong efficiency and also that  $u_b(0) = O(b \mathbb{P}(X_1 \in bA))$  as  $b \nearrow \infty$ .

**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 indicated before, in the heavy-tailed case the rare event is caused by a single large jump, while all the other increments prior to the jump evolve 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 the other one that basically evolves 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 set 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 may be several directions that one can take simultaneously.

The vectors  $v_1^*, \dots, v_m^*$  will be used as possible directions. Our family of changes of measure is constructed so that, with some probability, which may depend on the current position, the random walk reaches  $A$  in the next step. More precisely, given the current position of the walk is  $s$  and a constant  $a \in (0, 1)$ , we define

$$C_i^a(s, b) = \{x : v_i^{*\top} x > a(\delta^* b - v_i^{*\top} s)\}.$$

The parameter  $a \in (0, 1)$  helps to over sample those paths that take several jumps to reach  $A$ . We propose sampling the next increment according to a mixture density of the form

$$q_{X|s}(x) = p(s) \frac{f_X(x) \mathbf{1}(x \in \bigcup_{i=1}^m C_i^a(s, b))}{\mathbb{P}(X \in \bigcup_{i=1}^m C_i^a(s, b))} + (1 - p(s)) f_X(x), \tag{3}$$

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

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

Note that

$$q_{X|s}(x) = [p(s) + (1 - p(s)) P(s)] \frac{f_X(x) \mathbf{1}(x \in \bigcup_{i=1}^m C_i^a(s, b))}{P(s)} + (1 - p(s)) f_X(x) \mathbf{1}\left(x \in \bigcap_{i=1}^m \bar{C}_i^a(s, b)\right).$$

We find that the likelihood ratio corresponding to (3) takes the form (using notation consistent with the statement of Proposition 1)

$$\begin{aligned}
 r(s, x + s) &:= \frac{f_X(x)}{q_{X|s}(x)} \\
 &= \frac{P(s)}{(p(s) + (1 - p(s))P(s))} \mathbf{1}\left(x \in \bigcup_{i=1}^m C_i^a(s, b)\right) \\
 &\quad + \frac{1}{(1 - p(s))} \mathbf{1}\left(x \in \bigcap_{i=1}^m \bar{C}_i^a(s, b)\right). \tag{4}
 \end{aligned}$$

We will 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

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

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, we also need to propose a parametric expression for the candidate Lyapunov function. As discussed in Remark 1, if the proposed importance sampler is close enough to the zero-variance change of measure, we expect the Lyapunov inequality to be satisfied by a function that behaves like (or is an upper bound for)  $(P_s^2(T_{b,A} < \infty) : s \in \mathbb{R}^d)$ . Therefore, a natural strategy is to obtain a heuristic guess for  $P_s(T_{b,A} < \infty)$  and use this guess to propose an explicit form for the Lyapunov function which will then be tested rigorously. A rough analysis, known as ‘fluid analysis’, suggests that

$$\begin{aligned}
 P_s(T_{b,A} < \infty) &\approx \int_0^\infty P(X + s + \eta t \in bA) dt \\
 &\leq \sum_{i=1}^m \int_0^\infty P(v_i^{*\top}(X + s + \eta t) \geq b\delta^*) dt \\
 &= \sum_{i=1}^m \frac{1}{-v_i^{*\top}\eta} G_i(b\delta^* - v_i^{*\top}s), \tag{5}
 \end{aligned}$$

where

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

The idea behind the ‘fluid analysis’ is that, prior to the occurrence 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 given time,  $t$ , which takes the process to the target set. The approximation is constructed by integrating over  $t$ , thereby arriving at (5).

So, we propose our Lyapunov function to be

$$g_b(s) = \min(c_g h_b(s)^2, 1) \quad \text{for some } c_g > 0, \tag{6}$$

where

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

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

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

$$1 \geq E^q \left( \frac{g_b(s+X)}{g_b(s)} r(s, s+X)^2 \right) = E \left( \frac{g_b(s+X)}{g_b(s)} r(s, s+X) \right), \tag{7}$$

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

We will establish bound (7) first on  $\{s : g_b(s) < 1\}$ . Note that if  $g_b(s) = 1$ , inequality (7) is satisfied by selecting  $p(s) = 0$ . If  $g_b(s) < 1$ ,

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

Let us define

$$\begin{aligned} J_1 &= E \left( \frac{g_b(s+X)}{g_b(s)} ; X \in \bigcup_{i=1}^m C_i^a(s, b) \right) \frac{P(s)}{p(s) + (1-p(s))P(s)}, \\ J_2 &= E \left( \frac{g_b(s+X)}{g_b(s)} ; X \in \bigcap_{i=1}^m \bar{C}_i^a(s, b) \right) \frac{1}{1-p(s)}. \end{aligned} \tag{8}$$

An immediate upper bound is obtained for  $J_1$ , namely,

$$J_1 \leq \frac{P(s)^2}{c_g h_b(s)^2 p(s)}, \tag{9}$$

on the set that  $g_b(s) < 1$ . To handle  $J_2$ , the idea is to use a Taylor expansion with remainder. Recall that if  $f(\cdot)$  is absolutely continuous then by writing  $X = (X^{(1)}, \dots, X^{(d)})$  we have

$$\begin{aligned} & f(s_1 + X^{(1)}, \dots, s_d + X^{(d)}) - f(s_1, \dots, s_d) \\ &= \int_0^1 \frac{\partial f}{\partial s_1}(s_1 + X^{(1)}u, s_2, \dots, s_d) X^{(1)} du \\ &\quad + \int_0^1 \frac{\partial f}{\partial s_2}(s_1 + X^{(1)}, s_2 + X^{(2)}u, s_3, \dots, s_d) X^{(2)} du + \dots \\ &\quad + \int_0^1 \frac{\partial f}{\partial s_d}(s_1 + X^{(1)}, \dots, s_{d-1} + X^{(d-1)}u, s_d + X^{(d)}u) X^{(d)} du \end{aligned}$$

$$\begin{aligned}
 &= X^{(1)} \mathbb{E} \left( \frac{\partial f}{\partial s_1} (s_1 + X^{(1)}U, s_2, \dots, s_d) \mid X \right) \\
 &\quad + X^{(2)} \mathbb{E} \left( \frac{\partial f}{\partial s_2} (s_1 + X^{(1)}, s_2 + X^{(2)}U, s_3, \dots, s_d) \mid X \right) + \dots \\
 &\quad + X^{(d)} \mathbb{E} \left( \frac{\partial f}{\partial s_d} (s_1 + X^{(1)}, s_2 + X^{(2)}, \dots, s_d + X^{(d)}U) \mid X \right),
 \end{aligned}$$

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

$$g_b(s + X) = g_b(s) + \mathbb{E} \left( \sum_{j=1}^d \frac{\partial g_b}{\partial s_j} (s + D_j X) X^{(j)} \mid X \right), \tag{10}$$

where  $D_j$  denotes an appropriate diagonal matrix, in particular the  $j - 1$  first components in its main diagonal are 1s, the  $j$ th component in the main diagonal is  $U$ , and the remaining components of the matrix are 0s. The strategy then consists in using a bound of the form

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

for some  $\gamma_1 > 0$ , where the dot denotes the inner product, and noting that

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

A key observation is that the right-hand side of the previous inequality is a negative quantity which, combined with (10), can be used to develop inequality (7) via (8).

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

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

$$\mathbb{E} \left( \sum_{j=1}^d \partial_j g_b(s + D_j X) X^{(j)}; X \in \bigcap_{i=1}^m \bar{C}_i^a(s, b) \right) \leq \gamma_1 \nabla g_b(s) \cdot \eta,$$

where  $\partial_j g_b = \partial g_b / \partial s_j$ .

*Proof.* First, observe that if  $g_b(s) < 1$  then

$$\partial_j g_b(s) = 2h_b(s) \sum_{i=1}^m \mathbb{P}(v_i^{*\top} X > b\delta^* - v_i^{*\top} s) v_i^{*(j)}.$$

Furthermore, if  $X \in \bigcap_{i=1}^m \bar{C}_i^a(s, b)$ ,

$$\mathbb{P}(v_i^{*\top} X > b\delta^* - v_i^{*\top} s - v_i^{*\top} D_j X) \leq \mathbb{P}(v_i^{*\top} X > (b\delta^* - v_i^{*\top} s)(1 - a)) \tag{12}$$

for  $1 \leq i \leq m$ ; likewise

$$h_b(s + D_j X) \leq \sum_{i=1}^m G_i((b\delta^* - v_i^{*\top} s)(1 - a)). \tag{13}$$

The condition  $g_b(s) < 1$  implies that, for all  $1 \leq i \leq m$ ,  $b\delta^* - v_i^{*\top} s > G_i^{-1}(1/c_g^{1/2}) \nearrow \infty$  as  $c_g \nearrow \infty$ . As a consequence, if  $c_g$  is sufficiently large and  $X \in \bigcap_{i=1}^m \bar{C}_i^a(s, b)$ , then  $g_b(s + D_j X) < 1$  for all  $1 \leq j \leq m$  and, therefore,

$$\partial_j g_b(s + D_j X) X^{(j)} = 2h_b(s + D_j X) \sum_{i=1}^m \mathbf{P}(v_i^{*\top} X > b\delta^* - v_i^{*\top} s - v_i^{*\top} D_j X) v_i^{*(j)} X^{(j)}.$$

Moreover, because the  $G_i$ s and the distribution of the random variables  $v_i^{*\top} X$  are all regularly varying, then, by Karamata’s theorem, it follows from (12) and (13) that there exists a constant  $\gamma_0 \in (0, \infty)$  such that

$$\frac{h_b(s + D_j X)}{h_b(s)} \leq \gamma_0 \quad \text{and} \quad \frac{\mathbf{P}(v_i^{*\top} X > (b\delta^* - v_i^{*\top} s)(1 - a))}{\mathbf{P}(v_i^{*\top} X > b\delta^* - v_i^{*\top} s)} \leq \gamma_0 \tag{14}$$

for  $1 \leq i \leq m$ . On the other hand, because

$$\nabla h_b(s) \cdot \eta \leq -\delta^* \sum_{i=1}^m \mathbf{P}(v_i^{*\top} X > b\delta^* - v_i^{*\top} s) < 0$$

and the bounds in (14), we have

$$\frac{\sum_{j=1}^d \partial_j g_b(s + D_j X) X^{(j)} \mathbf{1}(X \in \bigcap_{i=1}^m \bar{C}_i^a(s, b))}{-\nabla g_b(s) \cdot \eta} \leq \frac{\gamma_0^2}{\delta^*} \sum_{j=1}^d \sum_{i=1}^m |v_i^{*(j)} X^{(j)}|.$$

Therefore, the left-hand side of the previous display is uniformly integrable as  $b \nearrow \infty$ . Moreover, note that

$$\frac{\sum_{j=1}^d \partial_j g_b(s + D_j X) X^{(j)} \mathbf{1}(X \in \bigcap_{i=1}^m \bar{C}_i^a(s, b))}{\nabla g_b(s) \cdot \eta} \rightarrow \frac{\sum_{j=1}^d \sum_{i=1}^m v_i^{*(j)} X^{(j)}}{\sum_{i=1}^m v_i^{*\top} \eta}$$

almost surely as  $c_g \nearrow \infty$  (as mentioned before,  $b\delta^* - v_i^{*\top} s > G_i^{-1}(1/c_g^{1/2}) \nearrow \infty$  as  $c_g \nearrow \infty$ ). Therefore, it follows from the dominated convergence theorem that there is a  $\gamma_1 \in (0, 1)$  such that

$$\mathbf{E} \left( \sum_{j=1}^d \partial_j g_b(s + D_j X) X^{(j)}; X \in \bigcap_{i=1}^m \bar{C}_i^a(s, b) \right) \leq \gamma_1 \nabla g_b(s) \cdot \eta,$$

as claimed.

**Remark 3.** The gray area in Figure 2 shows the form of the region where  $g_b(s) < 1$ .

It follows from Lemma 1 combined with (9) and the Taylor development applied to (8) that, if  $g_b(s) < 1$ ,

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

We select

$$p(s) = \min \left( \theta \frac{\mathbf{P}(s)}{h_b(s)}, \frac{1}{2} \right) \mathbf{1}(g_b(s) < 1) \quad \text{for some } \theta > 0. \tag{15}$$

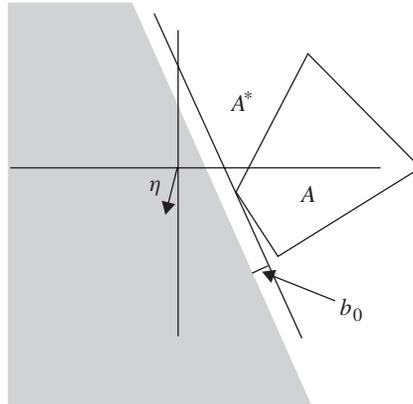


FIGURE 2: Two-dimensional diagram illustrating the region where Lemma 1 is satisfied. Here we define  $A^* = \{s : g(s) = 1\}$ , and  $g(s) < 1$  implies that  $b\delta^* - v_1^{*\top}s > b_0$ .

On the set that  $g_b(s) < 1$ , we have

$$J_1 + J_2 \leq 1 + 2\theta \frac{P(s)}{h_b(s)} + \frac{P(s)}{c_g \theta h_b(s)} - 2\gamma_1 \delta^* \frac{\sum_{i=1}^m P(v_i^{*\top} X > b\delta^* - v_i^{*\top} s)}{h_b(s)}. \tag{16}$$

Now, our task is to appropriately select  $\theta$  and  $c_g$  in order to make the right-hand side of the previous inequality less than 1. First we collect the following lemma.

**Lemma 2.** *There exists  $\gamma_2 \in (0, \infty)$  such that*

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

for all  $s \in \mathbb{R}^d$ .

*Proof.* By the union bound we obtain

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

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

Applying the previous result to (16) 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 the following proposition.

**Proposition 3.** *Given the family of importance samplers defined via (4), with  $p(s)$  defined according to (15) and  $g_b(s)$  satisfying (6), we can compute  $\theta$  and  $c_g$  such that the Lyapunov inequality (7) is satisfied for all  $s \in \mathbb{R}^d$ . In particular, any pair  $(\theta, c_g)$  such that  $\theta \leq \gamma_1 \delta^* / (2\gamma_2)$  and  $c_g \geq \gamma_2 / (\theta \gamma_1 \delta^*)$  are valid choices.*

*Proof.* First, for  $s$  such that  $g_b(s) < 1$ , we have, combining the bound in (16) and Lemma 2,

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

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

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

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 will address this problem in a later section.

**Algorithm 1.** (*Basic algorithm.*) Set  $b > 0$  and fix  $a \in (0, 1)$ . Initialize  $s = 0$ , REACH = 0, and  $Z = 1$ . Assume that  $c_g$  and  $\theta$  have been selected in order to satisfy (1) and set  $g_b(w)$  according to (6).

*Step 1.* **While** REACH = 0.

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

**Else set**

$$p \leftarrow \min\left(\theta \frac{P(s)}{h_b(s)}, \frac{1}{2}\right) \mathbf{1}(s : g_b(s) < 1)$$

and sample  $X$  as follows. With probability  $p$ , generate  $X$  with law  $\mathcal{L}(X \mid X \in \bigcup_{i=1}^m C_i^a(s, b))$ ; with probability  $1 - p$ , sample  $X$  from the nominal distribution. Then, update

$$Z_b \leftarrow Z_b \left\{ \frac{P(s) \mathbf{1}(X \in \bigcup_{i=1}^m C_i^a(s, b))}{p + (1 - p) P(s)} + \frac{\mathbf{1}(X \in \bigcap_{i=1}^m \bar{C}_i^a(s, b))}{1 - p} \right\}.$$

**End if**

(b) Update

$$s \leftarrow s + X.$$

(c) **If**  $s \in bA$  then

$$\text{REACH} \leftarrow 1.$$

**End if**

**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 Algorithm 1 has bounded relative error.*

*Proof.* The result follows by a straightforward application of the Lyapunov inequality. In particular, we obtain

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

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**

A couple of problems involving the implementation of the algorithm 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  $P(X \in \bigcup_{i=1}^m C_i^a(s, b))$  involved in computing  $r(\cdot)$ . Moreover, such a procedure can often be designed, parametrically in  $s$  and  $b$ , so that the acceptance probability remains bounded away from 0 uniformly over  $s$  and  $b > 0$ . This is explained in Blanchet and Glynn (2008) in the one-dimensional case, but the ideas can be adapted to higher dimensions. The most interesting problem arises when calculating  $P(X \in \bigcup_{i=1}^m C_i^a(s, b))$ , which is required to output the corresponding estimator. This is also a rare-event estimation, but it is of finite dimension. The design of efficient estimation procedures for computing these finite-dimensional probabilities are more standard and can be handcrafted in a case-by-case basis.

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 the particular problem structure to compute  $P(X \in \bigcup_{i=1}^m C_i^a(s, b))$ . Consider any parametric family of sets  $\{B(s, b) : s \in \mathbb{R}^d, b > 0\}$  such that  $\bigcup_{i=1}^m C_i^a(s, b) \subseteq B(s, b)$  and with the property that there exists a constant  $\tilde{\delta} > 0$  such that

$$\tilde{\delta} P(X \in B(s, b)) \leq P\left(X \in \bigcup_{i=1}^m C_i^a(s, b)\right), \tag{17}$$

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

$$\begin{aligned} \frac{f_X(x)}{\tilde{q}_{X|s}(x)} &:= \tilde{r}(s, s + x) \\ &= \frac{\tilde{P}(s)}{\tilde{p}(s) + (1 - \tilde{p}(s))\tilde{P}(s)} \mathbf{1}(x \in B(s, b)) + \frac{1}{1 - \tilde{p}(s)} \mathbf{1}(x \in B(s, b)^c). \end{aligned} \tag{18}$$

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

We 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 (6) can be used to construct a Lyapunov function with a suitable selection of the parameter  $c_g$ . Indeed, we can define

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

where

$$\begin{aligned} \tilde{J}_1 &= \mathbb{E}\left(\frac{g_b(s+X)}{g_b(s)}; X \in B(s, b)\right) \frac{\mathbb{P}(X \in B(s, b))}{\tilde{p}(s) + (1 - \tilde{p}(s))\tilde{P}(s)} \\ &\leq \mathbb{E}\left(\frac{g_b(s+X)}{g_b(s)}; X \in B(s, b)\right) \frac{\mathbb{P}(X \in B(s, b))}{\tilde{p}(s)}, \\ \tilde{J}_2 &= \mathbb{E}\left(\frac{g_b(s+X)}{g_b(s)}; X \notin B(s, b)\right) \frac{\mathbb{P}(X \notin B(s, b))}{1 - \tilde{p}(s)}, \end{aligned}$$

and

$$\tilde{p}(s) = \min\left(\theta \frac{\mathbb{P}(X \in B(s, b))}{h_b(s)}, \frac{1}{2}\right) \mathbf{1}(s: g_b(s) < 1). \tag{19}$$

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

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

which corresponds to (9). Similarly, the analysis of  $\tilde{J}_2$  follows exactly the same steps as that of  $J_2$  in Section 4, thereby obtaining the inequality

$$\begin{aligned} \tilde{J}_1 + \tilde{J}_2 &\leq 1 + 2\theta \frac{\mathbb{P}(X \in B(s, b))}{h_b(s)} + \frac{\mathbb{P}(X \in B(s, b))}{c_g \theta h_b(s)} \\ &\quad - 2\gamma_1 \delta^* \frac{\sum_{i=1}^m \mathbb{P}(v_i^{*\top} X > b\delta^* - v_i^{*\top} s)}{h_b(s)}. \end{aligned} \tag{20}$$

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

**Proposition 4.** *Given the family of importance samplers defined via (18), with  $\tilde{p}(s)$  defined according to (19), we can compute  $\theta$  and  $c_g$  such that  $\tilde{J}_1 + \tilde{J}_2 \leq 1$  for all  $s \in \mathbb{R}^d$ . In particular, any pair  $(\theta, c_g)$  such that  $\theta \leq \gamma_1 \tilde{\delta} \delta^* / (2\gamma_2)$  and  $c_g \geq \gamma_2 / (\theta \gamma_1 \tilde{\delta} \delta^*)$  are valid choices.*

*Proof.* First, for  $s \notin \tilde{R}_b$ , we have, combined with the development behind (20), Lemma 2, and the definition of  $\tilde{\delta}$ ,

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

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  $\mathbb{P}(X \in B(s, b))$  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 a situation, evaluation of constants such as  $\mathbb{P}(X \in \bigcup_{i=1}^m C_i^a(s, b))$  is not entirely straightforward. However, it is not difficult to construct  $B(s, b)$ , suitably expressed in polar coordinates, which contains  $\bigcup_{i=1}^m C_i^a(s, b)$  and satisfies (17); a diagram indicating the structure of  $B(s, b)$  in a two-dimensional case is given in Figure 3. The evaluation of  $\mathbb{P}(X \in B(s, b))$  and the corresponding path generation can then be easily done in polar coordinates. See Blanchet and Liu (2007) for details on this implementation.

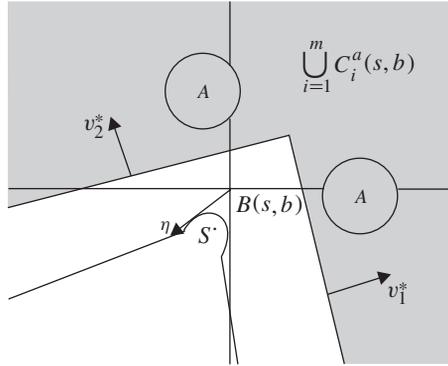


FIGURE 3: Illustration of a set  $B(s, b)$  that can be easily evaluated in polar coordinates for a bivariate  $t$  distribution.

### 6. Introducing a controlled bias

The analysis of Section 4 allows us to control the variance of the estimator given by Algorithm 1. However, nothing has been said about the termination time of the algorithm. In fact, even in the one-dimensional case, the zero-variance change of measure might have a termination time with infinite mean. This occurs when the increments have infinite variance. Therefore, we will develop a criterion that allows us to safely stop the path generation by introducing a controlled bias in our procedure. We will consider biased estimators for which the relative bias can be controlled in a suitable way. In particular, we will construct a set  $B_{\beta,b}$  depending on  $b$  and a suitable parameter  $\beta$  so that  $T_{B_{\beta,b}} < \infty$  with probability 1 under the change of measure. We note that

$$P_0(T_{bA} < \infty) = \nu_{\beta,b}(0) + P_0(T_{bA} < \infty, T_{B_{\beta,b}} < T_{bA}),$$

where  $\nu_{\beta,b}(0) = P_0(T_{bA} < T_{B_{\beta,b}})$ , and we will find a function  $\gamma(\cdot)$  such that

$$\frac{P_0(T_{bA} < \infty, T_{B_{\beta,b}} < T_{bA})}{P_0(T_{bA} < \infty)} \leq \gamma(\beta) \tag{21}$$

with  $\gamma(\beta) = O(\beta^{-p})$  for some  $p > 0$  as  $\beta \nearrow \infty$ . It will follow easily that, with the aid of a strongly efficient estimator for  $\nu_{\beta,b}(0)$  and a bound such as (21), we can efficiently estimate  $P_0(T_{bA} < \infty)$  with a controlled relative error. Then, to carry over our strategy we must:

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

Computing the function  $\gamma(\beta)$ . Jensen’s inequality combined with Proposition 1 yields

$$\begin{aligned} P_0(T_{bA} < \infty, T_{B_{\beta,b}} < T_{bA}) &= E_0(P(T_{bA} < \infty \mid S_{T_{B_{\beta,b}}}); T_{B_{\beta,b}} < T_{bA}) \\ &\leq \varepsilon^{-1/2} E(g_b(S_{T_{B_{\beta,b}}})^{1/2}; T_{B_{\beta,b}} < T_{bA}) \\ &\leq \varepsilon^{-1/2} \sup_{s \in B_{\beta,b}} g_b(s)^{1/2}. \end{aligned}$$

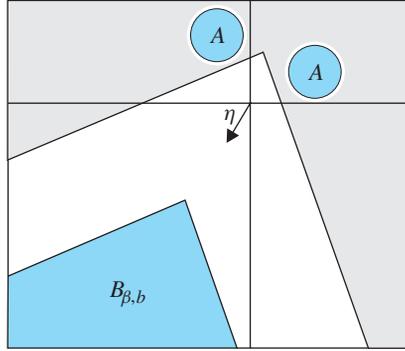


FIGURE 4: Illustration of the location of the set  $B_{\beta,b}$ .

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

$$\frac{P_0(T_{bA} < \infty, T_{B_{\beta,b}} < T_{bA})}{P_0(T_{bA} < \infty)} \leq \frac{\sup_{s \in B_{\beta,b}} g_b(s)^{1/2}}{v_{\beta,b}(0)} = c_0 \beta^{-p} \tag{22}$$

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

*Selecting  $B_{\beta,b}$ .* Our choice of  $B_{\beta,b}$  takes advantage of the fact that, prior to the big jump, the process drifts according to its unconditional mean,  $\eta$ . We then construct a suitable set  $B_{\beta,b}$  that intersects the fluid path,  $\{\eta t : t > 0\}$ . The parameter  $\theta$  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  $bA$  in an expected time of order  $O(b)$ . The selection of  $\theta$  might come at a price of potentially increasing the value of  $c_g$  (as indicated in the constraints 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

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

Figure 4 shows an illustration of the set  $B_{\beta,b}$ .

*Providing a bound for  $E_s^q T_{B_{\beta,b}}$ .* The following result, which follows exactly the same steps as in the first part of the proof in Theorem 11.3.4 of 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 we can find a nonnegative function  $\varrho(s)$  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  $s \notin B_{\beta,b}$ .

In order to construct the Lyapunov function  $\varrho(\cdot)$ , we observe that the boundary of the set  $B_{\beta,b}$  is

$$\partial B_{\beta,b} = \left\{ s : \max_{i=1}^m \left( v_i^{*\top} s - \frac{v_i^{*\top} \eta \beta b}{\|\eta\|_2} \right) = \delta^* b \right\}.$$

On the other hand, as mentioned before, we will choose  $\theta > 0$  sufficiently small so that  $E_s^q X \approx \eta$ . Therefore, looking at Figure 4 we expect the time it takes for the process to hit  $B_{\beta,b}$  starting at  $s \in \bar{B}_{\beta,b}$  to be roughly  $\max_{i=1}^m (v_i^{*\top} s - v_i^{*\top} \eta \beta b / \|\eta\|_2) - \delta^* b$ . Our Lyapunov function  $\varrho(\cdot)$  is constructed based on the contour plots of the previous expression as a function of  $s$ . However, since it is convenient to work with a smooth function,  $\varrho(\cdot)$ , instead of working with the maximum of affine functions as indicated above, we consider

$$\varrho(s) = \kappa b \log \left( 1 + \sum_{i=1}^m \exp \left( \frac{v_i^{*\top} s}{\kappa b} \right) \right)$$

for  $\kappa > 0$ . The following result summarizes some useful properties of  $\varrho(\cdot)$  and its derivatives; the proof is straightforward and therefore omitted.

**Lemma 4.** *Define*

$$w_j(s) = \frac{\exp(v_i^{*\top} s / (\kappa b))}{1 + \sum_{i=0}^m \exp(v_i^{*\top} s / (\kappa b))}.$$

Then

- (i)  $\max_{i=1}^m (v_i^{*\top} s) \leq \varrho(s) \leq \max_{i=1}^m (v_i^{*\top} s) + \kappa b \log(m + 1)$ ;
- (ii)  $(\partial \varrho)(s) = \sum_{j=0}^m v_j^* w_j(s)$ .

The following result shows that  $\kappa > 0$  can be selected for  $\varrho(\cdot)$  to satisfy the Lyapunov inequality in Lemma 1.

**Proposition 5.** *We can choose  $\theta > 0$  small enough in (15) so that the Lyapunov inequality in Lemma 3 holds for an appropriate selection of  $\rho > 0$ . As a consequence,  $E_0^q \min(T_{B_{\beta,b}}, T_{bA}) \leq \kappa b \log(m + 1) / \rho$ , where  $\kappa = c\beta$  for some sufficiently large  $c$ .*

*Proof.* First note the nonnegativity of  $\varrho$  is satisfied. Using Taylor’s expansion with remainder, as in Section 4 and Lemma 4, we obtain

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

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

$$\begin{aligned} & p(s) E \left( \sum_{j=1}^m w_j(s + DX) X^\top v_j^* \mid X \in \bigcup_{i=1}^m C_i^a(s, b) \right) \\ & \leq p(s) E \left( \left| \max_{j=1}^m X^\top v_j^* \right| \mid X \in \bigcup_{i=1}^m C_i^a(s, b) \right) < \frac{\delta^*}{8}. \end{aligned}$$

Now note that, for any  $\varepsilon_0 > 0$ , there exists  $\varepsilon > 0$  such that

$$|w_j(s + DX) - w_j(s)| \leq \varepsilon_0$$

whenever  $\|X\|_2 \leq \varepsilon\kappa b$ . Furthermore, by choosing  $\kappa = c\beta$  with sufficiently large  $c$  we can guarantee that  $w_j(s) > 1/(2m)$  uniformly over  $s \notin B_{\beta,b}$ . Given that  $E X^\top v_j^* \leq -\delta^*$ , we can then conclude that, for  $b \geq 1$ ,

$$(1 - p(s)) E \left( \sum_{j=1}^m w_j(s + DX) X^\top v_j^*; \|X\|_2 \leq \varepsilon\kappa b \right) \leq -\frac{\delta^*}{3},$$

$$(1 - p(s)) E \left( \sum_{j=1}^m w_j(s + DX) |X^\top v_j^*|; \|X\|_2 > \varepsilon\kappa b \right) \leq \frac{\delta^*}{8}.$$

This yields

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

with  $\rho = \delta^*/12$ , and the result follows.

We close the paper with a description of the algorithm and a summary of how the overall relative mean squared error can be reduced to a prescribed relative accuracy with high confidence at a computational expense that scales graciously as  $b \nearrow \infty$ .

**Algorithm 2.** (*Bias controlled algorithm.*) Set  $b > 0$  and fix  $a \in (0, 1)$ . Initialize  $s = 0$ , REACH = 0, and  $Z_\beta(b) = 1$ . Let  $\theta$  be selected according to Proposition 5, and let  $g_b(s)$  be defined as in (6) with an appropriate choice of  $c_g$  as in Proposition 3.

*Step 1.* **While** REACH = 0.

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

**Else set**

$$p \leftarrow \min \left( \theta \frac{P(s)}{h_b(s)}, \frac{1}{2} \right) \mathbf{1}(s : g_b(s) < 1)$$

and sample  $X$  as follows. With probability  $p$ , generate  $X$  with law  $\mathcal{L}(X | X \in \bigcup_{i=1}^m C_i^a(s, b))$ ; with probability  $1 - p$ , sample  $X$  from the nominal distribution. Then, update

$$Z_\beta(b) \leftarrow Z_\beta(b) \left\{ \frac{P(s) \mathbf{1}(X \in \bigcup_{i=1}^m C_i^a(s, b))}{p + (1 - p) P(s)} + \frac{\mathbf{1}(X \in \bigcap_{i=1}^m \bar{C}_i^a(s, b))}{1 - p} \right\}.$$

**End if**

(b) **Update**

$$s \leftarrow s + X.$$

(c) **If**  $s \in bA \cup B_{\beta,b}$  then

$$\text{REACH} \leftarrow 1.$$

**End if**

**Loop**

*Step 2.* **Return**  $Z_\beta(b) \leftarrow Z_\beta(b) \mathbf{1}(s \in bA)$ .

**Theorem 2.** Let  $Z_\beta^{(i)}(b)$ ,  $i = 1, \dots, n$ , be i.i.d. copies of  $Z_\beta(b)$ , generated according to Algorithm 2,

$$Y_{n,\beta}(b) = \frac{1}{n} \sum_{i=1}^n Z_\beta^{(i)}(b).$$

Then the following statements hold.

1. There exists  $c_1$  such that

$$\sup_{b \geq 1} \frac{E^q Z_\beta(b)^2}{v_{\beta,b}(0)^2} \leq c_1.$$

2. Moreover, for any  $\varepsilon > 0$ ,

$$P(|Y_{n,\beta}(b) - u_b(0)| \geq \varepsilon u_b(0)) \leq \frac{c_1}{(\varepsilon - c_0\beta^{-p})^2 n}.$$

3. Consequently, given any  $\tilde{\varepsilon} > 0$  and  $\delta > 0$ , we can choose  $\kappa_0, c_0, \beta > (2c_0/\tilde{\varepsilon})^{1/p}$  and  $n = \kappa_0 \tilde{\varepsilon}^{-2} \delta^{-1}$  so that

$$P(|Y_{n,\beta}(b) - u_b(0)| \geq \tilde{\varepsilon} u_b(0)) \leq \delta.$$

This implies that  $n = O(\tilde{\varepsilon}^{-2} \delta^{-1})$  replications suffice to obtain an estimator with  $\tilde{\varepsilon}$  relative error and  $(1 - \delta)100\%$  confidence, and the expected number of generated increments required for each replication is not more than  $\varrho(0) = O(\beta b)$ .

*Proof.* Note that  $Z_\beta(b)$  is stochastically dominated by  $Z_b$  simulated according to Algorithm 2 (this is because  $Z_b > 0$  and  $Z_\beta(b) \geq 0$ , and we can couple the simulation output so that if  $Z_\beta(b) > 0$  then  $Z_b = Z_\beta(b)$ ). Therefore,

$$E^q (Z_\beta^{(i)}(b))^2 \leq E^q Z_b^2 \leq g_b(0).$$

Together with (22), the bound on the coefficient of variation comes immediately. The second statement is obtained as follows. Given any  $\tilde{\varepsilon} > 0$ ,

$$\begin{aligned} P(|Y_{n,\beta}(b) - u_b(0)| \geq \tilde{\varepsilon} u_b(0)) &\leq P(|Y_{n,\beta}(b) - v_{\beta,b}(0)| + |u_b(0) - v_{\beta,b}(0)| \geq \tilde{\varepsilon} u_b(0)) \\ &\leq P(|Y_{n,\beta}(b) - v_{\beta,b}(0)| \geq (\tilde{\varepsilon} - \gamma(\beta))u_b(0)) \\ &\leq \frac{c_1}{(\tilde{\varepsilon} - \gamma(\beta))^2 n} \\ &\leq \frac{c_1}{(\tilde{\varepsilon} - c_0\beta^{-p})^2 n}. \end{aligned}$$

Together with the results in Proposition 5, the last statement is also straightforward by plugging in  $\beta > (2c_0/\tilde{\varepsilon})^{1/p}$ ,  $n = \kappa_0 \tilde{\varepsilon}^{-2} \delta^{-1}$ , and choosing large enough  $\kappa_0$ .

### Acknowledgements

This research was partially supported by NSF grants 0846816 and 0902075, and by the Institute of Education Sciences, US Department of Education, under grant R305D100017.

## References

- ASMUSSEN, S. (2003). *Applied Probability and Queues*, 2nd edn. Springer, New York.
- BLANCHET, J. AND GLYNN, P. (2008). Efficient rare event simulation for the maximum of heavy-tailed random walks. *Ann. Appl. Prob.* **18**, 1351–1378.
- BLANCHET, J. H. AND LIU, J. (2007). Rare-event simulation for a multidimensional random walk with  $t$  distributed increments. In *Proc. 39th Conf. Winter Simulation*, IEEE, pp. 395–402.
- BLANCHET, J. H. AND LIU, J. (2008). State-dependent importance sampling for regularly varying random walks. *Adv. Appl. Prob.* **40**, 1104–1128.
- BLANCHET, J., GLYNN, P. AND LIU, J. (2007). Fluid heuristics, Lyapunov bounds and efficient importance sampling for a heavy-tailed G/G/1 queue. *Queueing Systems* **57**, 99–113.
- COLLAMORE, J. (2002). Importance sampling techniques for the multidimensional ruin problem for general Markov additive sequences of random vectors. *Ann. Appl. Prob.* **12**, 382–421.
- 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 MCNEIL, A. (2003). Modeling dependence with copulas and applications to risk management. In *Handbook of Heavy Tailed Distributions in Finance*, ed. S. Rachev, Elsevier, pp. 329–384.
- GLASSERMAN, P. AND JUNEJA, S. (2008). Uniformly efficient importance sampling for the tail distribution of sums of random variables. *Math. Operat. Res.* **33**, 36–50.
- GLYNN, P. W. AND IGLEHART, D. L. (1989). Importance sampling for stochastic simulations. *Manag. Sci.* **35**, 1367–1392.
- HULT, H. AND LINDSKOG, F. (2006). Heavy-tailed insurance portfolios: buffer capital and ruin probabilities. Tech. Rep. 1441, School of ORIE, Cornell University.
- HULT, H., LINDSKOG, F., MIKOSCH, T. AND SAMORODNITSKY, G. (2005). Functional large deviations for multivariate regularly varying random walks. *Ann. Appl. Prob.* **15**, 2651–2680.
- KALLENBERG, O. (1983). *Random Measures*, 3rd edn. Akademie, Berlin.
- MEYN, S. P. AND TWEEDIE, R. L. (1993). *Markov Chains and Stochastic Stability*. Springer, London.
- SADOWSKY, J. S. AND BUCKLEW, J. A. (1990). On large deviations theory and asymptotically efficient Monte Carlo estimation. *IEEE Trans. Inf. Theory* **36**, 579–588.