

# Efficient Rare-Event Simulation for Multiple Jump Events in Regularly Varying Random Walks and Compound Poisson Processes

By Bohan Chen, Jose Blanchet, Chang-Han Rhee & Bert Zwart

Centrum Wiskunde & Informatica and Columbia University

June 13, 2017

## Abstract

We propose a class of strongly efficient rare event simulation estimators for random walks and compound Poisson processes with a regularly varying increment/jump-size distribution in a general large deviations regime. Our estimator is based on an importance sampling strategy that hinges on the heavy-tailed sample path large deviations result recently established in [23]. The new estimators are straightforward to implement and can be used to systematically evaluate the probability of a wide range of rare events with bounded relative error. They are “universal” in the sense that a single importance sampling scheme applies to a very general class of rare events that arise in heavy-tailed systems. In particular, our estimators can deal with rare events that are caused by multiple big jumps (therefore, beyond the usual principle of a single big jump) as well as multidimensional processes such as the buffer content process of a queueing network. We illustrate the versatility of our approach with several applications that arise in the context of mathematical finance, actuarial science, and queueing theory.

## 1 Introduction

In this paper, we develop a strongly efficient importance sampling scheme for computing rare-event probabilities involving path functionals of heavy-tailed random walks and compound Poisson processes in a general large deviations regime. Heavy-tailed distributions play an important role in many man-made stochastic systems. For example, they accurately model inputs to computer and communications networks (see e.g. [17]), and they are an essential component of the description of many financial risk processes (see e.g. [14]).

We focus on stochastic processes with a regularly varying increment/jump-size distribution. The estimators produced with our sampling scheme are straightforward to implement and can be used to estimate the likelihood of a wide range of rare events with bounded relative error. In particular, such a single sampling scheme applies to a very general class of rare events whose occurrence is caused by one or several components in the system which exhibit extreme behavior, while the rest of the system is operating in “normal” circumstances (therefore, beyond the usual principle of a single big jump). In particular, our results apply to a large class of continuous functionals of multiple random walks and compound Poisson processes.

Our estimators are based on importance sampling, a Monte Carlo technique which consists in biasing the nominal distribution of the underlying process in order to induce the rare event of interest. The estimator is obtained by weighting each sample by the corresponding likelihood ratio in order to obtain unbiased estimators. Our goal is to find biasing techniques leading to estimators which have bounded coefficient of variation uniformly as the probability of the event of interest tends to zero in a suitable large deviations regime. A brief review of importance sampling and the notion of strong efficiency will be given later in this paper; for a more in-depth discussion, see [1].

The construction of our sampling scheme is driven by recently developed sample path large deviations results in [23] for regularly varying Lévy processes and random walks. Specifically, let  $X(t)$ ,  $t \geq 0$  be a one-dimensional compensated compound Poisson process with unit arrival rate and a positive jump distribution  $W$  that is regularly varying at infinity. Define  $\bar{X}_n = \{\bar{X}_n(t)\}_{t \in [0,1]}$ , with  $\bar{X}_n(t) = X(nt)/n$ . For a measurable set  $A \subseteq \mathbb{D}$  satisfying a specific topological property, the large deviations results derived in [23] establish that

$$\mathbb{P}(\bar{X}_n \in A) = \Theta \left( (n\mathbb{P}(W \geq n))^{t^*} \right),$$

where precise details can be found in Section 2 below. In practice, exact estimates are often demanded. Hence, we design a sampling scheme for rare events that take the form  $\mathbb{P}(\bar{X}_n \in A)$ . We illustrate our approach with several applications that arise in the context of mathematical finance, actuarial science, and queueing theory.

In order to contextualize our contribution, let us provide a review of the theory and methods which are standard in rare event simulation settings similar to those studied in this paper.

In the context of stochastic processes with light-tailed characteristics, such as random walks with increments possessing a finite moment generating function in a neighborhood of the origin, large deviations theory can be used to bias the process of interest in order to induce the occurrence of the rare event in question. In fact, it is well known that a conventional type of proof of the asymptotic lower bound in large deviations analysis one can extract an exponential change of measure which can sometimes be proved to be efficient (for counterexamples see e.g. [18] and [19]). By connecting the design of efficient importance sampling estimators with a game theoretic formulation, [10], [11] and [12] provide the foundations for the use of large deviations theory in the construction and analysis of provably efficient rare event simulation estimators. Moreover, a weakly efficient “universal” sampler has been proposed by [13] for a general class of hitting sets in arbitrary Jackson network topologies.

The setting of stochastic processes with heavy-tailed increments brings up additional challenges compared to its light-tailed counterpart discussed in the previous paragraph. These challenges were exposed in [2]. First of all, typically, the asymptotic conditional distribution of any particular increment given the rare event of interest converges to the underlying nominal distribution. Intuitively, if a rare event is caused by a large jump that may occur in a single “unlucky” increment out of many possible alternatives, then the chance that any specific increment is precisely the unlucky one is, naturally, small. So, any particular increment is likely to behave “normally” and therefore, in contrast to the light-tailed setting, there is no direct way in which one might attempt to bias a particular increment in order to stir the process towards the rare event of interest.

Moreover, as pointed out in [2], the asymptotic description of the most likely way in which a rare event may occur, for example due to the presence of a single large jump, does not lead to a valid change of measure for importance sampling because it is possible that several large jumps (or no large jump at all) might actually produce the event of interest under the nominal dynamics of the system. In other words, the natural biasing mechanism induced by directly approximating the zero-variance importance sampling distribution in the heavy-tailed setting assigns zero probability to events which are possible under the nominal dynamics leading to an ill-defined likelihood ratio.

The use of state-dependent importance sampling provides a way to deal with these difficulties. In [4], the authors explain how approximating Doob’s  $h$ -transform can lead to a feasible change of measure which produces a strongly efficient importance sampling estimator in the setting of first passage time probabilities for one dimensional random walks. A Lyapunov technique was introduced for the analysis of state-dependent importance sampling estimators. But the direct approximation of Doob’s  $h$ -transform might be difficult to implement in higher dimensions both because of sampling implementation challenges and the evaluation of normalizing constants.

In the setting of one-dimensional compound sums of independent and identically distributed (i.i.d.) regularly varying random variables, [9] produced a state-dependent change of measure whose normalizing constant is straightforward to implement. Their idea can be described as follows: each increment is sampled by either the original measure or—with small probability, which is a design parameter—a different measure, which is essentially the original measure conditional on exhibiting a large jump. The advantage of the mixture samplers is that sampling implementation challenges and the evaluation of normalizing constants can often be addressed by choosing a suitable set of parameters.

Under the setting where the time horizon is growing in large and moderate deviation schemes, Blanchet and Liu show in [6] how to use Lyapunov inequalities to address both the parameter selection while enforcing a bounded relative error. A key step in the methodology is the construction of a suitable Lyapunov function (for an illustration of the technique in multidimensional settings, see [7]). Blanchet and Liu suggest using the type of fluid analysis which is prevalent in the large deviations literature of heavy-tailed stochastic processes (see e.g. [15] and [16]). However, the construction of the Lyapunov function and the verification of the Lyapunov inequality becomes highly non-trivial in settings involving multiple jumps and the presence of boundaries which are common in queueing systems, for an example of the types of complications which arise in queueing settings, see [5].

The idea of using mixtures, suggested in [9], is also used here. But, while [9] treats a particular one-dimensional setting involving a rare event that is caused by a single big jump during a bounded time horizon, our setting is more general. We allow for a wide range of events, which might be caused by multiple jumps during a growing time horizon in a large deviations scaling.

Recall that we are interested in estimating the probability  $\mathbb{P}(\bar{X}_n \in A)$ . The concept behind our sampling scheme can be described as follows. Based on the large deviations results derived in [23], we construct first an auxiliary set  $B^\gamma$  that is closely related to the optimization problem given by (2.1) below. Then, given a fixed mixture probability parameter  $w \in (0, 1)$ , we generate the sample path of  $\bar{X}_n$  under the nominal measure. And, with probability  $1 - w$  we generate the sample path of  $\bar{X}_n$  under the measure  $\mathbb{Q}^\gamma(\cdot) \triangleq \mathbb{P}(\cdot | \bar{X}_n \in B^\gamma)$ . Finally, as a consequence of applying the importance sampling technique, we scale our samplers with a suitable likelihood ratio given as in (3.3) below. It should be noted that the set  $A$  can be as general as in the setting of [23]. Therefore, our methodological contribution in this paper addresses precisely those types of difficulties mentioned in the previous paragraphs, such as, multiple jumps, growing time horizon, avoiding the evaluation of normalizing constants, and by-passing the verification of Lyapunov inequalities. The advantages of our sampling scheme are that the new estimators are strongly efficient and straightforward to implement. Moreover, they are “universal” in the sense that a single importance sampling scheme applies to a very general class of rare events involving multiple jumps that arise in heavy-tailed systems. As a final remark, it should be mentioned that constructing the auxiliary set  $B^\gamma$  requires choosing a set of suitable parameters  $\gamma$  whose existence is guaranteed by the topological property we impose on  $A$ . Hence, one of the main challenges is to select the set of parameters specifically for each application.

Our mathematical contributions in this paper can be summarized as follows.

- We propose a simulation algorithm for estimating the rare-event probability of  $\bar{X}_n \in A$ , together with a sampling scheme for  $\bar{X}_n \in \cdot$  given  $\bar{X}_n \in B^\gamma$ , which is based on a rejection sampling with an unconditional acceptance probability bounded away from zero as  $n \rightarrow \infty$ .
- By showing the existence of the parameter  $\gamma$ , we prove the strong efficiency of our sampling scheme under a very general setting (see Assumption 2 below).
- We showcase the versatility of the algorithm by illustrating the implementation of the proposed sampling scheme to the rare-events that arise in finance, actuarial science, and queueing theory.
- Especially, in the application to queueing networks (see Section 6 below), we show that the tail index of the rare-event probability—which usually exhibit a complex boundary behavior due to the nonlinear nature of the associated Skorokhod mapping—can be determined by solving knapsack problem with nonlinear constraints.

The rest of the paper is organized as follows. Section 2 deals with basic background and notations required to state our contributions. Section 3 introduces our estimators and describes the main result. Applications and numerical implementations are discussed in Section 4, Section 5, and Section 6. All the proofs of results presented in this paper are given in Section 7.

## 2 Notations and preliminaries

### 2.1 Notations

We start with a summary of notations that will be employed in this paper. Let  $\mathbb{Z}_+$  denote the set of non-negative integers, and let  $\mathbb{R}_+$  denote the set of non-negative real numbers. Let  $A^\circ$  and  $A^-$  denote the interior and the closure of  $A$ , respectively. Let  $(\mathbb{D}_{[0,1],\mathbb{R}}, d)$  be the metric space of real-valued RCLL functions on  $[0, 1]$ , denoted by  $\mathbb{D} = \mathbb{D}_{[0,1],\mathbb{R}}$ , equipped with the Skorokhod  $J_1$  metric on  $\mathbb{D}$  that is defined by

$$d(x, y) = \inf_{\lambda \in \Lambda} \|\lambda - id\|_\infty \vee \|x \circ \lambda - y\|_\infty, \quad x, y \in \mathbb{D},$$

where  $id$  denotes the identity mapping,  $\|\cdot\|_\infty$  denotes the uniform metric, i.e.,  $\|x\|_\infty \triangleq \sup_{t \in [0,1]} |x(t)|$ , and  $\Lambda$  denotes the set of all strictly increasing, continuous bijections from  $[0, 1]$  to itself. Let  $\mathbb{D}^k$  denote the  $k$ -fold product space of  $\mathbb{D}$ . Let  $\mathbb{D}_\uparrow^k$  denote the subset of functions in  $\mathbb{D}^k$  that are non-negative and nondecreasing in each coordinate. When it comes to the tail indices of a regularly varying distribution, we use  $\beta$  (or  $\beta_i$  in the multidimensional case) for the right tail and  $\alpha$  for the left tail. Let  $\mathbb{D}_l$  denote the subspace of  $\mathbb{D}$  consisting of non-decreasing step functions vanishing at time zero with  $l$  jumps, and let  $\mathbb{D}_{< l^*}$  denote the subspace of  $\mathbb{D}$  consisting of non-decreasing step functions vanishing at 0 with at most  $l - 1$  jumps, i.e.  $\mathbb{D}_{< l^*} = \bigcup_{l \leq l^* - 1} \mathbb{D}_l$ . Define

$$\mathbb{D}_{< (l_1^*, \dots, l_d^*)} \triangleq \bigcup_{(l_1, \dots, l_d) \in I_{< (l_1^*, \dots, l_d^*)}} \prod_{i=1}^d \mathbb{D}_{l_i},$$

where

$$I_{< (l_1^*, \dots, l_d^*)} \triangleq \{(l_1, \dots, l_d) \in \mathbb{Z}_+^d \setminus \{(l_1^*, \dots, l_d^*)\} \mid \mathcal{I}(l_1, \dots, l_d) \leq \mathcal{I}(l_1^*, \dots, l_d^*)\},$$

and  $\mathcal{I}(l_1, \dots, l_d) \triangleq (\beta_1 - 1)l_1 + \dots + (\beta_d - 1)l_d$ . Define a partial order  $\prec$  on  $\mathbb{Z}_+^d$  such that  $(l_1, \dots, l_d) \prec (m_1, \dots, m_d)$  if and only if  $\mathbb{C}_{(l_1, \dots, l_d)} \subsetneq \mathbb{C}_{(m_1, \dots, m_d)}$ , where  $\mathbb{C}_{(l_1, \dots, l_d)} \triangleq \bigcup_{i=1}^d \mathbb{D}^{i-1} \times \mathbb{D}_{< l_i} \times \mathbb{D}^{d-i}$ . Define

$$J_{(j_1, \dots, j_d)} \triangleq \{(l_1, \dots, l_d) \in \mathbb{Z}_+^d \setminus I_{< (j_1, \dots, j_d)} \mid (m_1, \dots, m_d) \prec (l_1, \dots, l_d) \text{ implies } (m_1, \dots, m_d) \in I_{< (j_1, \dots, j_d)}\}.$$

Let  $\mathbb{D}_{l_-; l_+}$  denote the subspace of the Skorokhod space consisting of step functions vanishing at the origin with exactly  $l_-$  downward jumps and  $l_+$  upward jumps. Define

$$\mathbb{D}_{< l_-^*; l_+^*} \triangleq \bigcup_{(l_-, l_+) \in I_{< l_-^*; l_+^*}} \mathbb{D}_{l_-; l_+}$$

where  $I_{< l_-^*; l_+^*} \triangleq \{(l_-, l_+) \in \mathbb{Z}_+^2 \setminus \{(l_-^*, l_+^*)\} \mid (\alpha - 1)l_- + (\beta - 1)l_+ \leq (\alpha - 1)l_-^* + (\beta - 1)l_+^*\}$ .

Given non-negative sequences of real numbers  $x_n$  and  $y_n$ , we write  $x_n = \mathcal{O}(y_n)$ ,  $x_n = o(y_n)$  and  $x_n = \Theta(y_n)$ , if  $\limsup_{n \rightarrow \infty} x_n/y_n < \infty$ ,  $\lim_{n \rightarrow \infty} x_n/y_n = 0$  and  $0 < \liminf_{n \rightarrow \infty} x_n/y_n \leq \limsup_{n \rightarrow \infty} x_n/y_n < \infty$ , respectively. Given two  $\mathbb{R}$ -valued functions  $f$  and  $g$ , we write  $f \propto g$ , if there exists  $c \in \mathbb{R}$  such that  $f = cg$ . For  $x = (x_1, \dots, x_k)$ ,  $y = (y_1, \dots, y_k) \in \mathbb{R}^k$ , we write  $x \leq y$ , if  $x_i \leq y_i$ , for all  $i \in \{1, \dots, k\}$ . Let the cardinality of  $S$  be denoted by  $|S|$  or  $\#S$ . Finally, let  $\mathcal{C}(S, k)$  and  $\mathcal{P}(S, k)$  denote the set of all  $k$ -combinations and  $k$ -permutations of a set  $S$ , respectively. Note that  $|\mathcal{C}(S, k)| = \binom{|S|}{k}$  and  $|\mathcal{D}(S, k)| = |\mathcal{C}(S, k)| * k!$

To describe the efficiency of a rare-event simulation algorithm, we adopt a widely applied criterion, which requires that the relative mean squared error of the associated estimator is appropriately controlled. To be more precise, suppose that we are interested in a sequence of rare events  $A_n$ , which becomes more and more rare as  $n \rightarrow \infty$ . For each  $n \in \mathbb{Z}_+$ , let  $L_n$  be an unbiased estimator of the rare-event probability  $\eta_n = \mathbb{P}(A_n)$ .  $L_n$  is said to be strongly efficient if  $\mathbb{E}L_n^2 = \mathcal{O}(\eta_n^2)$ . In particular, strong efficiency implies that the number of simulation runs required to estimate the target probability to a given relative accuracy is bounded with respect to (w.r.t.)  $n$ .

## 2.2 Preliminaries

As we will see, the simulation algorithm that we propose in this paper is constructed based on the asymptotic behavior of rare-event probabilities, therefore we review some well known large deviations results for scaled Lévy processes with heavy-tailed Lévy measures, introduced in [23]. To begin with, let  $X$  be a Lévy process with Lévy measure  $\nu$ , where  $\nu$  is spectrally positive and regularly varying (at infinity) with index  $-\beta < -1$ . Let  $\bar{X}_n \triangleq \{X(nt)/n\}_{t \in [0,1]}$  denote the associated scaled process. Let  $\nu_\beta^l$  denote the restriction of  $l$ -fold product measure of  $\nu_\beta$  to  $\{x \in \mathbb{R}_+^l : x_1 \geq x_2 \geq \dots \geq x_l\}$ , where  $\nu_\beta(x, \infty) \triangleq x^{-\beta}$ . For  $l \geq 1$ , define a (Borel) measure

$$C_l(\cdot) \triangleq \mathbb{E} \left[ \nu_\beta^l \left\{ y \in (0, \infty)^l \mid \sum_{i=1}^l y_i \mathbb{1}_{[U_i, 1]} \in \cdot \right\} \right]$$

where  $U_i, i \geq 1$  are i.i.d. uniformly distributed on  $[0, 1]$ . Note that  $C_l$  is concentrated on  $\mathbb{D}_l$ , i.e.,  $C_l(\mathbb{D}_l) = 1$ . Moreover, we make the convention that  $C_0$  is the Dirac measure concentrated on the zero function. The following result is useful in designing an efficient algorithm for rare events involving one-dimensional scaled processes. Throughout the rest of this paper, all measurable sets are understood to be Borel measurable.

**Result 2.1** (Theorem 3.1 of [23]). *Suppose that  $A$  is a measurable set. If  $A$  is bounded away from  $\mathbb{D}_{< l^*}$ , where  $l^* \triangleq \min\{l \in \mathbb{Z}_+ \mid \mathbb{D}_l \cap A \neq \emptyset\} < \infty$ , then we have that*

$$C_{l^*}(A^\circ) \leq \liminf_{n \rightarrow \infty} \frac{\mathbb{P}(\bar{X}_n \in A)}{(n\nu[n, \infty))^{l^*}} \leq \limsup_{n \rightarrow \infty} \frac{\mathbb{P}(\bar{X}_n \in A)}{(n\nu[n, \infty))^{l^*}} \leq C_{l^*}(A^-).$$

As one can see in Section 5 and Section 6 below, plenty of applications can be interpreted as sample-path rare events in a multidimensional setting. Therefore, it is particularly interesting to consider large deviations results for multidimensional processes. Let  $X^{(1)}, \dots, X^{(d)}$  be independent centered one-dimensional Lévy processes with spectrally positive Lévy measures  $\nu_1(\cdot), \dots, \nu_d(\cdot)$ , respectively, where each  $\nu_i$  is regularly varying with index  $-\beta_i < -1$  at infinity. Moreover, for the finite product of metric spaces, we use the maximum metric; i.e., we use  $d_{\mathbb{S}_1 \times \dots \times \mathbb{S}_d}((x_1, \dots, x_d), (y_1, \dots, y_d)) \triangleq \max_{i=1, \dots, d} d_{\mathbb{S}_i}(x_i, y_i)$  for the product  $\mathbb{S}_1 \times \dots \times \mathbb{S}_d$  of metric spaces  $(\mathbb{S}_i, d_{\mathbb{S}_i})$ . Finally, for  $(l_1, \dots, l_d) \in \mathbb{Z}_+^d$ , we define  $C_{l_1} \times \dots \times C_{l_d}(\cdot)$  (which is concentrated on  $\prod_{i=1}^d \mathbb{D}_{l_i}$ ) as the product measure of

$$C_{l_i}(\cdot) \triangleq \mathbb{E} \left[ \nu_{\beta_i}^{l_i} \left\{ y \in (0, \infty)^{l_i} \mid \sum_{j=1}^{l_i} y_j \mathbb{1}_{[U_j, 1]} \in \cdot \right\} \right].$$

Result 2.2 states a large deviations result for  $d$  dimensional process  $\bar{X}_n(t) \triangleq (X^{(1)}(nt)/n, \dots, X^{(d)}(nt)/n)$  for  $t \in [0, 1]$ .

**Result 2.2** (Theorem 3.6 of [23]). *Suppose that  $A$  is measurable. If  $A$  is bounded away from  $\mathbb{D}_{< (l_1^*, \dots, l_d^*)}$ , where*

$$(l_1^*, \dots, l_d^*) = \arg \min_{\substack{(l_1, \dots, l_d) \in \mathbb{Z}_+^d \\ \prod_{i=1}^d \mathbb{D}_{l_i} \cap A \neq \emptyset}} \mathcal{I}(l_1, \dots, l_d), \quad (2.1)$$

then we have that

$$\begin{aligned} C_{l_1^*} \times \dots \times C_{l_d^*}(A^\circ) &\leq \liminf_{n \rightarrow \infty} \frac{\mathbb{P}(\bar{X}_n \in A)}{\prod_{i=1}^d (n\nu_i[n, \infty))^{l_i^*}} \\ &\leq \limsup_{n \rightarrow \infty} \frac{\mathbb{P}(\bar{X}_n \in A)}{\prod_{i=1}^d (n\nu_i[n, \infty))^{l_i^*}} \leq C_{l_1^*} \times \dots \times C_{l_d^*}(A^-). \end{aligned}$$

Note that the assumption that  $A$  is bounded away from  $\mathbb{D}_{< (l_1^*, \dots, l_d^*)}$  guarantees the uniqueness of  $(l_1^*, \dots, l_d^*)$ . Finally, we conclude this section with an extension of Result 2.2, which will be useful in

constructing an efficient simulation algorithm for heavy-tailed random walks. Let  $S_k$ ,  $k \geq 0$ , be a random walk, set  $\bar{S}_n(t) = S_{\lfloor nt \rfloor} / n$ ,  $t \geq 0$ , and define  $\bar{S}_n = \{\bar{S}_n(t), t \in [0, 1]\}$ . Let  $\nu_\beta^l$  be as defined above. Similarly, let  $\nu_\alpha^m$  denote the restriction of  $m$ -fold product measure of  $\nu_\alpha$  to  $\{x \in \mathbb{R}_+^m : x_1 \geq x_2 \geq \dots \geq x_m\}$ , where  $\nu_\alpha(x, \infty) \triangleq x^{-\alpha}$ . Let  $C_{0,0}(\cdot) \triangleq \delta_0(\cdot)$  be the Dirac measure concentrated on the zero function. For each  $(l_-, l_+) \in \mathbb{Z}_+^2 \setminus \{(0, 0)\}$ , define a measure (which is concentrated on  $\mathbb{D}_{l_-, l_+}$ )

$$C_{l_-, l_+}(\cdot) \triangleq \mathbb{E} \left[ \nu_\alpha^{l_-} \times \nu_\beta^{l_+} \{ (x, y) \in (0, \infty)^{l_-} \times (0, \infty)^{l_+} : \sum_{i=1}^{l_-} x_i 1_{[U_i, 1]} - \sum_{i=1}^{l_+} y_i 1_{[V_i, 1]} \in \cdot \} \right],$$

where  $U_i$ 's and  $V_i$ 's are i.i.d. uniform on  $[0, 1]$ .

**Result 2.3.** *Suppose that  $\mathbb{P}(S_1 \leq -x)$  is regularly varying with index  $-\alpha$  and  $\mathbb{P}(S_1 \geq x)$  is regularly varying with index  $-\beta$ . Let  $A$  be a measurable set bounded away from  $\mathbb{D}_{<l^*, l_+^*}$ , where*

$$(l_-^*, l_+^*) = \underset{\substack{(l_-, l_+) \in \mathbb{Z}_+^2 \\ \mathbb{D}_{l_-, l_+} \cap A \neq \emptyset}}{\operatorname{arg\,min}} (\alpha - 1)l_- + (\beta - 1)l_+. \quad (2.2)$$

Then

$$\begin{aligned} C_{l_-^*, l_+^*}(A^\circ) &\leq \liminf_{n \rightarrow \infty} \frac{\mathbb{P}(\bar{S}_n \in A)}{(n\mathbb{P}(S_1 \leq -n))^{l_-^*} (n\mathbb{P}(S_1 \geq n))^{l_+^*}} \\ &\leq \limsup_{n \rightarrow \infty} \frac{\mathbb{P}(\bar{S}_n \in A)}{(n\mathbb{P}(S_1 \leq -n))^{l_-^*} (n\mathbb{P}(S_1 \geq n))^{l_+^*}} \leq C_{l_-^*, l_+^*}(A^-). \end{aligned}$$

### 3 Main results

In this section we present our main results. Although the large deviations results reviewed in Section 2 are stated for Lévy processes, we focus on compensated compound Poisson process for simulation purposes. Let  $X$  denote a  $d$ -dimensional compensated compound Poisson process, and recall that  $\bar{X}_n$  is the scaled process with  $\bar{X}_n(t) = X(nt)/n$ ,  $t \in [0, 1]$ . For a measurable set  $A \in \mathbb{D}^d$ , we are interested in estimating the probability of the event  $A_n \triangleq \{\bar{X}_n \in A\}$ , when  $n$  is large. Note that, in view of the law of large numbers, one can expect that  $\mathbb{P}(\bar{X}_n \in A) \rightarrow 0$  for  $A$ 's that are bounded away from the zero function, and hence,  $A_n$ 's are rare events for large  $n$ 's. In Section 3.1, we first illustrate the idea of our algorithm in the special case for  $d = 1$ , where the notations are simpler. In Section 3.2 we extend this result to general  $d$ .

#### 3.1 The one-dimensional case

Let  $\{X(t)\}_{t \geq 0}$  be a one-dimensional compensated compound Poisson process with jump sizes  $\{W(k)\}_{k \geq 1}$ . That is,

$$X(t) = \sum_{k=1}^{N(t)} W(k) - \lambda t \mathbb{E}W(1)$$

where  $\{N(t)\}_{t \geq 0}$  is a Poisson process with arrival rate  $\lambda$ , and let  $\bar{X}_n \triangleq \{X(nt)/n\}_{t \in [0, 1]}$  denote the associated scaled process. Moreover, let  $\mathbb{P}(X(1) > x)$  be regularly varying of index  $-\beta < -1$ . The following assumption is essential for analyzing the asymptotic behavior of the rare-event probability, and hence, deriving the strong efficiency of our estimator.

**Assumption 1.** Let  $A$  be a measurable set in  $\mathbb{D}$ . We assume that  $A$  is bounded away from  $\mathbb{D}_{<l^*}$ , where  $l^* = \min \{l \in \mathbb{Z}_+ \mid \mathbb{D}_l \cap A \neq \emptyset\}$  denotes the minimal number of upward jumps of a step function in  $A$ . Moreover, assume that  $C_{l^*}(A^\circ) > 0$ .

*Remark 1.* As one can see in Section 4, 5, and 6, one of the typical settings that arises in applications is that the set  $A$  can be written as a finite combination of unions and intersections of  $F_1^{-1}(A_1), \dots, F_m^{-1}(A_m)$ , where each  $F_i: \mathbb{D} \rightarrow \mathbb{S}_i$  is a continuous function, and all sets  $A_i$  are subsets of a general topological space  $\mathbb{S}_i$ . If we denote this operation of taking unions and intersections by  $\Psi$  (i.e.,  $A = \Psi(F_1^{-1}(A_1), \dots, F_m^{-1}(A_m))$ ), then it holds that

$$\Psi(F_1^{-1}(A_1^\circ), \dots, F_m^{-1}(A_m^\circ)) \subseteq A^\circ \subseteq A \subseteq A^- \subseteq \Psi(F_1^{-1}(A_1^-), \dots, F_m^{-1}(A_m^-)).$$

Hence,  $C_{l^*}(A^\circ) > 0$  holds if  $\hat{T}_{l^*}^{-1}(\Psi(F_1^{-1}(A_1^\circ), \dots, F_m^{-1}(A_m^\circ)))$  has positive Lebesgue measure, where  $\hat{T}_j: \hat{\mathbb{S}}_j \rightarrow \mathbb{D}_j$  is defined by  $\hat{T}_j(x, u) \triangleq \sum_{i=1}^j x_i \mathbb{1}_{[u_i, 1]}$  for  $j \in \mathbb{Z}_+$ , and

$$\hat{\mathbb{S}}_j \triangleq \left\{ (x, u) \in \mathbb{R}_+^j \times [0, 1]^j \mid x_1 \geq \dots \geq x_j, 0, 1, u_1, \dots, u_j \text{ are distinct} \right\}.$$

Analogously, one can derive a sufficient condition for  $C_{l_1^*} \times \dots \times C_{l_j^*}(A^\circ) > 0$  (see Assumption 2 below). More details about this discussion can be found in Section 3.1 of [23].

We design a rare-event simulation algorithm that estimates the probability of  $A_n \triangleq \{\bar{X}_n \in A\}$  efficiently, based on an importance sampling strategy. To construct an importance distribution, we introduce a constant  $\gamma > 0$  and define  $B_n^\gamma \triangleq \{\bar{X}_n \in B^\gamma\}$ , where  $B^\gamma$  is given by

$$B^\gamma \triangleq \left\{ \xi \mid \#\{t \mid \xi(t) - \xi(t^-) > \gamma\} \geq l^* \right\}.$$

In the construction of our rare-event simulation algorithm, we will take advantage of the fact that one can always choose  $\gamma$  so that  $B_n^\gamma$  is sufficiently ‘‘close’’ to  $A_n$ . The specific choice of  $\gamma$  will be further discussed later in Section 4, 5, and 6 for concrete examples. Let  $\mathbb{Q}_\gamma(\cdot) \triangleq \mathbb{P}(\cdot | B_n^\gamma)$  denote the conditional distribution given  $\bar{X}_n \in B^\gamma$ . One should notice that  $d\mathbb{Q}_\gamma/d\mathbb{P} = \mathbb{P}(B_n^\gamma)^{-1} \mathbb{1}_{B_n^\gamma}$ . Moreover, by the Fubini-Tonelli’s theorem, a closed-form expression for  $\mathbb{P}(B_n^\gamma)$  is given by

$$\begin{aligned} \mathbb{P}(B_n^\gamma) &= e^{-\lambda n} \sum_{i=l^*}^{\infty} \frac{(\lambda n)^i}{i!} \sum_{j=l^*}^i \binom{i}{j} \mathbb{P}(W(1) > n\gamma)^j (1 - \mathbb{P}(W(1) > n\gamma))^{i-j} \\ &= e^{-\lambda n} \sum_{j=l^*}^{\infty} \frac{(\lambda n)^j}{j!} \mathbb{P}(W(1) > n\gamma)^j \sum_{i=j}^{\infty} \frac{(\lambda n)^{i-j}}{(i-j)!} (1 - \mathbb{P}(W(1) > n\gamma))^{i-j} \\ &= 1 - \exp \left\{ -\lambda n \mathbb{P}(W(1) > n\gamma) \right\} \sum_{j=0}^{l^*-1} \frac{(\lambda n)^j}{j!} \mathbb{P}(W(1) > n\gamma)^j. \end{aligned} \quad (3.1)$$

From (3.1) one should recognize that  $B_n^\gamma$  can be interpreted as the event of a Poisson distributed random variable with rate  $\lambda n \mathbb{P}(W(1) > n\gamma)$  crossing the level  $l^*$ . Now, let  $w \in (0, 1)$  be arbitrary but fixed. We propose an importance distribution  $\mathbb{Q}_{\gamma, w}$  that is absolutely continuous w.r.t.  $\mathbb{P}$  and is given by

$$\mathbb{Q}_{\gamma, w}(\cdot) \triangleq w \mathbb{P}(\cdot) + (1-w) \mathbb{Q}_\gamma(\cdot). \quad (3.2)$$

We give here an algorithm for generating the sample path of  $\bar{X}_n$  under the probability measure  $\mathbb{Q}_\gamma(\cdot)$ . Since  $\{\bar{X}_n \in B^\gamma\} \subseteq \{N(n) \geq l^*\}$ , we observe that

$$\mathbb{Q}_\gamma(\bar{X}_n \in \cdot) = \frac{1}{\mathbb{P}(B_n^\gamma)} \mathbb{P}(\bar{X}_n \in \cdot, B_n^\gamma) = \sum_{m=l^*}^{\infty} h_m \mathbb{P}(\bar{X}_n \in \cdot | B_n^\gamma, N(n) = m),$$

where  $h_m = h_m(n) \triangleq \mathbb{P}(B_n^\gamma, N(n) = m) / \mathbb{P}(B_n^\gamma)$  satisfies  $\sum_{m \geq l^*} h_m = 1$ . Note that  $h_m$  can be computed, since

$$\begin{aligned} \mathbb{P}(B_n^\gamma, N(n) = m) &= \mathbb{P}(B_n^\gamma | N(n) = m) \mathbb{P}(N(n) = m) \\ &= e^{-\lambda n} \frac{(\lambda n)^m}{m!} \left( \sum_{i=l^*}^m \binom{m}{i} \mathbb{P}(W(1) > n\gamma)^i (1 - \mathbb{P}(W(1) > n\gamma))^{m-i} \right). \end{aligned}$$

Hence, it remains to discuss sampling from  $\mathbb{P}(\bar{X}_n \in \cdot | B_n^\gamma, N(n) = m)$ . It turns out that we can proceed a rejection sampling, where drawing from the proposal distribution can be achieved as follows.

1. Sample  $\{b_k\}_{k \leq l^*}$  uniformly from  $\mathcal{C}(\{1, \dots, m\}, l^*)$ ;
2. sample each  $W(b_k)$ ,  $k \leq l^*$ , conditional on  $W(1) > n\gamma$ ;
3. sample  $W(m')$ ,  $m' \leq m$ ,  $m' \notin \{b_k\}_{k \leq l^*}$ , under the nominal measure.

Note that the target density  $f_{\text{target};m}$ , defined by

$$f_{\text{target};m}(w_1, \dots, w_m) dw_1 \cdots dw_m \triangleq \mathbb{P}(W(1) \in w_1 + dw_1, \dots, W(m) \in w_m + dw_m \mid B_n^\gamma, N(n) = m),$$

can be bounded by  $M_m f_{\text{proposal};m}(w_1, \dots, w_m)$ , where

$$f_{\text{target};m}(w_1, \dots, w_m) \propto \mathbb{P}(B_n^\gamma \mid N(n) = m)^{-1} \prod_{j=1}^m \frac{d}{dw_j} \mathbb{P}(W(j) \leq w_j) \mathbb{1}_{B_n^\gamma}(w_1, \dots, w_m),$$

$$f_{\text{proposal};m}(w_1, \dots, w_m) = \binom{m}{l^*}^{-1} \mathbb{P}(W(1) > n\gamma)^{-l^*} \prod_{j=1}^m \frac{d}{dw_j} \mathbb{P}(W(j) \leq w_j) \sum_{\substack{(b_1, \dots, b_{l^*}) \in \\ \mathcal{C}(\{1, \dots, m\}, l^*)}} \mathbb{1}_{\{W(b_k) > n\gamma, \forall k \leq l^*\}},$$

and hence,

$$M_m = M_m(n) \triangleq \frac{\binom{m}{l^*} \mathbb{P}(W(1) > n\gamma)^{l^*}}{\mathbb{P}(B_n^\gamma \mid N(n) = m)}.$$

Now, it is natural to accept  $(W(1), \dots, W(m))$  with probability

$$a(W(1), \dots, W(m)) = \left( \frac{\#\{i \in \{1, \dots, m\} \mid W(i) > n\gamma\}}{l^*} \right)^{-1}.$$

Finally, we are able to formulate the pseudocode for generating the sample path of  $\bar{X}_n$  under  $\mathbb{Q}_\gamma$  in Algorithm 1. Moreover, we show in Proposition 3.1 that the expected running time of Algorithm 1 is uniformly bounded from above w.r.t.  $n$ .

**Proposition 3.1.** *Let  $T_{\text{alg1}}(n)$  denote the expected running time of Algorithm 1. Under the assumption that  $W(1)$  is regularly varying of index  $-\beta < -1$ , we have that  $T_{\text{alg1}}(n) = \sum_{m \geq l^*} h_m(n) M_m(n)$  is uniformly bounded from above w.r.t.  $n$ , i.e.  $\max_{n \geq 0} T_{\text{alg1}}(n) < \infty$ .*

*Proof.* See Section 7. □

In view of the observations we made so far, we propose an estimator  $Z_n$  for  $\mathbb{P}(A_n)$  that is given by

$$Z_n = \mathbb{1}_{A_n} \frac{d\mathbb{P}}{d\mathbb{Q}_{\gamma,w}} = \frac{\mathbb{1}_{A_n}}{w + \frac{1-w}{\mathbb{P}(B_n^\gamma)} \mathbb{1}_{B_n^\gamma}}. \quad (3.3)$$

Intuitively, an importance sampling technique is used to get more samples from the interesting region, by sampling from a distribution that overweights the important region. Based on this, the choice of  $B_n^\gamma$  can be “justified”, since  $B_n^\gamma$  is mimicking the asymptotic behavior of the probability of interest. However, as one can see in the proof of strong efficiency (see Theorem 3.6 below), we should analyze the second moment of our estimator to avoid “backfire”, yielding an estimator with larger or even infinite variance. It turns out that this intuition can be made rigorous by applying Result 2.1. We end this section with a theorem regarding to the strong efficiency of our estimator.

**Theorem 3.2.** *Under Assumption 1, there exists a  $\gamma > 0$  such that the estimator constructed in (3.3) is strongly efficient for estimating  $\mathbb{P}(A_n)$ .*

*Proof.* Analogous to the proof of the more general Theorem 3.6 presented below, where the existence of  $\gamma$  is also discussed. □



---

**Algorithm 1** Generating the sample path of  $\bar{X}_n$  under  $\mathbb{Q}_\gamma$

---

**Input:**  $\gamma$

sample  $m \sim h_m$   $\{m = m' \text{ with probability } h_{m'} = \mathbb{P}(N(n) = m' \mid B_n^\gamma)\}$

$R \leftarrow \text{true}$

**while**  $R = \text{true}$  **do**

sample  $\{b_k\}_{k \leq l^*} \sim \text{unif}(\mathcal{C}(\{1, \dots, m\}, k))$   $\{\text{uniform distribution on } \mathcal{C}(\{1, \dots, m\}, k)\}$

**for**  $i = 1$  to  $l^*$  **do**

**if**  $i \in \{b_k\}_{k \leq l^*}$  **then**

sample  $W(i) \sim W(1) \mid W(1) > n\gamma$

**else**

sample  $W(i) \sim W(1)$

**end if**

**end for**

$c \leftarrow \#\{j \in \{1, \dots, m\} \mid W(j) > n\gamma\}$

$a \leftarrow \binom{c}{l^*}^{-1}$

sample  $u \sim \text{uniform}[0, 1]$

**if**  $u < a$  **then**

$R \leftarrow \text{false}$

**else**

$R \leftarrow \text{true}$

**end if**

**end while**

**return**  $\bar{X}_n$

---

### 3.2 Extension to general $d$

In this section we extend the results in Section 3.1 to the case with arbitrary  $d$ . To be precise, let  $X \triangleq (X^{(1)}, \dots, X^{(d)})$  be a superposition of  $d$  independent compensated compound Poisson processes with upward jumps, where  $\{N^{(i)}(t)\}$  is a Poisson process with arrival rate  $\lambda_i$ , and

$$X^{(i)}(t) = \sum_{k=1}^{N^{(i)}(t)} W^{(i)}(k) - \lambda_i t \mathbb{E}W^{(i)}(1).$$

Moreover, let  $\mathbb{P}(X^{(i)}(1) > x)$  be regularly varying of index  $-\beta_i < -1$  at infinity. Finally, let  $\bar{X}_n$  denote the corresponding scaled process. As we can see in Result 2.2, the large deviations results for  $\mathbb{P}(\bar{X}_n \in A)$  depend heavily on the value of  $\mathcal{I}(l_1^*, \dots, l_d^*)$ , where  $(l_1^*, \dots, l_d^*)$  is as defined in (2.1). However, for  $c \in \mathbb{R}$ , the grid  $(l_1, \dots, l_d) \in \mathbb{Z}_+^d$  satisfying  $\mathcal{I}(l_1, \dots, l_d) = c$  is not unique in general. Therefore assuming  $A$  being bounded away from  $\prod_{i=1}^d \mathbb{D}_{< l_i}$  is not sufficient for our purposes. The following assumption, which is slightly different from the one we made in Section 3.1, corresponds to the extension of Result 2.1 to Result 2.2.

**Assumption 2.** Let  $A$  be a measurable set. Assume that  $A$  is bounded away from  $\mathbb{D}_{< (l_1^*, \dots, l_d^*)}$ , where  $(l_1^*, \dots, l_d^*)$  is the unique solution of the minimization problem given by (2.1). Moreover, assume that  $C_{l_1^*} \times \dots \times C_{l_d^*}(A^\circ) > 0$ .

If the solution to (2.1) is not unique, we may partition  $A$ . As in Section 3.1, we focus now on constructing the auxiliary set  $B^\gamma$  for the importance distribution. Define  $A_n \triangleq \{\bar{X}_n \in A\}$  and  $B_n^\gamma \triangleq \{\bar{X}_n \in B^\gamma\}$ . As one can see in the proof of Theorem 3.6, controlling the probability of  $A_n \cap (B_n^\gamma)^c$  should be taken into account in choosing the auxiliary set  $B^\gamma$ . In the one-dimensional case, letting  $B^\gamma$  mimic the optimal path leading to the rare event makes us capable of controlling the relative error of our estimator. However, the same strategy will fail in the multidimensional case, since the rare event can be reached through other feasible but not necessarily optimal paths. Thus, we require a more complicated construction of  $B^\gamma$ .

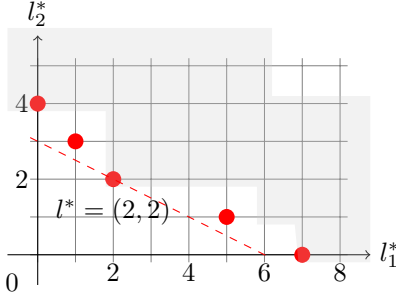


Figure 1: An example of  $J_{(l_1^*, \dots, l_d^*)}$  as in Definition 1. For  $(\beta_1 - 1)/(\beta_2 - 1) = 2$ , a given set  $A \in \mathbb{D}^2$  and the corresponding  $(l_1^*, l_2^*) = (2, 2)$ , we mark the elements in  $J_{(l_1^*, \dots, l_d^*)}$  with points. Moreover, the shaded area contains all those points  $(l_1, l_2)$  such that  $A \cap \mathbb{D}_{l_1} \times \mathbb{D}_{l_2} \neq \emptyset$ .

**Definition 1.** Let  $A$  be a measurable set in  $\mathbb{D}^d$ , and let  $(l_1^*, \dots, l_d^*)$  denote the unique solution to (2.1). Let  $\gamma \in \mathbb{R}^d$  with  $\gamma_i > 0$  for all  $i \in \{1, \dots, d\}$ , and define

$$B^\gamma \triangleq \bigcup_{(l_1, \dots, l_d) \in J_{(l_1^*, \dots, l_d^*)}} B^{\gamma; l}, \quad (3.4)$$

where  $B^{\gamma; l}$  is the set of càdlàg functions on  $\mathbb{R}^d$  that have greater or equal to than  $l_i$  number of jumps with size larger than  $\gamma_i$  in its  $i$ -th coordinate, i.e.,

$$B^{\gamma; l} \triangleq \left\{ \left( \xi^{(1)}, \dots, \xi^{(d)} \right) \in \mathbb{D}^d \mid \#\{t \mid \xi^{(i)}(t) - \xi^{(i)}(t^-) > \gamma_i\} \geq l_i, \forall i \in \{1, \dots, d\} \right\}.$$

*Remark 2.* Note that the cardinality of  $J_{(l_1^*, \dots, l_d^*)}$  is finite. To design a strongly efficient simulation algorithm for estimating  $\mathbb{P}(A_n)$ , we will take advantages of an important property of  $J_{(l_1^*, \dots, l_d^*)}$ . That is, for all  $\xi \in A$  with  $A$  being bounded away from  $\mathbb{D}_{<(l_1^*, \dots, l_d^*)}$ , there exists an index  $(l_1, \dots, l_d) \in J_{(l_1^*, \dots, l_d^*)}$ , such that the path of  $\xi$  in its  $i$ -th coordinate is bounded away from  $\mathbb{D}_{<l_i^*}$ , for every  $i \in \{1, \dots, d\}$ . An illustration of  $J_{(l_1^*, \dots, l_d^*)}$  can be found in Figure 1.

Let  $\mathbb{Q}_\gamma(\cdot) \triangleq \mathbb{P}(\cdot \mid B_n^\gamma)$  and let  $\mathbb{Q}_{\gamma, w}$  be as defined in (3.2), following the same strategy as in Section 3.1 we propose an estimator that takes the same form as in (3.3). Before turning to the efficiency analysis of our estimator, we summarize the findings above in Algorithm 2.

---

**Algorithm 2** Efficient sampling of  $\mathbb{P}(\bar{X}_n \in A)$

---

**Input:**  $\gamma \in \mathbb{R}^d$ ,  $w \in (0, 1)$   
sample  $u \sim \text{uniform}[0, 1]$  {uniform distribution on  $[0, 1]$ }  
**if**  $u < w$  **then**  
  sample  $\bar{X}_n \sim \mathbb{P}(\bar{X}_n \in \cdot)$   
**else**  
  sample  $\bar{X}_n \sim \mathbb{P}(\bar{X}_n \in \cdot \mid \bar{X}_n \in B^\gamma)$   
**end if**  
 $L \leftarrow [w + (1 - w)\mathbb{1}_{B_n^\gamma} / \mathbb{P}(B_n^\gamma)]^{-1}$   
**if**  $\bar{X}_n \in A$  **then**  
  **return**  $L$   
**else**  
  **return** 0  
**end if**

---

*Remark 3.* The strategy of choosing  $\gamma$  will be explained specifically via the examples presented in Section 4, Section 5 and Section 6.

In order to complete our algorithm, we need to discuss the computation of  $\mathbb{P}(B_n^\gamma)$ , as well as the strategy of sampling from the conditional distribution  $\mathbb{Q}_\gamma(\cdot)$ . Since  $B^\gamma$  constructed in Definition 1 is the union of  $B^{\gamma;l}$  with  $l = (l_1, \dots, l_d) \in J_{(l_1^*, \dots, l_d^*)}$ , by the inclusion-exclusion principle, it is sufficient to discuss computing the probability of sets of the form  $\bigcap_{(l_1, \dots, l_d) \in I} B^{\gamma;l}$ , where  $I$  is a finite collection of elements in  $\mathbb{Z}_+^d$ . It turns out that the probability of such a set can be computed similarly as in Section 3.1. Based on this observation, we give the following proposition.

**Proposition 3.3.** *The probability of  $B_n^\gamma$  is equal to*

$$\sum_{k=1}^{|J_{(l_1^*, \dots, l_d^*)}|} \left( (-1)^{k-1} \sum_{\substack{|I|=k \\ I \subseteq J_{(l_1^*, \dots, l_d^*)}}} \prod_{i=1}^d \left( 1 - \exp \left\{ -\lambda_i n \mathbb{P}(W^{(i)}(1) > n\gamma_i) \right\} \sum_{j=0}^{\hat{l}_{i;I}-1} \frac{(\lambda_i n)^j}{j!} \mathbb{P}(W^{(i)}(1) > n\gamma_i)^j \right) \right),$$

where  $\hat{l}_{i;I} \triangleq \max_{(l_1, \dots, l_d) \in I} l_i$ .

*Proof.* See Section 7. □

*Remark 4.* It should be mentioned that the complexity of computing  $\mathbb{P}(B_n^\gamma)$  can be reduced rapidly in the case, where, for example, one is able to take a smaller (in the sense of cardinality) set than  $J_{(l_1^*, \dots, l_d^*)}$  (see e.g. Corollary 3.8, Section 5 and 6 below).

As in Section 3.1, we now discuss generating the sample path of  $\bar{X}_n$  under  $\mathbb{Q}_\gamma$  in the next step. To begin with, we need the following lemma, which shows that  $B^\gamma$  can be decomposed into finitely many disjoint sets.

**Lemma 3.4.** *Let  $B^{\gamma;l}(i, j) \triangleq \{\xi \in \mathbb{D}^d \mid \#\{t \mid \xi^{(i)}(t) - \xi^{(i)}(t^-) > \gamma_i\} \geq (l(j))_i\}$ . Let the elements in  $J_{(l_1^*, \dots, l_d^*)}$ , denoted by  $l(1), \dots, l(|J_{(l_1^*, \dots, l_d^*)}|)$ , be ordered such that  $(l(1))_d \leq \dots \leq (l(|J_{(l_1^*, \dots, l_d^*)}|))_d$ . Define*

$$\Delta B^{\gamma;l}(i, j) \triangleq B^{\gamma;l}(i, j) \setminus \left( \bigcup_{m=1}^{j-1} B^{\gamma;l}(i, m) \right), \quad i \in \{1, \dots, d-1\}. \quad (3.5)$$

*Then, we have that*

$$B^\gamma = \bigcup_{m_1=1}^{|J_{(l_1^*, \dots, l_d^*)}|} \bigcup_{m_2=1}^{m_1} \dots \bigcup_{m_{d-1}=1}^{m_{d-2}} \left( \left( \bigcap_{i=1}^{d-1} \Delta B^{\gamma;l}(i, m_i) \right) \cap B^{\gamma;l}(d, 1) \right).$$

*Proof.* See Section 7. □

Lemma 3.4 shows that  $B^\gamma$  can be decomposed into finitely many disjoint sets. This implies that

$$\mathbb{Q}_\gamma(\bar{X}_n \in \cdot) = \sum_{m_1=1}^{|J_{(l_1^*, \dots, l_d^*)}|} \sum_{m_2=1}^{m_1} \dots \sum_{m_{d-1}=1}^{m_{d-2}} h_{1; m_1, \dots, m_{d-1}} \mathbb{P}(\bar{X}_n \in \cdot \mid \bar{X}_n \in B^\gamma(m_1, \dots, m_{d-1})),$$

where

$$B^\gamma(m_1, \dots, m_{d-1}) \triangleq \left( \bigcap_{i=1}^{d-1} \Delta B^{\gamma;l}(i, m_i) \right) \cap B^{\gamma;l}(d, 1),$$

and  $h_{1; m_1, \dots, m_{d-1}} \triangleq \mathbb{P}(\bar{X}_n \in B^\gamma(m_1, \dots, m_{d-1})) / \mathbb{P}(\bar{X}_n \in B^\gamma)$  satisfies that

$$\sum_{m_1=1}^{|J_{(l_1^*, \dots, l_d^*)}|} \sum_{m_2=1}^{m_1} \dots \sum_{m_{d-1}=1}^{m_{d-2}} h_{1; m_1, \dots, m_{d-1}} = 1.$$

Hence, it remains to design a sampling scheme for generating the sample path of  $\bar{X}_n$  under  $\mathbb{P}(\cdot \mid \bar{X}_n \in B^\gamma(m_1, \dots, m_{d-1}))$  (for details about generating multi-dimensional discrete random numbers, see e.g.

[21]). Due to the special structure of  $B^\gamma(m_1, \dots, m_{d-1})$ , we are able to generate  $\bar{X}_n^{(1)}, \dots, \bar{X}_n^{(d)}$  independently under  $\mathbb{P}(\cdot | \bar{X}_n \in B^\gamma(m_1, \dots, m_{d-1}))$ . To see this, first note that sampling  $\bar{X}_n^{(d)}$  is trivial due to the discussion in Section 3.1. Define

$$\check{l}(m_i; i) \triangleq \min_{\xi \in \Delta B^{\gamma;l}(i, m_i)} \#\{t | \xi(t) - \xi(t^-) > \gamma_i\},$$

and

$$\hat{l}(m_i; i) \triangleq \max_{\xi \in \Delta B^{\gamma;l}(i, m_i)} \#\{t | \xi(t) - \xi(t^-) > \gamma_i\},$$

for  $i \in \{1, \dots, d-1\}$ . By (3.5), we have that

$$\mathbb{P}(\bar{X}_n^{(i)} \in \cdot | \bar{X}_n \in B^\gamma(m_1, \dots, m_{d-1})) = \sum_{q_i = \check{l}(m_i; i)}^{\infty} h_{2; q_i} \mathbb{P}(\bar{X}_n^{(i)} \in \cdot | \Delta B^{\gamma;l}(i, m_i), N^{(i)}(n) = q_i),$$

where  $h_{2; q_i} \triangleq \mathbb{P}(\Delta B^{\gamma;l}(i, m_i), N^{(i)}(n) = q_i) / \mathbb{P}(\Delta B^{\gamma;l}(i, m_i))$  satisfies  $\sum_{q_i \geq \check{l}(m_i; i)} h_{2; q_i} = 1$ . Note that

$$\mathbb{P}(\Delta B^{\gamma;l}(i, m_i), N^{(i)}(n) = q_i) = e^{-\lambda_i n} \frac{(\lambda_i n)^{q_i}}{q_i!} \left( \sum_{i = \check{l}(m_i; i)}^{\hat{l}(m_i; i) \wedge q_i} \binom{q_i}{i} \mathbb{P}(W^{(i)}(1) > n\gamma)^i \left(1 - \mathbb{P}(W^{(i)}(1) > n\gamma)\right)^{q_i - i} \right).$$

Therefore, it suffices to consider sampling  $\bar{X}_n^{(i)}$  under  $\mathbb{P}(\cdot | \Delta B^{\gamma;l}(i, m_i), N^{(i)}(n) = q_i)$ . Again, we can proceed a similar approach as in Section 3.1:

1. Sample  $\{b_k\}_{k \leq l}$  uniformly from  $\mathcal{C}(\{1, \dots, q_i\}, \check{l}(m_i; i))$ ;
2. sample each  $W^{(i)}(b_k)$ ,  $k \leq q_i$ , conditional on  $W^{(i)}(1) > n\gamma_i$ ;
3. sample  $W^{(i)}(q'_i)$ ,  $q'_i \leq q_i$ ,  $q'_i \notin \{b_k\}_{k \leq l^*}$ , under the nominal measure.
4. accept  $(W^{(i)}(1), \dots, W^{(i)}(q_i))$  with probability

$$a(W^{(i)}(1), \dots, W^{(i)}(q_i)) = \left( \frac{\#\{j \in \{1, \dots, q_i\} | W^{(i)}(j) > n\gamma_i\}}{\check{l}(m_i; i)} \right)^{-1} \mathbb{1}_{\{\#\{j \in \{1, \dots, q_i\} | W^{(i)}(j) > n\gamma_i\} \leq \hat{l}(m_i; i)\}}.$$

Finally, we are able to give the pseudocode of this sampling scheme in Algorithm 3 below. For its expected running time, an analogous result to Proposition 3.1 is formulated in Proposition 3.5.

**Proposition 3.5.** *Let  $T_{\text{alg3}}(n)$  denote the expected running time of Algorithm 3. Under the assumption that  $W^{(i)}(1)$  is regularly varying of index  $-\beta_i < -1$ , for all  $i \in \{1, \dots, d\}$ , we have that  $T_{\text{alg3}}(n)$  is uniformly bounded from above w.r.t.  $n$ , i.e.  $\max_{n \geq 0} T_{\text{alg3}}(n) < \infty$ .*

*Proof.* Analogous to the proof Proposition 3.1. □

The discussion above shows that sampling from the conditional distribution  $\mathbb{Q}_\gamma(\cdot)$  is tractable. As we mentioned in the introduction, our estimator is straightforward to implement. Moreover, its strong efficiency, which is formulated in Theorem 3.6, can be proved based on Lemma 3.7. Without introducing any new notations, we formulate a corollary to address a special case, where it is sufficient to consider a smaller (in the sense of cardinality) set than  $J_{(l_1^*, \dots, l_d^*)}$  as in Definition 1.

**Theorem 3.6.** *Let  $B_n^\gamma \triangleq \{\bar{X}_n \in B^\gamma\}$ , where  $B^\gamma$  is as defined in (3.4). Under Assumption 2, there exists  $\gamma$  such that the estimator given by*

$$Z_n = \frac{\mathbb{1}_{A_n}}{w + \frac{1-w}{\mathbb{P}(B_n^\gamma)} \mathbb{1}_{B_n^\gamma}}$$

*is strongly efficient for estimating  $\mathbb{P}(A_n)$ .*

---

**Algorithm 3** Generating the sample path of  $\bar{X}_n^{(1)}, \dots, \bar{X}_n^{(d)}$  under  $\mathbb{Q}_\gamma$

---

**Input:**  $\gamma \in \mathbb{R}^d$   
sample  $(m_1, \dots, m_{d-1}) \sim h_{1;m_1, \dots, m_{d-1}}$   
**for**  $i = 1$  to  $d$  **do**  
  sample  $q_i \sim h_{2;q_i}$   
   $R \leftarrow \text{true}$   
  **while**  $R = \text{true}$  **do**  
    sample  $\{b_k\}_{k \leq \tilde{l}(m_i; i)} \sim \text{unif}(\mathcal{C}(\{1, \dots, q_i\}, \tilde{l}(m_i; i)))$   
    **for**  $j = 1$  to  $q_i$  **do**  
      **if**  $j \in \{b_k\}_{k \leq \tilde{l}(m_i; i)}$  **then**  
        sample  $W^{(i)}(j) \sim W^{(i)}(1) \mid W^{(i)}(1) > n\gamma_i$   
      **else**  
        sample  $W^{(i)}(j) \sim W^{(i)}(1)$   
      **end if**  
    **end for**  
     $c \leftarrow \#\{j \in \{1, \dots, q_i\} \mid W^{(i)}(j) > n\gamma_i\}$   
    **if**  $c < \hat{l}(m_i; i)$  **then**  
       $a \leftarrow (\hat{l}(m_i; i)^c)^{-1}$   
    **else**  
       $a \leftarrow 0$   
    **end if**  
    sample  $u \sim \text{uniform}[0, 1]$   
    **if**  $u < a$  **then**  
       $R \leftarrow \text{false}$   
    **else**  
       $R \leftarrow \text{true}$   
    **end if**  
  **end while**  
**end for**  
**return**  $\bar{X}_n^{(1)}, \dots, \bar{X}_n^{(d)}$

---

*Proof.* See Section 7. □

**Lemma 3.7.** *Let  $B^\gamma$  be as defined in (3.4). Under Assumption 2, we have that*

$$\mathbb{P}(\bar{X}_n \in B^\gamma) = \mathcal{O}(\mathbb{P}(\bar{X}_n \in A)).$$

*Moreover, there exists  $\gamma$ , such that*

$$\mathbb{P}(\bar{X}_n \in A \cap (B^\gamma)^c) = o(\mathbb{P}(\bar{X}_n \in A)^2).$$

*Proof.* See Section 7. □

**Corollary 3.8.** *Along with Assumption 2, we assume additionally that there exists an index set  $I \subseteq J_{(l_1^*, \dots, l_d^*)}$  and  $r > 0$  such that: for every  $\xi \in A$ , there exists  $(l_1, \dots, l_d) \in I$  satisfying  $d(\xi, \mathbb{C}_{(l_1, \dots, l_d)}) \geq r$ . Set  $\bar{J}_{(l_1^*, \dots, l_d^*)} = I \setminus \Delta I$ , where  $(l_1, \dots, l_d) \in \Delta I$  if and only if*

- $(l_1, \dots, l_d) \in I$  satisfies that  $\mathcal{I}(l_1, \dots, l_d) > 2\mathcal{I}(l_1^*, \dots, l_d^*)$ ; and
- for every  $(l'_1, \dots, l'_d) \in I \setminus \{(l_1, \dots, l_d)\}$ , we have that  $\mathcal{I}(l_1, \dots, l_d) \neq \mathcal{I}(l'_1, \dots, l'_d)$ .

*Setting  $B_n^\gamma = \{\bar{X}_n \in B^\gamma\}$  with*

$$B^\gamma \triangleq \bigcup_{(l_1, \dots, l_d) \in \bar{J}_{(l_1^*, \dots, l_d^*)}} B^{\gamma; l},$$

*there exists  $\gamma$  such that the estimator given by*

$$Z_n = \frac{\mathbb{1}_{A_n}}{w + \frac{1-w}{\mathbb{P}(B_n^\gamma)} \mathbb{1}_{B_n^\gamma}}$$

*is strongly efficient for estimating  $\mathbb{P}(A_n)$ .*

*Proof.* Analogous to the proofs of Lemma 3.7 and Theorem 3.6. □

*Remark 5.* Even though our simulation algorithm is constructed in the context of Poisson processes with positive jump distributions, it can be easily generalized to the case, where the jump distributions are regularly varying at both  $-\infty$  and  $\infty$  (for details see the proof of Theorem 3.5 in [23] and the references therein).

### 3.3 Extension to random walks

Let  $S_k$ ,  $k \geq 0$  be a centered random walk with increments  $\{Y_k\}_{k \geq 1}$ . Let  $\mathbb{P}(Y_1 \leq -x)$  be regularly varying with index  $-\alpha$  and let  $\mathbb{P}(Y_1 \geq x)$  be regularly varying with index  $-\beta$ . Define  $\bar{S}_n(t) = S_{\lfloor nt \rfloor} / n$ ,  $t \geq 0$ , where

$$\bar{S}_n(t) = S_{\lfloor nt \rfloor} / n, \quad t \geq 0. \tag{3.6}$$

In this subsection, we want to design an efficient simulation algorithm for estimating the probability of  $\bar{S}_n \in A$ . As in Section 3.1 and Section 3.2, we make the following assumption for the set  $A$ .

**Assumption 3.** Let  $A$  be a measurable set. Assume that  $A$  is bounded away from  $\mathbb{D}_{< l_-^*; l_+^*}$ , where  $(l_-^*, l_+^*)$  is the unique solution of the minimization problem given by (2.2). Moreover, assume that  $C_{l_-^*, l_+^*}(A^\circ) > 0$ .

Then, we construct the auxiliary set  $B^\gamma$  as follows.

**Definition 2.** Let  $(l_-^*, l_+^*)$  denote the unique solution to (2.2), and let

$$J_{l_-^*; l_+^*} \triangleq \left\{ (l_-, l_+) \in \mathbb{Z}_+^2 \setminus I_{< l_-^*; l_+^*} \mid (m_-, m_+) \prec (l_-, l_+) \text{ implies } (m_-, m_+) \in I_{< l_-^*; l_+^*} \right\},$$

where  $I_{<l_-^*, l_+^*} \triangleq \{(l_-, l_+) \in \mathbb{Z}_+^2 \setminus \{(l_-^*, l_+^*)\} \mid (\alpha - 1)l_- + (\beta - 1)l_+ \leq (\alpha - 1)l_-^* + (\beta - 1)l_+^*\}$ . For  $\gamma_- > 0$  and  $\gamma_+ > 0$ , define

$$B^\gamma \triangleq \bigcup_{(l_-, l_+) \in I_{<l_-^*, l_+^*}} B^{\gamma; l_-^*, l_+^*}, \quad (3.7)$$

where  $B^{\gamma; l_-^*, l_+^*} \triangleq \{\xi \in \mathbb{D} \mid \#\{t \mid \xi(t^-) - \xi(t) > \gamma_-\} \geq l_-^*, \#\{t \mid \xi(t) - \xi(t^-) > \gamma_+\} \geq l_+^*\}$ .

Defining  $A_n \triangleq \{\bar{S}_n \in A\}$  and  $B_n^\gamma \triangleq \{\bar{S}_n \in B^\gamma\}$ , we propose an estimator for  $\mathbb{P}(\bar{S}_n \in A)$  that is given by (3.3). Note that, computing  $\mathbb{P}(\bar{S}_n \in B^\gamma)$ , as well as generating the sample path  $\bar{S}_n$  under  $\mathbb{Q}^\gamma$  can be achieved by following similar strategies as in Section 3.1 and Section 3.2. Hence, the details are omitted (for examples, see Section 4 and Section 5 below). We state the strong efficiency of our estimator in the following theorem.

**Theorem 3.9.** *Let  $B_n^\gamma \triangleq \{\bar{X}_n \in B^\gamma\}$ , where  $B^\gamma$  is as defined in (3.7). Under Assumption 3, there exists  $\gamma_-$  and  $\gamma_+$  such that the estimator given by*

$$Z_n = \frac{\mathbb{1}_{A_n}}{w + \frac{1-w}{\mathbb{P}(B_n^\gamma)} \mathbb{1}_{B_n^\gamma}}$$

is strongly efficient for estimating  $\mathbb{P}(A_n)$ .

*Proof.* It is analogous to the proofs of Theorem 3.6 and Lemma 3.7, by using Result 2.3.  $\square$

With the results presented in this section at hand, we are able to apply our general simulation algorithm to three examples in the next sections. These examples can be found in the applications of mathematical finance, actuarial science and queueing networks.

## 4 An application to finite-time ruin probabilities

### Problem settings

Let  $S_k$ ,  $k \geq 0$ , be a centered random walk with increments  $\{Y_k\}_{k \geq 1}$ . Moreover, let  $\mathbb{P}(Y_1 > x)$  be regularly varying at  $+\infty$  with index  $-\beta$ . For  $c \geq 0$ , define

$$A_n \triangleq \left\{ \max_{0 \leq k \leq n} Y_k \leq nb, \max_{0 \leq k \leq n} S_k - ck \geq na \right\}.$$

Additionally, we make a technical assumption that  $a$  is not a multiple of  $b$ . We are interested in computing  $\mathbb{P}(A_n)$ . This probability is particularly interesting, since it is related to, for example, insurance, where huge claims may be reinsured and therefore are irrelevant in the sense of estimating the finite time ruin probability of an insurance company.

### Large deviations results

The rare-event probability can be estimated efficiently using the technique introduced in Section 3. To see this, define

$$A \triangleq \left\{ \xi \in \mathbb{D}: \sup_{t \in [0,1]} [\xi(t) - ct] \geq a; \sup_{t \in [0,1]} [\xi(t) - \xi(t^-)] \leq b \right\},$$

and  $\bar{S}_n \triangleq \{\bar{S}_n(t)\}_{t \in [0,1]}$ , where  $\bar{S}_n(t) = S_{[nt]}/n$  for  $t \geq 0$ . Note that  $\mathbb{P}(A_n) = \mathbb{P}(\bar{S}_n \in A)$ . Set  $l^* = \lceil a/b \rceil$ . Intuitively,  $l^*$  should be the key parameter, as it takes at least  $l^*$  jumps of size  $b$  to cross level  $a$ . This intuition has been made rigorous by Rhee et al. in [23, Section 5.1], where the authors show that  $A$  is bounded away from  $\mathbb{D}_{<l^*}$ , and hence,  $\mathbb{P}(A_n) = \Theta(n^{l^*} \mathbb{P}(S_1 \geq n)^{l^*})$ .

## Construction of $B^\gamma$

Since  $A$  is bounded away from  $\mathbb{D}_{<l^*}$ , we can set

$$B^\gamma = \{\xi \in \mathbb{D} \mid \#\{t \mid \xi(t) - \xi(t^-) > \gamma\} \geq l^*\},$$

and

$$B_n^\gamma = \{\bar{S}_n \in B^\gamma\} = \{\#\{k \in \{1, \dots, n\} \mid Y_k > n\gamma\} \geq l^*\},$$

where  $\gamma$  is the parameter that needs to be tuned. For the completeness of our algorithm, we give a closed-form expression for  $\mathbb{P}(B_n^\gamma)$ . Let  $p$  denote the probability of  $\mathbb{P}(Y_1 > \gamma n)$ , then we have that

$$\mathbb{P}(B_n^\gamma) = \sum_{i=l^*}^n \binom{n}{i} p^i (1-p)^{n-i} = 1 - \sum_{i=0}^{l^*-1} \binom{n}{i} p^i (1-p)^{n-i}, \quad (4.1)$$

where the latter representation in (4.1) is for numerical purposes.

## Choice of $\gamma$

As we mentioned in Remark 3, a strategy of choosing the parameters  $\gamma$  needs to be discussed in the next step. From the proof of Theorem 3.6, it is sufficient to select  $\gamma$  such that  $\mathbb{P}(A_n \cap (B_n^\gamma)^c) = o(\mathbb{P}(A_n)^2)$ . We propose to select  $\gamma$  such that  $(a - (l^* - 1)b)/\gamma \notin \mathbb{Z}_+$ , and that

$$\left\lceil \frac{a - (l^* - 1)b}{\gamma} \right\rceil > l^* + 1. \quad (4.2)$$

In view of Theorem 3.9, it is sufficient to show that  $A \cap (B^\gamma)^c$  is bounded away from  $\mathbb{D}_{<2l^*+1}$  with  $\gamma$  satisfying (4.2). To see this, choose  $\theta$  with  $d(\theta, \mathbb{D}_{<2l^*+1}) < r$ . This implies that there exists  $\xi \in \mathbb{D}_{<2l^*+1}$  satisfying  $d(\theta, \xi) < r$  and  $\xi(t) = \sum_{j=1}^{2l^*} x_j \mathbb{1}_{[u_j, 1]}(t)$ . In particular, there exists a homeomorphism  $\lambda: [0, 1] \rightarrow [0, 1]$  satisfying

$$\|\lambda - id\|_\infty \vee \|\xi \circ \lambda - \theta\|_\infty < r. \quad (4.3)$$

For  $\theta \in A$ , using (4.3) and the identity  $\xi \circ \lambda = \theta + (\xi \circ \lambda - \theta)$ , we conclude that the following holds:

1.  $x_j < b + 2r$ , for every  $j \in \{1, \dots, 2l^*\}$ ; and
2. there exists  $t'$  such that

$$\sum_{u_j \leq 1} x_j \geq \sum_{u_j \leq \lambda(t')} x_j > a - 2r.$$

This implies that

$$\sum_{j \geq l^*} x_j > a - 2r - (l^* - 1)(b + 2r).$$

Moreover, for  $\theta \in (B^\gamma)^c$ , every jump of  $\xi$  should be bounded by  $\gamma + 2r$  after having  $l^* - 1$  jumps with size bigger than  $b$ . Due to the fact that  $\gamma$  satisfies (4.2) and that  $a$  is not a multiple of  $b$ , we obtain the result by choosing  $r$  sufficiently small.

## Sampling from $\mathbb{Q}_\gamma$

Summarizing the discussion from previous paragraphs, we are able to propose a strongly efficient estimator for  $\mathbb{P}(A_n)$  that is given by (3.3). As the last ingredient of our simulation algorithm, a strategy of sampling from  $\mathbb{Q}_\gamma(\cdot)$  ( $= \mathbb{P}(\cdot | B_n^\gamma)$ ) needs to be discussed. We use a similar strategy as in Algorithm 3 and formulate the pseudocode in Algorithm 4.



---

**Algorithm 4**

---

```
R  $\leftarrow$  true
while R = true do
  sample  $(i_1, \dots, i_{l^*})$  uniformly from  $\mathcal{C}(\{1, \dots, n\}, l^*)$ 
  for  $j = 1$  to  $n$  do
    if  $j \in \{i_1, \dots, i_{l^*}\}$  then
      sample  $Y_j \sim Y_1 \mid \gamma n < Y_1 \leq bn$ 
    else
      sample  $Y_j \sim Y_1$ 
    end if
  end for
  sample  $u \sim \text{uniform}[0, 1]$ 
   $c \leftarrow \#\{m \in \{1, \dots, n\} \mid \gamma n < Y_1 \leq bn\}$ 
   $a \leftarrow \binom{c}{l^*}$ 
  if  $u < a^{-1}$  then
    R  $\leftarrow$  false
  else
    R  $\leftarrow$  true
  end if
end while
```

---

## Numerical results

Finally, we investigate our algorithm numerically based on a concrete example. Let  $Y_1 = Y'_1 - \mathbb{E}Y'_1$ , where  $\mathbb{P}(Y'_1 > t) = (t_r/t)^\beta$ , i.e.  $Y'_1$  follows a Pareto distribution with scale parameter  $t_r$  and shape parameter  $\beta$ . Let  $t_r = 1$ ,  $a = 2$  and  $b = 1.2$ . In Table 1 we select  $c = 0.05$ ,  $w = 0.05$ ,  $\gamma = 0.13$  and summarize the estimated probability and the level of precision (ratio between the radius of the 95% confidence interval and the estimated value) for different combinations of  $n$  and  $\beta$  (based on 200000 samples). We observe that, for different values of  $\beta$ , the precision stays roughly constant as  $n$  grows. This confirms our theoretical results.

## 5 An application in barrier option pricing

In this section we consider an application that arises in the context of financial mathematics; in particular we consider a down-in barrier option (see Section 11.3 in [26]).

### Problem settings

Let  $S_k$ ,  $k \geq 0$ , be a centered random walk with increments  $\{Y_k\}_{k \geq 1}$ . Let  $\mathbb{P}(Y_1 \leq -x)$  be regularly varying with index  $-\alpha$  and let  $\mathbb{P}(Y_1 \geq x)$  be regularly varying with index  $-\beta$ . Let  $a$ ,  $b$  and  $c$  be positive real numbers. We provide a strongly efficient estimator for the probability of

$$A_n \triangleq \left\{ S_n \geq bn, \min_{0 \leq k \leq n} S_k + ck \leq -an \right\},$$

which can be interpreted as the chance of exercising a down-in barrier option. This application is interesting, since, as we will see, the large deviations behavior of  $\mathbb{P}(A_n)$  is caused by two large jumps.

### Large deviations results

To begin with, define

$$A \triangleq \left\{ \xi \in \mathbb{D} : \xi(1) \geq b, \inf_{0 \leq t \leq 1} \xi(t) + ct \leq -a \right\}.$$

Est PR	$n = 1100$	$n = 1400$	$n = 1700$	$n = 2000$	$n = 2300$	$n = 2600$
$\beta = 1.45$	$2.188 \times 10^{-4}$ 0.052	$1.691 \times 10^{-4}$ 0.054	$1.376 \times 10^{-4}$ 0.055	$1.204 \times 10^{-4}$ 0.055	$1.076 \times 10^{-4}$ 0.055	$9.603 \times 10^{-5}$ 0.055
Est PR	$n = 1100$	$n = 1400$	$n = 1700$	$n = 2000$	$n = 2300$	$n = 2600$
$\beta = 1.60$	$3.243 \times 10^{-5}$ 0.062	$2.335 \times 10^{-5}$ 0.064	$1.509 \times 10^{-5}$ 0.071	$1.324 \times 10^{-5}$ 0.069	$1.150 \times 10^{-5}$ 0.069	$8.809 \times 10^{-6}$ 0.071
Est PR	$n = 2400$	$n = 2600$	$n = 2800$	$n = 3000$	$n = 3200$	$n = 3400$
$\beta = 1.75$	$8.903 \times 10^{-7}$ 0.095	$8.416 \times 10^{-7}$ 0.092	$7.261 \times 10^{-7}$ 0.094	$6.166 \times 10^{-7}$ 0.096	$5.605 \times 10^{-7}$ 0.096	$5.576 \times 10^{-7}$ 0.093
Est PR	$n = 3000$	$n = 3500$	$n = 4000$	$n = 4500$	$n = 5000$	$n = 5500$
$\beta = 1.90$	$5.813 \times 10^{-8}$ 0.119	$3.994 \times 10^{-8}$ 0.125	$3.045 \times 10^{-8}$ 0.126	$2.766 \times 10^{-8}$ 0.120	$2.077 \times 10^{-8}$ 0.126	$1.758 \times 10^{-8}$ 0.126

Table 1: Estimated rare-event probability and level of precision for the application as described in Section 4 w.r.t. different combinations of  $n$  and  $\beta$ .

Obviously, we have that  $(l_-^*, l_+^*) = (1, 1)$ , where  $(l_-^*, l_+^*)$  denotes the solution to (2.2). To verify the topological property of  $A$ , we define  $m, \pi_1: \mathbb{D} \rightarrow \mathbb{R}$  by  $m(\xi) = \inf_{0 \leq t \leq 1} \{\xi(t) + ct\}$ , and  $\pi_1(\xi) = \xi(1)$ . Note that  $F, \pi_1$  and  $m$  are continuous, therefore

$$F^{-1}(A) = m^{-1}(-\infty, -a] \cap \pi_1^{-1}[b, \infty)$$

is a closed set. By adapting the results in [23, Section 5.2], it can be shown that, for any arbitrary  $i \geq 0$ ,  $\mathbb{D}_{i;0}$  and  $\mathbb{D}_{0;i}$  are bounded away from  $m^{-1}(-\infty, -a]$  and  $\pi_1^{-1}[b, \infty)$ , respectively. Hence,  $A$  is bounded away from  $\mathbb{D}_{<1;1}$ . Applying Result 2.3, we obtain that  $\mathbb{P}(X_n \in A) = \Theta(n^2 \mathbb{P}(S_1 \geq n) \mathbb{P}(S_1 \leq n))$ .

### Construction of $B^\gamma$

Now we are in the framework of Theorem 3.9. Note that, by Definition 2, we have that  $J_{1;1} = \{(1, 1), (l, 0), (0, m)\}$ , where

$$l = \min \{l' \in \mathbb{Z}_+ \mid (l' - 1)(\beta - 1) > (\alpha - 1)\}, \quad m = \min \{m' \in \mathbb{Z}_+ \mid (m' - 1)(\alpha - 1) > (\beta - 1)\}.$$

However, adapting the idea behind Corollary 3.8 together with the fact that  $A$  is bounded away from both  $\mathbb{D}_{i;0}$  and  $\mathbb{D}_{0;i}$ , it is sufficient to consider  $\tilde{J}_{1;1} = \{(1, 1)\}$ . Hence, we can set

$$B^\gamma = \{\xi \in \mathbb{D} \mid \#\{t \mid \xi(t^-) - \xi(t) > \gamma_-\} \geq 1, \#\{t \mid \xi(t) - \xi(t^-) > \gamma_+\} \geq 1\}.$$

As we mentioned in the introduction, it is possible that estimators may be crafted specifically for the events of interest, in order to obtain (up to constant factors) better performance. Due to the fact that at least one downward jump should happen *before* upward jumps, without introducing new notations, we can modify  $B^\gamma$  such that

$$B^\gamma = \{\xi \in \mathbb{D} \mid \exists t_1 < t_2 : \xi(t_1^-) - \xi(t_1) > \gamma_-, \xi(t_2) - \xi(t_2^-) > \gamma_+\}.$$

This implies that  $B_n^\gamma = \{\exists i < j : Y_i < -\gamma_- n, Y_j > \gamma_+ n\}$ . By a straightforward computation, we obtain that

$$\mathbb{P}(B_n^\gamma) = 1 - \frac{p_2}{p_2 - p_1} (1 - p_1)^n + \frac{p_1}{p_2 - p_1} (1 - p_2)^n,$$

where  $p_1 \triangleq \mathbb{P}(Y_1 > \gamma_+ n)$  and  $p_2 \triangleq \mathbb{P}(Y_1 < -\gamma_- n)$ .

## Choice of $\gamma_-$ and $\gamma_+$

We discuss here the strategy of choosing the parameters  $\gamma_-$  and  $\gamma_+$ . From the proof of Theorem 3.6, it is sufficient to select  $\gamma_-, \gamma_+$  such that  $\mathbb{P}(A_n \cap (B_n^\gamma)^c) = o(\mathbb{P}(A_n)^2)$ . Hence, we propose to choose  $\gamma_-$  and  $\gamma_+$  such that  $((a+b)/\gamma_+, a/\gamma_-) \notin \mathbb{Z}_+^2$ , and that

$$\min \left\{ (\alpha - 1) + \left\lceil \frac{a+b}{\gamma_+} \right\rceil (\beta - 1), \left\lceil \frac{a}{\gamma_-} \right\rceil (\alpha - 1) + (\beta - 1) \right\} > 2(\alpha + \beta - 2). \quad (5.1)$$

W.l.o.g. we assume that  $\lceil a/\gamma_- \rceil (\alpha - 1) + (\beta - 1)$  is the unique minimum of (5.1). It suffices to prove that  $A \cap (B^\gamma)^c$  is bounded away from  $\mathbb{D}_{\lceil a/\gamma_2 \rceil; 1}$ . To show that  $\bigcup_{(l_-, l_+)} \mathbb{D}_{\langle l_-; l_+}$  with  $l_- \leq \lceil a/\gamma_2 \rceil - 1$  is bounded away from  $A \cap (B^\gamma)^c$ , choose  $\theta$  with  $d\left(\theta, \bigcup_{(l_-, l_+)} \mathbb{D}_{\langle l_-; l_+}\right) < r$ . This implies that there exists  $\xi \in \bigcup_{(l_-, l_+)} \mathbb{D}_{\langle l_-; l_+}$  satisfying  $d(\theta, \xi) < r$ , where  $\xi = \sum_{k=1}^{l_+} x_k \mathbb{1}_{[u_k, 1]}(t) - \sum_{k=1}^{l_-} y_k \mathbb{1}_{[v_k, 1]}(t)$ . In particular, there exists homeomorphism  $\lambda: [0, 1] \rightarrow [0, 1]$  satisfying

$$\|\lambda - id\|_\infty \vee \|\xi \circ \lambda - \theta\|_\infty < r. \quad (5.2)$$

Using (5.2) and the identity  $\xi \circ \lambda = \theta + (\xi \circ \lambda - \theta)$ , we conclude that, for  $\theta \in (B^\gamma)^c$  and  $t \in [0, 1]$ , at least one of the following holds:

- $x_k \leq \gamma_+ + 2r$ , for every  $u_k \geq t$ ; or
- $y_k \leq \gamma_- + 2r$ , for every  $v_k < t$ .

For  $\theta \in m^{-1}(-\infty, -a]$ , by (5.2) there exists  $t'$  such that

$$\sum_{u_j \leq \lambda(t')} x_j - \sum_{v_j \leq \lambda(t')} y_j < -a + 3r. \quad (5.3)$$

Moreover, we can assume that  $y_j \leq \gamma_- + 2r$  for  $j$  satisfying  $v_j \leq \lambda(t')$ . Otherwise  $x_j$  is bounded by  $\gamma_+ + 2r$  for  $j$  satisfying  $v_j > \lambda(t')$ . By choosing  $r$  sufficiently small, this leads to a contradiction of that  $\theta \in \pi_1^{-1}[b, \infty)$  and  $\lceil a/\gamma_- \rceil (\alpha - 1) + (\beta - 1)$  is the minimum of (5.1). Hence, (5.3) implies that

$$\left( \left\lceil \frac{a}{\gamma_-} \right\rceil - 1 \right) (\gamma_- + 2r) > a - 3r.$$

Since  $(\lceil a/\gamma_- \rceil - 1)\gamma_- < a$ , choosing  $r$  sufficiently small we obtain the result. Similarly, it can be shown that  $A \cap (B^\gamma)^c$  is bounded away from  $\bigcup_{(l_-, l_+)} \mathbb{D}_{\langle l_-; l_+}$  for  $l_+ \leq \lceil (a+b)/\gamma_+ \rceil - 1$ .

## Sampling from $\mathbb{Q}_\gamma$

Summarizing the discussion in the previous paragraphs, we are able to propose a strongly efficient estimator for  $\mathbb{P}(A_n)$  that is given by (3.3). As in Section 4, a strategy of sampling from  $\mathbb{Q}_\gamma(\cdot)$  needs to be discussed. Even though  $B^\gamma$  is modified to obtain smaller relative error, a similar strategy as in Algorithm 3 can be used here. We formulate the pseudocode in Algorithm 5.

## Numerical results

We end this section with a numerical investigation (based on 200000 samples). Let  $Y_1 = Y_1' - \mathbb{E}Y_1'$ , where  $Y_1'$  is a random variable with density function  $f_Y$  that is given by

$$f_Y = p_1 \left( \frac{t_r}{y} \right)^\beta \mathbb{1}_{(t_r, \infty)}(y) + p_2 \left( \frac{t_l}{y} \right)^\alpha \mathbb{1}_{(-\infty, t_l)}(y) + (1 - p_1 - p_2) \frac{1}{t_r - t_l} \mathbb{1}_{[t_l, t_r]}(y),$$

where  $t_r > 0$  and  $t_l < 0$ . In the following example we choose  $t_r = -t_l = 1$ ,  $p_1 = p_2 = 1/3$ ,  $a = 2$  and  $b = 1.5$ . In Table 2 we compare the estimated rare-event probability and precision w.r.t. different values of  $n$ ,  $\alpha$  and  $\beta$ . We observe that the precision stays roughly constant as  $n$  increases for different combinations of  $\alpha$  and  $\beta$ , which suggests the strong efficiency of our estimator.

---

**Algorithm 5** Sampling from  $\mathbb{Q}_\gamma$ : a modification of Algorithm 3

---

**Input:**  $\gamma_-, \gamma_+$   
 $R \leftarrow \text{true}$   
**while**  $R = \text{true}$  **do**  
  sample  $(i_1, i_2)$  uniformly from  $\mathcal{P}(\{1, \dots, n\}, 2)$   
  **for**  $j = 1$  to  $n$  **do**  
    **if**  $j \neq i_1$  **and**  $j \neq i_2$  **then**  
      sample  $Y_j \sim Y_1$   
    **else if**  $j = i_1$  **then**  
      sample  $Y_{i_1} \sim Y_1 \mid Y_1 < -\gamma_- n$   
    **else**  
      sample  $Y_{i_2} \sim Y_2 \mid Y_2 > \gamma_+ n$   
    **end if**  
  **end for**  
  sample  $u \sim \text{uniform}[0, 1]$   
  **if**  $(Y_1, \dots, Y_n) \in B_n^\gamma$  **then**  
     $a \leftarrow (\#\{(l, m) \in \{1, \dots, n\}^2 \mid Y_l < -\gamma_- n, Y_m > \gamma_+ n\})^{-1}$   
  **else**  
     $a \leftarrow 0$   
  **end if**  
  **if**  $u < a$  **then**  
     $R \leftarrow \text{false}$   
  **else**  
     $R \leftarrow \text{true}$   
  **end if**  
**end while**

---

Est PR	$n = 250$	$n = 500$	$n = 750$	$n = 1000$	$n = 1250$	$n = 1500$
$\alpha = 2, \beta = 1.5$	$3.913 \times 10^{-7}$	$1.370 \times 10^{-7}$	$6.992 \times 10^{-8}$	$4.539 \times 10^{-8}$	$3.305 \times 10^{-8}$	$2.471 \times 10^{-8}$
	0.043	0.043	0.044	0.044	0.044	0.044
$\alpha = 1.8, \beta = 1.7$	$3.322 \times 10^{-7}$	$1.154 \times 10^{-7}$	$6.040 \times 10^{-8}$	$3.840 \times 10^{-8}$	$2.870 \times 10^{-8}$	$2.225 \times 10^{-8}$
	0.037	0.037	0.038	0.038	0.038	0.037
$\alpha = 2.3, \beta = 2$	$1.923 \times 10^{-9}$	$4.004 \times 10^{-10}$	$1.491 \times 10^{-10}$	$7.601 \times 10^{-11}$	$4.632 \times 10^{-11}$	$3.072 \times 10^{-11}$
	0.053	0.053	0.054	0.054	0.054	0.054
$\alpha = 2.7, \beta = 1.8$	$6.838 \times 10^{-10}$	$1.121 \times 10^{-10}$	$4.092 \times 10^{-11}$	$2.079 \times 10^{-11}$	$1.105 \times 10^{-11}$	$6.896 \times 10^{-12}$
	0.068	0.070	0.070	0.069	0.071	0.071

Table 2: Estimated rare-event probability and level of precision for the application as described in Section 5 w.r.t. different combinations of  $n$ ,  $\alpha$  and  $\beta$ .

## 6 An application to queueing networks

In this section, an application to queueing networks is considered. More specifically, the probability of the number of customers in a subset of the system crossing a high level is estimated. Although some particular cases exist that allow for an explicit analysis (see e.g. Section 13 in [8]), it is hard to come up with exact results for the distribution of the workload process in general. Hence, implementing our algorithm in such a context is particularly interesting.

### 6.1 Model description and preliminaries

To be specific, we consider a  $d$ -dimensional stochastic fluid model. Suppose that jobs arrive to the  $i$ -th station in the network according to a Poisson process with unit rate, which is denoted by  $\{N^{(i)}(t)\}_{t \geq 0}$  and independent of  $\{N^{(j)}\}$  for  $j \neq i$ . Moreover, the  $k$ -th arrival of the  $i$ -th station brings a job of size  $W^{(i)}(k)$ . We are assuming that  $\{W(k) \triangleq (W_1(k), \dots, W_d(k))^T\}_{k \geq 1}$  is a sequence of i.i.d. positive random vectors and that  $\{W(k)\}_{k \geq 1}$  is independent of  $\{N(t)\}_{t \geq 0}$ . Therefore, the total amount of external work that arrives to the  $i$ -th station up to time  $t$  is given by  $J^{(i)}(t) = \sum_{k=1}^{N^{(i)}(t)} W^{(i)}(k)$ . Now, assume that the workload at the  $i$ -th station is processed as a fluid by the server at a rate  $r_i$  and that a proportion  $Q_{ij} \geq 0$  of the fluid processed by the  $i$ -th station is routed to the  $j$ -th server. Moreover, we assume that  $Q$  is a substochastic matrix with  $Q_{ii} = 0$  and that  $Q^n \rightarrow 0$  as  $n \rightarrow \infty$ . The dynamics of the model are expressed formally by the so-called Skorokhod map (for details see e.g. [24], [25], [20] etc.), that is defined in terms of a pair of processes  $(Z, Y)$  satisfying a stochastic differential equation that we shall describe now. Let  $R = (I - Q)^T$ ,  $r = (r_1, \dots, r_d)^T$ ,  $X(t) \triangleq J(t) - Rrt$  and  $Z^{(i)}(t)$  denote the workload of the  $i$ -th station at time  $t$ , for given  $Z^{(i)}(0)$ , we have that

$$dZ(t) = dX(t) + RdY(t), \quad (6.1)$$

where  $Y(\cdot)$  encodes the minimal amount of pushing required to keep  $Z(\cdot)$  non-negative. In order to describe how to characterize the solution  $(Z, Y)$  to (6.1), we need to introduce some notations. Let  $\psi : \mathbb{D}^d \rightarrow \mathbb{D}_+^d$  with

$$\psi(x) \triangleq \inf \{w \in \mathbb{D}_+^d \mid x + Rw \geq 0\},$$

i.e.,

$$\psi^{(i)}(x)(t) \triangleq \inf \left\{ w^{(i)}(t) \in \mathbb{R} \mid w \in \mathbb{D}_+^d, x + Rw \geq 0 \right\}, \quad \text{for all } i \text{ and } t,$$

and  $\phi : \mathbb{D}^d \rightarrow \mathbb{D}^d$  with  $\phi(x) \triangleq x + R\psi(x)$ . The following results summarize useful properties and characterizations of the Skorokhod mappings  $\psi$ ,  $\phi$ , as well as the workload process  $Z(t)$ .

**Result 6.1** (Theorem 14.2.1, Theorem 14.2.5 and Theorem 14.2.7 of [27]). *For all  $x \in \mathbb{D}^d$ , the mappings  $\psi$  and  $\phi$  are well-defined. Moreover,  $\psi$  and  $\phi$  are Lipschitz continuous w.r.t. both the uniform metric and the Skorokhod  $J_1$  metric. If  $Y(t) \triangleq \psi(X)(t)$  and  $Z(t) \triangleq \phi(X)(t)$ , then  $(Y(t), Z(t))$  solve the Skorokhod problem given by (6.1).*

**Result 6.2** (Lemma 14.3.3, Corollary 14.3.4 and Corollary 14.3.5 of [27]). *Let  $x \in \mathbb{D}^d$ . For the discontinuity points of  $\psi(x)$  (denoted by  $\text{Disc}(\psi(x))$ ) and  $\phi(x)$ , we have that  $\text{Disc}(\psi(x)) \cup \text{Disc}(\phi(x)) = \text{Disc}(x)$ . Moreover, if  $x$  has only positive jumps, then  $\psi(x)$  is continuous and  $\phi(x)(t) - \phi(x)(t^-) = x(t) - x(t^-)$ .*

**Result 6.3** (Theorem 14.2.2 of [27]). *The regulator map  $y = \psi(x)$  can be characterized as the unique fixed point of the map  $\pi_{x,Q} : \mathbb{D}_+^d \rightarrow \mathbb{D}_+^d$ , which is defined by*

$$\pi_{x,Q}(w)(t) = \max \left\{ 0, \sup_{s \in [0,t]} Q^T w(s) - x(s) \right\}.$$

**Result 6.4** (Consequence of Theorem 4.1 of [22]). *Let  $\Delta \in \mathbb{D}^d$  be a non-decreasing function such that  $\Delta(0) \geq 0$ . Then, for  $x \in \mathbb{D}^d$ , we have that*

$$\psi(x) \geq \psi(x + \Delta), \quad \phi(x) \leq \phi(x + \Delta),$$

and

$$\phi(x)(t_2) - \phi(x)(t_1) \leq \phi(x + \Delta)(t_1) - \phi(x + \Delta)(t_2),$$

for any  $0 \leq t_1 \leq t_2 \leq 1$ .

Finally, we assume that the right tail of  $W^{(i)}(1)$  is regularly varying with index  $-\beta_i$  and that the stability condition holds, i.e.  $R^{-1}\rho < r$ , where  $\rho \triangleq \mathbb{E}J(1)$ . Let  $\bar{Z}_n(t) \triangleq Z(nt)/n$  and  $\bar{X}_n(t) \triangleq X(nt)/n$ . Let  $c \in \{0, 1\}^d$  be a binary vector, and let  $\mathcal{J}_c$  denote the index set encoded by  $c$ , i.e.,  $j \in \mathcal{J}_c$  if  $c_j = 1$ . Set  $\bar{Z}_n(t) \triangleq Z(nt)/n$  and  $\bar{X}_n(t) \triangleq X(nt)/n$ . Define  $l_c : \mathbb{R}^d \rightarrow \mathbb{R}$  by  $l_c(x) = c^T x$  and  $\pi_1 : \mathbb{D}^d \rightarrow \mathbb{R}^d$  by  $\pi_1(\xi) = \xi(1)$ . Moreover, let  $F \triangleq l_c \circ \pi_1 \circ \phi$ . We are interested in estimating the probability of  $\mathbb{P}(c^T \bar{Z}_n(1) \geq a)$ . By Theorem 14.2.6 (iii) of [27], we have that  $\bar{Z}_n = \phi(\bar{X}_n)$ , and hence it holds that

$$\mathbb{P}(c^T \bar{Z}_n(1) \geq a) = \mathbb{P}(F(\bar{X}_n) \geq a) = \mathbb{P}(\bar{X}_n \in A), \quad (6.2)$$

where  $a > 0$  and  $A \triangleq \{\xi \in \mathbb{D} : F(\xi) \geq a\}$ .

## 6.2 Large deviations results

To obtain the large deviations asymptotics for the rare-event probability as in (6.2), we proceed the following.

- To determine the tail index of the rare-event probability, we study first the optimization problem given by (2.1) and transform it into a (nonstandard) knapsack problem with nonlinear constraints (see (6.9) and Proposition 6.1 below).
- Under a certain assumption (see Assumption 4 below), we show that  $A$ , as defined in (6.2), is bounded away from  $\mathbb{D}_{<(l_1^*, \dots, l_d^*)}$ , where  $l_1^*, \dots, l_d^*$  is the optimal solution to the knapsack problem derived in the first step.
- Finally, we derive a large deviations result for  $\mathbb{P}(\bar{X}_n \in A)$  by applying Result 2.2.

We start with the optimization problem given by (2.1). Due to the fact that  $X(t)$  is in general not a compensated compound Poisson process but one with certain drift, it is convenient to consider a slightly different problem, which is given by

$$\arg \min_{\substack{(l_1, \dots, l_d) \in \mathbb{Z}_+^d \\ \prod_{i=1}^d \mathbb{L}_{l_i}(\mu_i) \cap A \neq \emptyset}} \mathcal{I}(l_1, \dots, l_d), \quad (6.3)$$

where  $\mu \triangleq \mathbb{E}X(1) = \rho - Rr$ ,  $r' = r - R^{-1}\rho > 0$  due to the stability condition, and  $\mathbb{L}_{l_i}(\mu_i) \triangleq \{\xi \mid \exists \xi' \in \mathbb{D}_{l_i} : \xi(t) = \xi'(t) + \mu_i t = \xi'(t) - (Rr')_i t\}$ . Define  $E_0 \triangleq \{(l_1, \dots, l_d) \in \mathbb{Z}_+^d \mid l_i = 0, \forall i \in \mathcal{J}_c\}$  and  $E_1 \triangleq \{e_i \mid i \in \mathcal{J}_c\}$ , where  $e_i$  denotes the unit vector with entries 0 except for the  $i$ -th coordinate. By Result 6.2, instead of (6.3) we can solve two separate problems that are given by

$$\arg \min_{\substack{(l_1, \dots, l_d) \in E_0 \\ \prod_{i=1}^d \mathbb{L}_{l_i}(\mu_i) \cap A \neq \emptyset}} \mathcal{I}(l_1, \dots, l_d), \quad (6.4)$$

and

$$\arg \min_{\substack{(l_1, \dots, l_d) \in E_1 \\ \prod_{i=1}^d \mathbb{L}_{l_i}(\mu_i) \cap A \neq \emptyset}} \mathcal{I}(l_1, \dots, l_d). \quad (6.5)$$

Note that the optimization problem given by (6.5) can be solved easily by considering  $\min_{i \in \mathcal{J}_c} \beta_i - 1$ , therefore we focus on the optimization problem given by (6.4). Let  $\mathcal{J}$  be a subset of  $(\mathcal{J}_c)^c$ . Moreover, let  $\theta \in \mathbb{D}_1$  and let  $\xi \in \mathbb{D}^d$  be such that

$$\xi^{(i)}(t) = \begin{cases} -(Rr')_i t, & t \in [0, 1], & \text{for } i \notin \mathcal{J}, \\ \theta^{(i)} - (Rr')_i t, & t \in [0, 1], & \text{for } i \in \mathcal{J}. \end{cases} \quad (6.6)$$

A necessary and sufficient condition for the existence of  $\xi \in A$  is given in the following Proposition.

**Proposition 6.1.** Let  $\mathcal{J} \subseteq (\mathcal{J}_c)^c$ . Moreover, let  $\{r_i^*\}_{i \notin \mathcal{J}}$  be such that

$$r_i^* = \max \left\{ r'_i - \sum_{j \neq i} Q_{ji} r'_j + \sum_{\substack{j \neq i \\ j \notin \mathcal{J}}} Q_{ji} r_j^*, 0 \right\}, \quad \text{for } i \notin \mathcal{J}. \quad (6.7)$$

Define

$$\partial_z(\mathcal{J}) \triangleq \sum_{i \in \mathcal{J}_c} \left( r_i^* - r'_i + \sum_{j \neq i} Q_{ji} r'_j - \sum_{\substack{j \neq i \\ j \notin \mathcal{J}}} Q_{ji} r_j^* \right). \quad (6.8)$$

If  $\partial_z(\mathcal{J}) \neq a$ , then there exists  $\xi$  satisfying (6.6) and  $c^T \phi(\xi)(1) \geq a$ , if and only if  $\partial_z(\mathcal{J}) > a$ . Additionally, if  $\mathcal{J}_1 \subseteq \mathcal{J}_2 \subseteq (\mathcal{J}_c)^c$ , then we have that  $\partial_z(\mathcal{J}_1) \leq \partial_z(\mathcal{J}_2)$ .

*Proof.* We give here a sketch of the proof, where a detailed one can be found in Section 7. Note that  $\partial_z(\mathcal{J})$  given by (6.8) is the increasing rate of the subset  $\mathcal{I}_c$  of the workload process, whose associated input process does not have any jumps but starts with sufficiently large initial value. Based on this observation, a  $\xi$  can be constructed for the “if”-part of the first statement. For the “only if”-part, suppose that there exists a  $\xi$  satisfying  $c^T \phi(\xi)(1) \geq a$ . By Result 6.4, enlarging the size of jumps in  $\xi$  will preserve the fact that  $c^T \phi(\xi)(1) \geq a$ . Hence, we can construct a new  $\xi$ , such that

- the associated workload process  $\phi(\xi)$  is piecewise linear between two neighboring discontinuity points; and
- the increasing rate of  $c^T \phi(\xi)$  is always smaller or equal than  $\partial_z(\mathcal{J})$  given by (6.8).

□

*Remark 6.* Note that (6.7) can be written in a matrix notation that is given by

$$r^* = \max \left\{ ((I - Q^T)r')_{\notin \mathcal{J}} + (Q_{\notin \mathcal{J}})^T r^*, 0 \right\} = \max \left\{ (Rr - \rho)_{\notin \mathcal{J}} + (Q_{\notin \mathcal{J}})^T r^*, 0 \right\},$$

where  $(Rr - \rho)_{\notin \mathcal{J}}$  and  $Q_{\notin \mathcal{J}}$  denote the vector and matrix respectively with its  $i$ -th row and column being removed for all  $i \in \mathcal{J}$ . Using the Banach fixed-point theorem, we obtain that  $r^* = \lim_{n \rightarrow \infty} \bar{\pi}^n(0)$ , where  $\bar{\pi}^n \triangleq \bar{\pi} \circ \bar{\pi}^{n-1}$  and  $\bar{\pi}(x) \triangleq \max \left\{ (Rr - \rho)_{\notin \mathcal{J}} + (Q_{\notin \mathcal{J}})^T x, 0 \right\}$ .

Define  $E'_0 \triangleq E_0 \cap \{(l_1, \dots, l_d) \in \mathbb{Z}_+^d \mid l_i \in \{0, 1\}, \forall i \notin \mathcal{J}_c\}$  and

$$E_{\mathcal{J}} \triangleq \{(l_1, \dots, l_d) \in E'_0 \mid \partial_z(\mathcal{J}_{(l_1, \dots, l_d)}) > a\},$$

where  $\partial_z(\mathcal{J}_{(l_1, \dots, l_d)})$  is as defined in (6.8) with  $\mathcal{J}_{(l_1, \dots, l_d)}$  denoting the index set encoded by  $(l_1, \dots, l_d) \in E'_0$ . By Proposition 6.1, we conclude that the optimization problem formulated in (6.4) is equivalent to

$$\arg \min_{(l_1, \dots, l_d) \in E_{\mathcal{J}}} \mathcal{I}(l_1, \dots, l_d). \quad (6.9)$$

Thanks to the last statement of Proposition 6.1, it is unnecessary to check every  $(l_1, \dots, l_d) \in E_{\mathcal{J}}$  for solving (6.9). Although, the optimization problem formulated in (6.9) is a nonstandard knapsack problem with nonlinear constraints. In the following example, we consider a specific fluid network and illustrate how to solve the optimization problem given by (6.9) using Proposition 6.1.

**Example 1.** Consider the fluid network given by  $\rho = (0.8 \ 0.8 \ 1)^T$ ,  $r = (1 \ 1 \ 2.5)^T$  and

$$Q = \begin{bmatrix} 0 & 0.1 & 0.8 \\ 0.1 & 0 & 0.8 \\ 0 & 0 & 0 \end{bmatrix}.$$

We are interested in the probability of the rare event that the third station crosses the level  $na$  at time  $n$  for large  $n$ , i.e.  $\mathcal{J}_c = \{3\}$ . It is easy to check that the stability condition holds. By an easy computation, we obtain that  $\partial_z(\{1, 2\}) = 0.1$  and  $\partial_z(\{1\}) = \partial_z(\{2\}) = 0.02$ . For  $a = 0.05$ , the optimal solution to (6.9) is given by  $(1, 1, 0)$ .

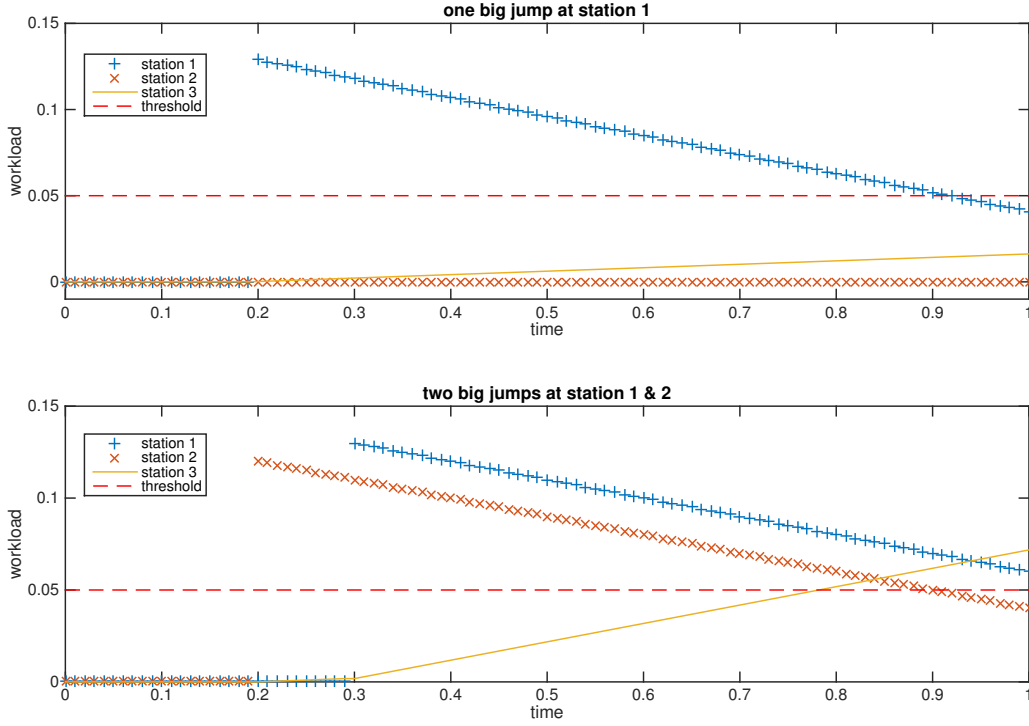


Figure 2: An illustration of two different sample paths of workload processes (under the setting of Example 1), whose associated input processes have the form as in (6.6).

Suppose that we have solved the optimization problem given by (6.9). To obtain the large deviations results, the following technical assumption needs to be made.

**Assumption 4.** Assume that the optimization problem formulated in (6.9) satisfies the conditions as follows.

- a) The optimization problem given by (6.9) has a unique solution.
- b) For every  $\mathcal{J} \subseteq (\mathcal{J}_c)^c$ , it holds that  $\partial_z(\mathcal{J}) \neq a$ .
- c) Let  $(l_1^*, \dots, l_d^*)$  denote the optimal solution to (6.9). We assume that

$$\mathcal{I}(l_1^*, \dots, l_d^*) < \min_{i \in \mathcal{J}_c} \beta_i - 1.$$

By Result 2.2, Assumption 4 c) implies that the rare event is caused by multiple large jumps. Throughout the rest of this section, we assume that Assumption 4 holds. We end this subsection with a large deviations result for  $\mathbb{P}(\bar{X}_n \in A) = \mathbb{P}(c^T \bar{Z}_n(1) \geq a)$ , which is formulated in the following proposition.

**Proposition 6.2.** *Suppose that Assumption 4 holds. Let  $F$  be as defined in (6.2). Then  $A = F^{-1}[a, \infty)$  is bounded away from*

$$\bigcup_{(l_1, \dots, l_d) \in I_{<(l_1^*, \dots, l_d^*)}} \prod_{i=1}^d \mathbb{L}_{l_i}(\mu_i),$$



where  $(l_1^*, \dots, l_d^*)$  denotes the unique optimal solution of (6.9). Moreover, we have that

$$\begin{aligned} C_{l_1^*} \times \dots \times C_{l_d^*} ((F^{-1}[a, \infty))^{\circ}) &\leq \liminf_{n \rightarrow \infty} \frac{\mathbb{P}(\bar{X}_n \in A)}{\prod_{i=1}^d (n\nu_i[n, \infty))^{l_i^*}} \\ &\leq \limsup_{n \rightarrow \infty} \frac{\mathbb{P}(\bar{X}_n \in A)}{\prod_{i=1}^d (n\nu_i[n, \infty))^{l_i^*}} \leq C_{l_1^*} \times \dots \times C_{l_d^*} (F^{-1}[a, \infty)). \end{aligned}$$

*Proof.* See Section 7. □

### 6.3 Simulation

Again, we are in the setting of Theorem 3.6. To be able to discuss the choice of  $J_{(l_1^*, \dots, l_d^*)}$  and the parameter  $\gamma$  in a more precise context, let us consider the stochastic fluid network introduced in Example 1.

**Example 1** (continued). Recall that, for  $a = 0.05$ , the optimal solution of (6.3) is given by  $\beta_1 + \beta_2 - 2$ , if we assume that  $\beta_1 + \beta_2 - 2 < \beta_3 - 1$ . Moreover, it can be easily shown that  $A$  is bounded away from both  $\mathbb{D}_{<i} \times \mathbb{D}_0 \times \mathbb{D}_0$  and  $\mathbb{D}_0 \times \mathbb{D}_{<j} \times \mathbb{D}_0$ . Combining this with  $\mathcal{I}(1, 1, 1) > 2\mathcal{I}(1, 1, 0)$ , as well as Corollary 3.8, it is sufficient to take  $\tilde{J}_{(l_1^*, \dots, l_d^*)} = \{(1, 1, 0), (0, 0, 1)\}$ . This implies that

$$B_n^\gamma = \left\{ \#\{k \mid W^{(i)}(k) > n\gamma_i, k \leq N^{(i)}(n)\} \geq 1, \forall i \in \{1, 2\} \right\} \cup \left\{ \#\{k \mid W^{(3)}(k) > n\gamma_3, k \leq N^{(3)}(n)\} \geq 1 \right\},$$

and hence

$$(B_n^\gamma)^c = \left\{ \exists i \in \{1, 2\} : W^{(i)}(k) \leq n\gamma_i, \forall 1 \leq k \leq N^{(i)}(n) \right\} \cap \left\{ W^{(3)}(k) \leq n\gamma_3, \forall 1 \leq k \leq N^{(3)}(n) \right\}.$$

We choose  $\gamma$  such that  $\mathbb{P}(A_n \cap (B_n^\gamma)^c) = o(\mathbb{P}(A_n)^2)$ . To begin with, we assume w.l.o.g. that  $\beta_3 - 1 \leq 2(\beta_1 + \beta_2 - 2)$ , otherwise we can simply set  $J_{(l_1^*, \dots, l_d^*)} = \{(1, 1, 0)\}$ , since  $\mathcal{I}(0, 0, 1) > 2\mathcal{I}(1, 1, 0)$ . Now the parameter  $\gamma_3$  can be chosen such that

$$\left\lceil \frac{1/20}{\gamma_3} \right\rceil (\beta_3 - 1) > 2(\beta_1 + \beta_2 - 2).$$

For the choice of  $\gamma_1$ , we observe that the job arriving at the second station can have arbitrarily large size. Hence, it is sufficient to consider the inequality  $\partial_z(\{1, 2\})t' + \partial_z(\{2\})(1 - t') > a$ , where  $\partial_z(\{1, 2\}) = 0.1$  and  $\partial_z(\{2\}) = 0.02$ . Solving the inequality we obtain that  $t' < 3/8$ . This simply means that the workload process of the third station cannot exceed the level  $a$  at time 1 if we keep both of the first and the second stations overloaded less than  $3/8$  of the time. Since the workload process of the first station decays at rate  $1/10$ , one can choose  $\gamma_1$  such that

$$\left\lceil \frac{3/80}{\gamma_1} \right\rceil (\beta_1 - 1) + (\beta_2 - 1) > 2(\beta_1 + \beta_2 - 2).$$

Analogously, it is sufficient to set  $\gamma_2$  such that

$$(\beta_1 - 1) + \left\lceil \frac{3/80}{\gamma_2} \right\rceil (\beta_2 - 1) > 2(\beta_1 + \beta_2 - 2).$$

We give a closed-form expression for  $\mathbb{P}(B_n^\gamma)$ . By assumption  $\{W^{(i)}(k)\}_{1 \leq i \leq d}$  are mutually independent, therefore we have that

$$\begin{aligned} \mathbb{P}((B_n^\gamma)^c) &= \mathbb{P}\left(\exists i \in \{1, 2\} : W^{(i)}(k) \leq n\gamma_i, \forall 1 \leq k \leq N^{(i)}(n)\right) \mathbb{P}\left(W^{(3)}(k) \leq n\gamma_3, \forall 1 \leq k \leq N^{(3)}(n)\right) \\ &= \left[1 - \prod_{i=1}^2 \left(1 - \mathbb{P}\left(W^{(i)}(k) \leq n\gamma_i, \forall 1 \leq k \leq N^{(i)}(n)\right)\right)\right] \mathbb{P}\left(W^{(3)}(k) \leq n\gamma_3, \forall 1 \leq k \leq N^{(3)}(n)\right). \end{aligned}$$

Est PR	$n = 1200$	$n = 1600$	$n = 2000$	$n = 2400$
$\beta_1 = 1.5, \beta_2 = 1.5, \beta_3 = 2.2$	$7.719 \times 10^{-2}$ 0.045	$6.228 \times 10^{-2}$ 0.058	$4.541 \times 10^{-2}$ 0.057	$3.973 \times 10^{-2}$ 0.057
Est PR	$n = 800$	$n = 1200$	$n = 1600$	$n = 2000$
$\beta_1 = 2.5, \beta_2 = 2.3, \beta_3 = 4$	$2.894 \times 10^{-2}$ 0.325	$1.686 \times 10^{-2}$ 0.404	$6.153 \times 10^{-3}$ 0.445	$2.023 \times 10^{-3}$ 0.448
Est PR	$n = 600$	$n = 1000$	$n = 1400$	$n = 1800$
$\beta_1 = 2.2, \beta_2 = 2.9, \beta_3 = 4.5$	$5.139 \times 10^{-2}$ 0.249	$1.858 \times 10^{-2}$ 0.347	$9.987 \times 10^{-3}$ 0.351	$1.028 \times 10^{-3}$ 0.377

Table 3: Estimated rare-event probability and level of precision for the application as described in Section 6.3 w.r.t. different combinations of  $n$  and  $\beta_1, \beta_2, \beta_3$ .

Conditional on  $N^{(i)}(n)$ , we obtain that

$$\begin{aligned} \mathbb{P}\left(W^{(i)}(k) \leq n\gamma_i, \forall 1 \leq k \leq N^{(i)}(n)\right) &= e^{-n} \sum_{m=0}^{\infty} \frac{n^m}{m!} \mathbb{P}\left(W^{(i)}(1) \leq n\gamma_i\right)^m \\ &= \exp\left\{-n\left(1 - \mathbb{P}\left(W^{(i)}(1) \leq n\gamma_i\right)\right)\right\}. \end{aligned}$$

Summarizing the findings from above, we are able to propose a strongly efficient estimator for  $\mathbb{P}(A_n)$  that is given by

$$Z_n = \frac{\mathbb{1}_{A_n}}{w + \frac{1-w}{\mathbb{P}(B_n^\gamma)} \mathbb{1}_{B_n^\gamma}}.$$

Moreover, Algorithm 3 can be used to sample from  $\mathbb{Q}^\gamma$ . To see this, we decompose  $B_n^\gamma$  into two disjoint sets  $B_n^\gamma(1)$  and  $B_n^\gamma(2)$  that are given by

$$B_n^\gamma(1) \triangleq \left\{ \#\{k \mid W^{(3)}(k) > n\gamma_3, k \leq N^{(3)}(n)\} \geq 1 \right\},$$

and

$$B_n^\gamma(2) \triangleq \left\{ \#\{W^{(i)}(k) > n\gamma_i, k \leq N^{(i)}(n)\} \geq 1, \forall i \in \{1, 2\} \right\} \cap \{W^{(3)}(k) \leq n\gamma_3, \forall 1 \leq k \leq N^{(3)}(n)\},$$

respectively. Using Algorithm 3, the sample path of  $\bar{X}_n^{(1)}, \bar{X}_n^{(2)}, \bar{X}_n^{(3)}$  can be simulated independently on both  $B_n^\gamma(1)$  and  $B_n^\gamma(2)$ . We present the numerical results based on 20000 samples in Table 3. We choose  $W^{(i)}(1)$  such that  $\mathbb{P}(W^{(i)}(1) > t) = (t_{r,i}/t)^{\beta_i}$  and  $t_{r,i} = \rho_i(\beta_i - 1)/\beta_i$ , for  $i \in \{1, 2, 3\}$ . As one can see, the numerical results suggest again what our theory predicts.

## 7 Proofs

In this section we provide proofs of the results presented in this paper.

*Proof of Proposition 3.1.* Recall that the expected running time of the rejection sampling (see Algorithm 1 above), which is used to generate the jumps of  $\bar{X}_n$ , is bounded from above by

$$M_l = \frac{\binom{l}{l_*} \mathbb{P}(W(1) > n\gamma)^{l_*}}{\mathbb{P}(B_n^\gamma | N(n) = l)}.$$

Hence, for the expected running time of Algorithm 1, denoted by  $T_{\text{alg1}}(n)$ , we have that

$$\begin{aligned}
T_{\text{alg1}}(n) &= \sum_{l \geq l^*} h_l M_l = \mathbb{P}(B_n^\gamma)^{-1} \sum_{l \geq l^*} \mathbb{P}(B_n^\gamma | N(n) = l) \mathbb{P}(N(n) = l) M_l \\
&= \mathbb{P}(B_n^\gamma)^{-1} \sum_{l \geq l^*} \mathbb{P}(N(n) = l) \binom{l}{l^*} \mathbb{P}(W(1) > n\gamma)^{l^*} \\
&= C(l^*) \frac{n^{l^*} (\lambda \mathbb{P}(W(1) > n\gamma))^{l^*}}{\mathbb{P}(B_n^\gamma)} e^{-\lambda n} \sum_{l \geq l^*} \frac{(\lambda n)^{l-l^*}}{(l-l^*)!} \\
&= C(l^*) \frac{n^{l^*} (\lambda \mathbb{P}(W(1) > n\gamma))^{l^*}}{\mathbb{P}(B_n^\gamma)}.
\end{aligned}$$

Using Result 2.1, we obtain that  $\limsup_{n \rightarrow \infty} T_{\text{alg1}}(n) < \infty$ .  $\square$

*Proof of Proposition 3.3.* Let  $B^\gamma$  be as defined in (3.4). For  $I \subseteq J_{(l_1^*, \dots, l_d^*)}$ , define

$$B_I^{\gamma;l} \triangleq \bigcap_{(l_1, \dots, l_d) \in I} B^{\gamma;l}.$$

By the inclusion-exclusion principle, we have that

$$\mathbb{P}(\bar{X}_n \in B^\gamma) = \sum_{k=1}^{|J_{(l_1^*, \dots, l_d^*)}|} (-1)^{k-1} \sum_{\substack{|I|=k \\ I \subseteq J_{(l_1^*, \dots, l_d^*)}}} \mathbb{P}(\bar{X}_n \in B_I^{\gamma;l}). \quad (7.1)$$

Moreover, for any finite collection  $I$  of elements in  $\mathbb{Z}_+^d$  with  $I \subseteq J_{(l_1^*, \dots, l_d^*)}$ , we have that

$$\begin{aligned}
B_I^{\gamma;l} &= \bigcap_{i=1}^d \bigcap_{(l_1, \dots, l_d) \in I} \left\{ \left( \xi^{(1)}, \dots, \xi^{(d)} \right) \mid \#\{t \mid \xi^{(i)}(t) - \xi^{(i)}(t^-) > \gamma_i\} \geq l_i \right\} \\
&= \bigcap_{i=1}^d \left\{ \left( \xi^{(1)}, \dots, \xi^{(d)} \right) \mid \#\{t \mid \xi^{(i)}(t) - \xi^{(i)}(t^-) > \gamma_i\} \geq \hat{l}_{i;I} \right\},
\end{aligned} \quad (7.2)$$

where  $\hat{l}_{i;I} \triangleq \max_{(l_1, \dots, l_d) \in I} l_i$ . Since  $\bar{X}_n^{(1)}, \dots, \bar{X}_n^{(d)}$  are independent processes, we obtain that

$$\mathbb{P}(B_I^{\gamma;l}) = \prod_{i=1}^d \left( 1 - \exp \left\{ -\lambda_i n \mathbb{P}(W^{(i)}(1) > n\gamma_i) \right\} \sum_{j=0}^{\hat{l}_{i;I}-1} \frac{(\lambda_i n)^j}{j!} \mathbb{P}(W^{(i)}(1) > n\gamma_i)^j \right).$$

$\square$

*Proof of Lemma 3.4.* Recall that

$$B^{\gamma;l}(i, j) \triangleq \left\{ \xi \in \mathbb{D}^d \mid \#\{t \mid \xi^{(i)}(t) - \xi^{(i)}(t^-) > \gamma_i\} \geq (l(j))_i \right\}.$$

Hence, we have that

$$B^\gamma = \bigcup_{j=1}^{|J_{(l_1^*, \dots, l_d^*)}|} \bigcap_{i=1}^d B^{\gamma;l}(i, j) = \bigcup_{j=1}^{|J_{(l_1^*, \dots, l_d^*)}|} \left( B^{\gamma;l}(1, j) \cap \bigcap_{i=2}^d B^{\gamma;l}(i, j) \right). \quad (7.3)$$

By definition

$$\Delta B^{\gamma;l}(i, j) \triangleq B^{\gamma;l}(i, j) \setminus \left( \bigcup_{m=1}^{j-1} B^{\gamma;l}(i, m) \right).$$

Therefore, we have that

$$B^{\gamma;l}(i, j) = \bigcup_{m_i=1}^j \Delta B^{\gamma;l}(i, j). \quad (7.4)$$

Plugging (7.4) into (7.3), we obtain that

$$\begin{aligned} B^\gamma &= \bigcup_{j=1}^{|J(t_1^*, \dots, t_d^*)|} \left( \left( \bigcup_{m_1=1}^j \Delta B^{\gamma;l}(1, m_1) \right) \cap \bigcap_{i=2}^d B^{\gamma;l}(i, j) \right) \\ &= \bigcup_{m_1=1}^{|J(t_1^*, \dots, t_d^*)|} \left( \bigcup_{j=m_1}^{|J(t_1^*, \dots, t_d^*)|} \left( \Delta B^{\gamma;l}(1, m_1) \cap \bigcap_{i=2}^d B^{\gamma;l}(i, j) \right) \right) \\ &= \bigcup_{m_1=1}^{|J(t_1^*, \dots, t_d^*)|} \left( \Delta B^{\gamma;l}(1, m_1) \cap \left( \bigcup_{j=1}^{m_1} \bigcap_{i=2}^d B^{\gamma;l}(i, j) \right) \right). \end{aligned}$$

Applying the same procedure to  $\bigcup_{j=1}^{m_1} \bigcap_{i=2}^d B^{\gamma;l}(i, j)$ , we obtain that

$$B^\gamma = \bigcup_{m_1=1}^{|J(t_1^*, \dots, t_d^*)|} \bigcup_{m_2=1}^{m_1} \left( \Delta B^{\gamma;l}(1, m_1) \cap \Delta B^{\gamma;l}(2, m_2) \cap \left( \bigcup_{j=1}^{m_2} \bigcap_{i=3}^d B^{\gamma;l}(i, j) \right) \right).$$

Iterating the same procedure  $d - 1$  times, we obtain that

$$B^\gamma = \bigcup_{m_1=1}^{|J(t_1^*, \dots, t_d^*)|} \bigcup_{m_2=1}^{m_1} \dots \bigcup_{m_{d-1}=1}^{m_{d-2}} \left( \left( \bigcap_{i=1}^{d-1} \Delta B^{\gamma;l}(i, m_i) \right) \cap \left( \bigcup_{j=1}^{m_{d-1}} B^{\gamma;l}(d, j) \right) \right). \quad (7.5)$$

Since  $l(1), \dots, l(|J(t_1^*, \dots, t_d^*)|)$  are ordered such that  $(l(1))_d \leq (l(2))_d \leq \dots \leq (l(|J(t_1^*, \dots, t_d^*)|))_d$ , we obtain that

$$\bigcup_{j=1}^{m_{d-1}} B^{\gamma;l}(d, j) = B^{\gamma;l}(d, 1). \quad (7.6)$$

Plugging (7.6) into (7.5), we obtain that

$$B^\gamma = \bigcup_{m_1=1}^{|J(t_1^*, \dots, t_d^*)|} \bigcup_{m_2=1}^{m_1} \dots \bigcup_{m_{d-1}=1}^{m_{d-2}} \left( \left( \bigcap_{i=1}^{d-1} \Delta B^{\gamma;l}(i, m_i) \right) \cap B^{\gamma;l}(d, 1) \right).$$

□

*Proof of Theorem 3.6.* For the second moment of  $Z$  (under the change of measure) we have that

$$\begin{aligned} \mathbb{E}^{\mathbb{Q}_{\gamma, w}} [Z_n^2] &= \mathbb{E}[Z_n] \\ &= \mathbb{E}[Z_n \mathbb{1}_{B_n^\gamma}] + \mathbb{E}[Z_n \mathbb{1}_{(B_n^\gamma)^c}] \\ &\leq \frac{1}{1-w} \mathbb{P}(A_n \cap B_n^\gamma) \mathbb{P}(B_n^\gamma) + \frac{1}{w} \mathbb{P}(A_n \cap (B_n^\gamma)^c) \\ &\leq \frac{1}{1-w} \mathbb{P}(A_n) \mathbb{P}(B_n^\gamma) + \frac{1}{w} \mathbb{P}(A_n \cap (B_n^\gamma)^c). \end{aligned}$$

Combining this with Lemma 3.7 we obtain the strong efficiency of our estimator.

□

*Proof of Lemma 3.7.* First, note that  $\mathbb{P}(\bar{X}_n \in B^\gamma) = \mathcal{O}(\mathbb{P}(\bar{X}_n \in A))$  follows immediately from Result 2.2.

We need to show the existence of  $\gamma$  such that  $\mathbb{P}(\bar{X}_n \in A \cap (B^\gamma)^c) = o(\mathbb{P}(\bar{X}_n \in A)^2)$ . Since  $A$  is bounded away from  $\mathbb{D}_{<(l_1^*, \dots, l_d^*)}$  by assumption, there exists  $r$  such that  $d\left(A, \mathbb{D}_{<(l_1^*, \dots, l_d^*)}\right) \geq r$ . On the one hand, from [23] we have that

$$A \subseteq \left\{ \left( \xi^{(1)}, \dots, \xi^{(d)} \right) \mid \exists (l_1, \dots, l_d) \in J_{(l_1^*, \dots, l_d^*)} : d\left(\xi^{(i)}, \mathbb{D}_{<l_i}\right) \geq r, \forall i \in \{1, \dots, d\} \right\}. \quad (7.7)$$

On the other hand, we have that

$$(B^\gamma)^c = \left\{ \left( \xi^{(1)}, \dots, \xi^{(d)} \right) \mid \forall (l_1, \dots, l_d) \in J_{(l_1^*, \dots, l_d^*)} : \right. \\ \left. \exists i \in \{1, \dots, d\} : \#\{t \mid \xi^{(i)}(t) - \xi^{(i)}(t^-) > \gamma_i\} \leq l_i - 1 \right\}. \quad (7.8)$$

Let  $\xi = (\xi^{(1)}, \dots, \xi^{(d)}) \in A \cap (B^\gamma)^c$  be a step function in the set  $\prod_{i=1}^d \mathbb{D}_{l_i}$ . By (7.7), there exists  $(l_1, \dots, l_d) \in J_{(l_1^*, \dots, l_d^*)}$ , such that  $\xi^{(i)} = \sum_{j=1}^{l_i+m_i} c_j^{(i)} \mathbb{1}_{[t_j^{(i)}, 1]}$ ,  $m_i \in \mathbb{Z}_+$  and  $d(\xi^{(i)}, \mathbb{D}_{<l_i}) \geq r$  for all  $i \in \{1, \dots, d\}$  with  $l_i \neq 0$ . Combining  $d(\xi^{(i)}, \mathbb{D}_{<l_i}) \geq r$  with the fact that  $\xi^{(i)} = \sum_{j=1}^{l_i-1} c_j^{(i)} \mathbb{1}_{[t_j^{(i)}, 1]} \in \mathbb{D}_{<l_i}$ , we conclude that

$$\sum_{j=l_i}^{l_i+m_i} c_j^{(i)} \geq d\left(\sum_{j=1}^{l_i+m_i} c_j^{(i)} \mathbb{1}_{[t_j^{(i)}, 1]}, \sum_{j=1}^{l_i-1} c_j^{(i)} \mathbb{1}_{[t_j^{(i)}, 1]}\right) \geq r, \quad (7.9)$$

or in other words, the sum of  $m_i + 1$  smallest jump is bounded from below by  $r$  for each  $\xi^{(i)}$  of  $\{\xi^{(i)}\}_{i \in \{1, \dots, d\}}$  satisfying  $l_i \neq 0$ . Combining (7.8) with (7.9), as well as choosing  $\gamma_k$  sufficiently small, there exists at least one  $k \in \{1, \dots, d\}$  such that the smallest jump of  $\xi^{(k)}$  is bounded from below by  $r' > 0$  for arbitrary but fixed  $m_k$ . Repeating the same procedure as described above, we can construct  $(m_1, \dots, m_d)$  for every  $(l_1, \dots, l_d) \in J_{(l_1^*, \dots, l_d^*)}$ , such that the optimization problem, given by

$$\arg \min_{\substack{(l_1, \dots, l_d) \in \mathbb{Z}_+^d \\ \prod_{i=1}^d \mathbb{D}_{l_i} \cap A \cap (B^\gamma)^c \neq \emptyset}} \mathcal{I}(l_1, \dots, l_d), \quad (7.10)$$

has a unique solution  $(l_1^{**}, \dots, l_d^{**})$  satisfying  $\mathcal{I}(l_1^{**}, \dots, l_d^{**}) > 2\mathcal{I}(l_1^*, \dots, l_d^*)$ . We denote this specific choice of  $(m_1, \dots, m_d)$  for every  $(l_1, \dots, l_d) \in J_{(l_1^*, \dots, l_d^*)}$  by  $\{m^{(l_1, \dots, l_d)}\}_{(l_1, \dots, l_d) \in J_{(l_1^*, \dots, l_d^*)}}$ . It should be noted that the existence and the uniqueness of  $(l_1^{**}, \dots, l_d^{**})$  can be guaranteed by enlarging the set  $A$  (since we are looking for an upper bound for  $\mathbb{P}(\bar{X}_n \in A \cap (B^\gamma)^c)$ ), together with choosing the corresponding  $\gamma_i$  sufficiently small. Therefore, it remains to show that, under the chosen  $\gamma$ , the set  $A \cap (B^\gamma)^c$  is bounded away from  $\mathbb{D}_{<(l_1^{**}, \dots, l_d^{**})}$ . Select  $\xi$  satisfying  $d(\xi, \mathbb{D}_{<(l_1^{**}, \dots, l_d^{**})}) < \delta$ , and hence, there exists  $\theta \in \mathbb{D}_{<(l_1^{**}, \dots, l_d^{**})}$  such that  $d(\xi, \theta) < \delta$ . On the one hand, combining  $d(\xi, \theta) < \delta$  with (7.7), there exists  $(l_1, \dots, l_d) \in J_{(l_1^*, \dots, l_d^*)}$  such that  $d(\theta^{(i)}, \mathbb{D}_{<l_i}) > r - \delta$ , for all  $i \in \{1, \dots, d\}$ . Hence, we have that  $\theta^{(i)} = \sum_{j=1}^{l_i+m_i} c_j^{(i)} \mathbb{1}_{[t_j^{(i)}, 1]}$ ,  $m_i \in \mathbb{Z}_+$ , satisfying

$$\sum_{j=l_i}^{l_i+m_i} c_j^{(i)} \geq d\left(\sum_{j=1}^{l_i+m_i} c_j^{(i)} \mathbb{1}_{[t_j^{(i)}, 1]}, \sum_{j=1}^{l_i-1} c_j^{(i)} \mathbb{1}_{[t_j^{(i)}, 1]}\right) \geq r - \delta, \quad (7.11)$$

for all  $i \in \{1, \dots, d\}$  with  $l_i \neq 0$ . On the other hand, there exist homeomorphisms  $\{\lambda_i\}_{i \in \{1, \dots, d\}}$  such that

$$\|\lambda_i - id\|_\infty \vee \|\theta^{(i)} \circ \lambda_i - \xi^{(i)}\|_\infty < \delta, \quad (7.12)$$

for  $i \in \{1, \dots, d\}$ . Combining (7.12) with (7.8), we conclude the existence of at least one  $i \in \{1, \dots, d\}$  such that

$$\#\{t \mid \xi^{(i)}(t) - \xi^{(i)}(t^-) > \gamma_i - \delta\} \leq l_i - 1. \quad (7.13)$$

Since  $\theta \in \mathbb{D}_{<(l_1^{**}, \dots, l_d^{**})}$ , we have that

$$m_i \leq (m^{(l_1, \dots, l_d)})_i - 1, \quad (7.14)$$

for all  $i \in \{1, \dots, d\}$  with  $l_i \neq 0$ . Finally, by (7.11), (7.13) and the choice of  $\gamma$ , we conclude that choosing  $\delta$  sufficiently small leads to contradiction of (7.14).  $\square$

*Proof of Proposition 6.1.* We derive a necessary and sufficient condition for  $c^T \phi(\xi)(1) \geq a$  with  $\xi$  (6.6).

For the ‘‘only if’’-part, suppose that  $\partial_z(\mathcal{J}) > a$ . Let  $(v_1, \dots, v_d) \in \mathbb{R}_+^d$ ,  $\delta \in (0, 1)$  and  $\xi$  be such that

$$\xi^{(i)}(t) = \begin{cases} -(Rr')_i t, & t \in [0, 1], & \text{for } i \notin \mathcal{J}, \\ v_i \mathbb{1}_{[\delta, 1]}(t) - (Rr')_i t, & t \in [0, 1], & \text{for } i \in \mathcal{J}. \end{cases}$$

Obviously  $\xi$  satisfies (6.6). For  $t \in [0, \delta)$ , by Result 6.3, the regulator process  $y_\xi \triangleq \psi(\xi)$  should satisfy the fixed point equation that is given by

$$y_\xi^{(i)}(t) = \max \left\{ 0, \sup_{s \in [0, t]} \sum_{j \neq i} Q_{ji} y_\xi^{(j)}(s) + (Rr')_i s \right\}, \text{ for all } i \in \{1, \dots, d\}.$$

Using the fact that  $r' > 0$ , we obtain that  $y_\xi(t) = r't$ , for  $t \in [0, \delta)$ . For  $t \in [\delta, 1]$ , again by Result 6.3, it holds that

$$y_\xi^{(i)}(t) = \max \left\{ 0, -v_i + \sup_{s \in [0, t]} r'_i s + \sum_{j \neq i} Q_{ji} (y_\xi^{(j)}(s) - r'_j s) \right\}, \text{ for all } i \in \mathcal{J}, \quad (7.15)$$

and

$$y_\xi^{(i)}(t) = \max \left\{ 0, \sup_{s \in [0, t]} r'_i s + \sum_{j \in \mathcal{J}} Q_{ji} (y_\xi^{(j)}(s) - r'_j s) + \sum_{\substack{j \neq i \\ j \notin \mathcal{J}}} Q_{ji} (y_\xi^{(j)}(s) - r'_j s) \right\}, \text{ for all } i \notin \mathcal{J}. \quad (7.16)$$

Since  $\{v_i\}_{i \in \mathcal{J}}$  are non-negative, by Result 6.2, we conclude that  $y_\xi(s)$ , as well as  $r'_i s + \sum_{j \neq i} Q_{ji} (y_\xi^{(j)}(s) - r'_j s)$  are continuous in  $s$  on  $[0, 1]$ . Using the Bolzano-Weierstrass theorem, there exists a set of sufficiently large  $\{v_i\}_{i \in \mathcal{J}}$  (depending on  $y_\xi$ ), such that  $y_\xi^{(i)}(t) = y_\xi^{(i)}(\delta) = r'_i \delta$  for  $i \in \mathcal{J}$ . Plugging this into (7.16) along with setting  $y_\xi^{(i)}(t) = r'_i \delta + r_i^*(t - \delta)$  for  $i \notin \mathcal{J}$ ,  $t \in [\delta, 1]$ , we obtain that

$$\begin{aligned} r'_i \delta + r_i^*(t - \delta) &= \max \left\{ 0, \sup_{s \in [0, t]} r'_i s + \sum_{j \in \mathcal{J}} Q_{ji} (y_\xi^{(j)}(s) - r'_j s) + \sum_{\substack{j \neq i \\ j \notin \mathcal{J}}} Q_{ji} (y_\xi^{(j)}(s) - r'_j s) \right\} \\ &= \max \left\{ r'_i \delta, r'_i \delta + \max_{s \in [\delta, t]} r'_i (s - \delta) - \sum_{j \neq i} Q_{ji} r'_j (s - \delta) + \sum_{\substack{j \neq i \\ j \notin \mathcal{J}}} Q_{ji} r_j^* (s - \delta) \right\}, \text{ for all } i \in \mathcal{J}. \end{aligned} \quad (7.17)$$

Note that (7.17) is solved by  $r_i^*$  satisfying (6.7). Moreover, by a straightforward computation, for the workload process  $z_\xi \triangleq \phi(\xi)$ , we obtain that  $c^T z_\xi(1) = \partial_z(\mathcal{J})(1 - \delta)$ . Since by assumption  $\partial_z(\mathcal{J}) > a$ , we can choose  $\delta$  such that  $c^T z_\xi(1) \geq a$ .

For the other direction of the proof, suppose that  $c^T \phi(\xi)(1) \geq a$  for some  $\xi$  satisfying (6.6). Let the jump sizes and the associated jump times of  $\xi$  be denoted by  $\{u_i\}_{i \in \mathcal{J}}$  and  $\{t_i\}_{i \in \mathcal{J}}$ , respectively. First we should mention that, by Result 6.4, enlarging  $\{u_i\}_{i \in \mathcal{J}}$  will preserve the fact that  $c^T \phi(\xi)(1) \geq a$ . Moreover, let  $d_1 < \dots < d_m$  denote the discontinuity points of  $\xi$  with  $m \leq |\mathcal{J}|$  and define  $\mathcal{J}_i \triangleq \{k \mid t_k \leq d_i\}$ , for every  $i \in \{1, \dots, m\}$ . Now observe that  $y_\xi(t) = r't$ ,  $t \in [0, d_1)$ . Hence, we have that  $z'_\xi(t) = 0 \leq \partial_z(\mathcal{J})$ , for  $t \in [0, d_1)$ . For  $y_\xi(t)$ ,  $t \in [d_1, d_2)$ , we can easily check that

$$y_\xi^{(i)}(t) = \begin{cases} r'_i d_1, & \text{for all } i \in \mathcal{J}_1, \\ r'_i d_1 + r_i^{*,1}(t - d_1), & \text{for all } i \notin \mathcal{J}_1, \end{cases}$$

by taking sufficiently large  $\{u_i\}_{i \in \mathcal{J}_1}$ , where

$$r_i^{*,1} = \max \left\{ r'_i - \sum_{j \neq i} Q_{ji} r'_j + \sum_{\substack{j \neq i \\ j \notin \mathcal{J}_1}} Q_{ji} r_j^{*,1}, 0 \right\}, \quad \text{for } i \notin \mathcal{J}_1.$$

Since  $\mathcal{J}_1 \subseteq \mathcal{J}$ , by Result 6.4 and (6.7), we conclude that  $z'_\xi(t) \leq \partial_z(\mathcal{J})$ , for  $t \in [d_1, d_2)$ . Defining  $\mathcal{J}'_1 \triangleq \mathcal{J}_1 \cup \{k \mid r_k^{*,1} = 0\}$ , we consider  $y_\xi(t)$  for  $t \in [d_2, d_3)$ . Following a similar argument as above, we claim that

$$y_\xi^{(i)}(t) = \begin{cases} r'_i d_1, & \text{for all } i \in \mathcal{J}'_1, \\ r'_i d_1 + r_i^{*,1}(d_2 - d_1), & \text{for all } i \in \mathcal{J}_2 \setminus \mathcal{J}'_1, \\ r'_i d_1 + r_i^{*,1}(d_2 - d_1) + r_i^{*,2}(t - d_2), & \text{for all } i \notin \mathcal{J}'_1 \cup \mathcal{J}_2, \end{cases}$$

for sufficiently large  $\{u_i\}_{i \in \mathcal{J}_1 \cup \mathcal{J}_2}$ , where

$$r_i^{*,2} = \max \left\{ r'_i - \sum_{j \neq i} Q_{ji} r'_j + \sum_{\substack{j \neq i \\ j \notin \mathcal{J}'_1 \cup \mathcal{J}_2}} Q_{ji} r_j^{*,2}, 0 \right\}, \quad \text{for } i \notin \mathcal{J}'_1 \cup \mathcal{J}_2.$$

Consider the fixed point equation that is given by

$$\tilde{r}_i^{*,2} = \max \left\{ r'_i - \sum_{j \neq i} Q_{ji} r'_j + \sum_{\substack{j \neq i \\ j \notin \mathcal{J}_1 \cup \mathcal{J}_2}} Q_{ji} \tilde{r}_j^{*,2}, 0 \right\}, \quad \text{for } i \notin \mathcal{J}_1 \cup \mathcal{J}_2. \quad (7.18)$$

Since  $\mathcal{J}_1 \subseteq \mathcal{J}_1 \cup \mathcal{J}_2$ , by Result 6.4, we obtain that  $\tilde{r}_k^{*,2} = 0$ , for every  $k \in \mathcal{J}'_1 \setminus \mathcal{J}_1$ . By making the convention that  $r_k^{*,2} = 0$  for  $k \in \mathcal{J}'_1 \setminus \mathcal{J}_1$ , we claim that  $r_i^{*,2} = \tilde{r}_i^{*,2}$ , for  $i \notin \mathcal{J}_1 \cup \mathcal{J}_2$ . Since  $\mathcal{J}_1 \cup \mathcal{J}_2 \subseteq \mathcal{J}$ , by Result 6.4, (7.18) and (6.7), we conclude that  $z'_\xi(t) \leq \partial_z(\mathcal{J})$ , for  $t \in [d_2, d_3)$ . Iterating the same procedure  $m$  more times, we can construct a  $\xi$  (by taking  $\{u_i\}_{i \in \mathcal{J}}$  sufficiently large) such that  $z_\xi$  is piecewise linear between neighboring discontinuity points. Moreover, the increasing rate of  $z_\xi$  is less than  $\partial_z(\mathcal{J})$ , i.e.  $z'_\xi(t) \leq \partial_z(\mathcal{J})$ , for  $t \in [0, 1]$ . Therefore, we obtain that  $\partial_z(\mathcal{J}) > a$ .

The last statement of Proposition 6.1 is a consequence of Result 6.4.  $\square$

*Proof of Proposition 6.2.* Let the unique optimal solution of (6.9) be denoted by  $(l_1^*, \dots, l_d^*)$ . To prove that  $A$  is bounded away from

$$\bigcup_{(l_1, \dots, l_d) \in \mathcal{I}_{< l_1^*, \dots, l_d^*}} \prod_{i=1}^d \mathbb{L}_{l_i}(\mu_i),$$

it is sufficient to show that  $A = F^{-1}[a, \infty)$  is bounded away from  $\prod_{i=1}^d \mathbb{L}_{l_i}(\mu_i)$  for all  $(l_1, \dots, l_d) \in \mathcal{I}_{< l_1^*, \dots, l_d^*}$ . To begin with, let  $(l_1, \dots, l_d) \in \mathcal{I}_{< l_1^*, \dots, l_d^*}$ . Under Assumption 4 we have that  $\partial_z(\mathcal{I}) < a$ ,

where  $j \in \mathcal{I}$  if and only if  $l_j \neq 0$ . Applying a similar approach as in the proof of Proposition 6.1, it can be shown that  $F\left(\prod_{i=1}^d \mathbb{L}_{l_i}(\mu_i)\right) \subseteq (-\infty, \partial_z(\mathcal{I})]$ . This implies that there exists  $\delta > 0$  satisfying

$$d\left(F\left(\prod_{i=1}^d \mathbb{L}_{l_i}(\mu_i)\right), [a, \infty)\right) > \delta. \quad (7.19)$$

Moreover, by Result 6.1 we conclude that the mapping  $F$  as composition of Lipschitz continuous mappings (for continuity of  $\pi_1$  see e.g. Theorem 12.5 in [3]) is again Lipschitz continuous. Let  $K_F$  denote the Lipschitz constant of  $F$ . Combining this with (7.19) we conclude that  $d\left(\prod_{i=1}^d \mathbb{L}_{l_i}(\mu_i), F^{-1}([a, \infty))\right) > \delta/K_F$ , hence the second statement is obtained by applying Result 2.2.  $\square$

## References

- [1] S. Asmussen and P. Glynn. *Stochastic simulation: Algorithms and analysis*, volume 57 of *Stochastic Modelling and Applied Probability*. Springer-Verlag New York, 1st edition, 2007.
- [2] S. Asmussen, H. Schmidli, and V. Schmidt. Tail probabilities for non-standard risk and queueing processes with subexponential jumps. *Advances in Applied Probability*, 31(2):422–447, 1999.
- [3] P. Billingsley. *Convergence of probability measures*. John Wiley & Sons, 2013.
- [4] J. Blanchet and P. Glynn. Efficient rare-event simulation for the maximum of heavy-tailed random walks. *The Annals of Applied Probability*, 18(4):1351–1378, August 2008.
- [5] J. Blanchet, P. Glynn, and J. Liu. Efficient rare event simulation for heavy-tailed multiserver queues. Technical report, 2007.
- [6] J. Blanchet and J. Liu. State-dependent importance sampling for regularly varying random walks. *Advances in Applied Probability*, 40(4):1104–1128, Dezember 2008.
- [7] J. Blanchet and J. Liu. Efficient importance sampling in ruin problems for multidimensional regularly varying random walks. *Journal of Applied Probability*, 47(2):301–322, June 2010.
- [8] K. Debicki and M. Mandjes. *Queues and Lévy Fluctuation Theory*. Universitext. Springer International Publishing, 1st edition, 2015.
- [9] P. Dupuis, K. Leder, and H. Wang. Importance sampling for sums of random variables with regularly varying tails. *ACM Transactions on Modeling and Computer Simulation*, 17(3), July 2007.
- [10] P. Dupuis, A. D. Sezer, and H. Wang. Dynamic importance sampling for queueing networks. *The Annals of Applied Probability*, 17(4):1306–1346, August 2007.
- [11] P. Dupuis and H. Wang. Importance sampling, large deviations, and differential games. *Stochastics and Stochastic Reports*, 76(6):481–508, 2004.
- [12] P. Dupuis and H. Wang. On the convergence from discrete to continuous time in an optimal stopping problem. *The Annals of Applied Probability*, 15(2):1339–1366, 2005.
- [13] P. Dupuis and H. Wang. Importance sampling for Jackson networks. *Queueing Systems*, 62(1):113–157, 2009.
- [14] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling extremal events for insurance and finance*, volume 33 of *Stochastic Modelling and Applied Probability*. Springer-Verlag Berlin Heidelberg, 1st edition, 1997.
- [15] S. Foss and D. Korshunov. Heavy tails in multi-server queue. *Queueing Systems: Theory and Applications*, 52(1):31–48, January 2006.



- [16] S. Foss and D. Korshunov. On large delays in multi-server queues with heavy tails. *Mathematics of Operations Research*, 37(2):201–218, 2012.
- [17] S. Foss, D. Korshunov, and S. Zachary. *An introduction to heavy-tailed and subexponential distributions*, volume 38 of *Springer Series in Operations Research and Financial Engineering*. Springer-Verlag New York, 2nd edition, 2013.
- [18] P. Glasserman and S.-G. Kou. Analysis of an importance sampling estimator for tandem queues. *ACM Transactions on Modeling and Computer Simulation*, 5(1):22–42, January 1995.
- [19] P. Glasserman and Y. Wang. Counterexamples in importance sampling for large deviations probabilities. *The Annals of Applied Probability*, 7(3):731–746, August 1997.
- [20] J. M. Harrison and M. I. Reiman. Reflected Brownian motion on an orthant. *The Annals of Probability*, 9(2):302–308, April 1981.
- [21] X. Hu and H. Cui. Generating multi-dimensional discrete distribution random number. In *2010 Sixth International Conference on Natural Computation*, volume 3, pages 1102–1104, August 2010.
- [22] S. Ramasubramanian. A subsidy-surplus model and the skorokhod problem in an orthant. *Mathematics of Operations Research*, 25(3):509–538, 2000.
- [23] C.-H. Rhee, J. Blanchet, and B. Zwart. Sample path large deviations for heavy-tailed Lévy processes and random walks. *eprint arXiv:1606.02795*, 2016.
- [24] A. V. Skorokhod. Stochastic equations for diffusion processes in a bounded region. *Theory of Probability & Its Applications*, 6(3):264–274, 1961.
- [25] A. V. Skorokhod. Stochastic equations for diffusion processes in a bounded region. II. *Theory of Probability & Its Applications*, 7(1):3–23, 1962.
- [26] P. Tankov and R. Cont. *Financial Modelling with Jump Processes, Second Edition*. Chapman and Hall/CRC Financial Mathematics Series. Taylor & Francis, 2nd edition, 2015.
- [27] W. Whitt. *Stochastic-Process Limits*. Springer Series in Operations Research and Financial Engineering. Springer-Verlag New York, 1st edition, 2002.