Proceedings of the 2023 Winter Simulation Conference
C. G. Corlu, S. R. Hunter, H. Lam, B. S. Onggo, J. Shortle, and B. Biller, eds.

CONDITIONAL IMPORTANCE SAMPLING FOR CONVEX RARE-EVENT SETS

Dohyun Ahn
Lewen Zheng

Department of Systems Engineering and Engineering Management
The Chinese University of Hong Kong
Shatin, New Territories, HONG KONG

ABSTRACT

This paper studies the efficient estimation of expectations defined on convex rare-event sets using importance sampling. Classical importance sampling methods often neglect the geometry of the target set, resulting in a significant number of samples falling outside the target set. This can lead to an increase in the relative error of the estimator as the target event becomes rarer. To address this issue, we develop a conditional importance sampling scheme that achieves bounded relative error by changing the sampling distribution to ensure that a majority of samples lie inside the target set. The proposed method is easy to implement and significantly outperforms the existing approaches in various numerical experiments.

1 INTRODUCTION

In this paper, we consider the problem of estimating expectations over rare-event sets. This problem arises in, for example, finance (Glasserman et al. 2000; Glasserman and Li 2005; Chen and Glasserman 2008; Bassamboo et al. 2008), insurance (Asmussen and Albrecher 2010; Bauer et al. 2012), queueing system (Dupuis et al. 2007; Blanchet and Lam 2014; Ma and Whitt 2018), communication network (Haraszti and Townsend 1999; Lokshina and Bartolacci 2012), and machine learning (O'Kelly et al. 2018; Bai et al. 2022). In these contexts, several studies improve the tractability and efficiency of simulation algorithms by focusing on particular geometric forms of a target event set; for example, a union of halfspaces (Ahn and Kim 2018; Owen et al. 2019), a hypercube (Botev et al. 2017; Botev 2017), and a union of polyhedra (Bai et al. 2022; Ahn, D., and L. Zheng. 2023). Nevertheless, few studies consider general convex rare-event sets constructed by nonlinear constraints, which we will explore in this paper.

The primary motivation for this problem stems from distributionally robust rare-event simulation. According to Blanchet and Murthy (2019), the distributionally robust bound of the probability of a convex rare event can be represented as the probability of another convex rare event. In this case, given an uncertainty set of distributions, rare-event simulation of expectations over convex sets turns out to be useful not only to identify the new target set but also to estimate the distributionally robust bound; see Section 5 for more details. Furthermore, our problem has potential applications in pricing deep-out-of-the-money options and evaluating the reliability of transportation networks. It may also be relevant in cases where the structural information of the rare-event set is not available, but the decision maker believes that it is convex (Huang et al. 2018).

When estimating expectations evaluated on rare events, the crude Monte Carlo method produces significantly large simulation errors since it seldom generates samples belonging to the target set. To tackle this issue, we often resort to a well-known variance reduction technique, called importance sampling, which changes the sampling distribution to increase the likelihood of hitting the target set and assigns weights to the generated samples via a likelihood ratio to retain unbiasedness. Finding suitable sampling distributions is a challenging task that has been typically addressed by using the theory of large deviations (Siegmund 1976; Glynn and Iglehart 1989; Heidelberger 1995; Juneja and Shahabuddin 2006). However, despite being

979-8-3503-6966-3/23/$31.00 ©2023 IEEE

provably effective in many settings, this traditional approach usually neglects salient geometric structures of the target set. As a result, the relative errors of the corresponding estimators could grow indefinitely as the target set becomes rarer.

In this paper, we propose to identify prominent geometric information of the target set and integrate them into the sampling distribution. Ahn and Zheng. (2023) use this idea to estimate expectations defined on "polyhedral" rare-event sets and develop a conditional importance sampling method with bounded relative error based on sharp asymptotics. However, it is unclear how and under what conditions this method can be adapted to the setting with "general convex" rare-event sets in a way that the associated relative error remains bounded. In this regard, the main contribution of this paper lies in providing an easy-to-implement adaptation and proving that the bounded relative error property is preserved under mild conditions on the structure of the target set.

Our approach is related to Dupuis and Wang (2004), Glasserman and Li (2005), and Bucklew (2005) in terms of combining importance sampling and conditional Monte Carlo. However, those methods are designed only for specific problem setups such as the evaluation of portfolio credit risk and the computation of tail probabilities associated with random walks. Further, they do not guarantee bounded relative error. In contrast, our method is developed for fairly general objectives and achieves bounded relative error.

The remainder of the paper is organized as follows. Section 2 describes the main problem. In Section 3, we formally review the conditional importance sampling method designed for polyhedral target sets. The modified algorithm for general convex target sets and its theoretical performance guarantee are discussed in Section 4. Section 5 addresses the issue of distributional uncertainty in estimating the probability over convex rare-event sets. In Section 6, we numerically validate the effectiveness of the algorithm. Section 7 concludes the paper.

2 PROBLEM FORMULATION

For each $i = 1, \dots, m$, let $f_i : \mathbb{R}^n \to \mathbb{R}$ be a convex differentiable function that is assumed to be known. Define $\mathscr{E} = \{\mathbf{x} \in \mathbb{R}^n \,|\, f_i(\mathbf{x}) \le 0, i = 1, \dots, m\}$. We assume that $\mathscr{E}$ is full-dimensional and that a random vector $\mathbf{X} = (X_1, \dots, X_n)^\top$ follows the n-dimensional standard normal distribution. We focus on the case where $\mathbf{0} \notin \mathscr{E}$ so that the event $\{\mathbf{X} \in \mathscr{E}\}$ is typically rare. Our main objective is to efficiently estimate the quantity

$$\mathsf{E}[h(\mathbf{X})\mathbb{1}_{\mathscr{E}}(\mathbf{X})], \tag{1}$$

where $h : \mathbb{R}^n \to \mathbb{R}$ is a positive measurable function and for any $\mathbf{v} \in \mathbb{R}^d$ and $\mathscr{A} \subseteq \mathbb{R}^d$, $\mathbb{1}_{\mathscr{A}}(\mathbf{v})$ yields 1 if $\mathbf{v} \in \mathscr{A}$ and 0 otherwise. When $h \equiv 1$, this problem reduces to estimating the probability $\mathsf{P}(\mathbf{X} \in \mathscr{E})$. Although the standard normality assumption is seemingly restrictive, it can be easily relaxed to a general normal distribution with a nonzero mean vector and a positive definite covariance matrix since the convexity of $\mathscr{E}$ is preserved under affine transformations. Furthermore, the analysis in this paper can be extended to normal mixture models, which greatly broadens the applicability of the current setting.

Consider the optimization problem $\min_{\mathbf{x} \in \mathscr{E}} \|\mathbf{x}\|^2$, and denote by $\mathbf{x}^\star$ and η_i its optimal solution and the Karush–Kuhn–Tucker (KKT) multipliers associated with the constraints $f_i(\mathbf{x}) \le 0$, $i = 1, \dots, m$ (Rockafellar 1970, Section 28), respectively. We impose the following conditions throughout the paper.

- (A1) For each $i = 1, \dots, m$, f_i has continuous second order derivatives on a neighborhood of $\mathbf{x}^\star$.
- (A2) There exists $k \in \{1, \dots, m\}$ such that $\eta_i > 0$ if and only if $i \le k$.
- (A3) $(\mathbf{x}^\star, \nabla f_1(\mathbf{x}^\star), \dots, \nabla f_k(\mathbf{x}^\star))$ is an upper triangular matrix with all zero rows at the bottom.

Condition (A1) ensures the constraint functions $f_1, \dots, f_m$ are sufficiently smooth near the point $\mathbf{x}^\star$. Condition (A2) can be satisfied by rearranging the constraints. The last condition can be satisfied through a suitable rotation of the coordinates and a rearrangement of the variables, which does not affect (1) because the standard normal distribution is invariant under such transformations.

To analyze the setting with rare events, let $\mathscr{E}_1 = \{\mathbf{x} \,|\, f_i(\mathbf{x}) \le 0, i = 1,\dots,m\}$ be an arbitrary convex set satisfying conditions (A1) to (A3) with $\mathbf{x}^\star = \mathbf{e}_1$, where $\mathbf{e}_1$ is the unit vector in the direction of the x_1-axis, and define $\mathscr{E}_r = \{\mathbf{x} \in \mathbb{R}^n \,|\, f_i(r^{-1}\mathbf{x}) \le 0, i = 1,\dots,m\}$ for all $r > 0$. Note that $\mathscr{E}_r$ also satisfies conditions (A1) to (A3) with $\mathbf{x}^\star = r\mathbf{e}_1$. Moreover, $\mathsf{E}[h(\mathbf{X})\mathbb{1}_{\mathscr{E}_r}(\mathbf{X})] \to 0$ as $r \to \infty$ as long as $\mathsf{E}[h(\mathbf{X})] < \infty$. In evaluating the efficiency of an estimator for $\mathsf{E}[h(\mathbf{X})\mathbb{1}_{\mathscr{E}_r}(\mathbf{X})]$, we use the following notion in the rare event simulation literature (Bucklew 2004).

Definition 1 Let $Z(\mathscr{E}_r;h)$ be an unbiased estimator for $\mathsf{E}[h(\mathbf{X})\mathbb{1}_{\mathscr{E}_r}(\mathbf{X})]$. We say that $Z(\mathscr{E}_r;h)$ has *bounded relative error* as $r \to \infty$ if $\limsup_{r\to\infty} \mathbb{V}ar(Z(\mathscr{E}_r;h))^{1/2}/\mathsf{E}[h(\mathbf{X})\mathbb{1}_{\mathscr{E}_r}(\mathbf{X})] < \infty$.

An estimator with bounded relative error is said to be strongly efficient in the sense that regardless of the rarity of the target event, only a fixed number of samples is required to maintain a specified accuracy level. However, it is widely known that in many problems, finding a strongly efficient estimator is extremely difficult (Asmussen and Glynn 2007).

3 CONDITIONAL IMPORTANCE SAMPLING FOR POLYHEDRAL TARGET SETS

This section briefly reviews the conditional importance sampling (CIS) method in Ahn and Zheng. (2023) that is designed for the case when $\mathscr{E}$ is a polyhedron, i.e., $\mathscr{E} = \{\mathbf{x} \,|\, \mathbf{a}_i^\top \mathbf{x} \ge b_i, i = 1,\dots,m\}$ for $\mathbf{a}_i \in \mathbb{R}^n$ and $b_i \in \mathbb{R}$, $i = 1,\dots,m$. In the next section, we will extend this method to the case with general convex rare-event sets.

Basically, the CIS method in Ahn and Zheng. (2023) increases the likelihood of hitting the rare-event set $\mathscr{E}_r = \{\mathbf{x} \,|\, \mathbf{a}_i^\top \mathbf{x} \ge rb_i\}$ by combining conditional Monte Carlo and importance sampling. To be more specific, firstly, the method generates a sample x_1 of X_1 by sampling a standard exponential random sample y and transforming it into $x_1 = r + y/r$. This is based on the fact that $\{\mathbf{X} \in \mathscr{E}_r\} \subseteq \{X_1 \ge r\}$ and the observation that $r(X_1 - r)$ conditioned on $\{X_1 \ge r\}$ converges in distribution to a standard exponential random variable when $r \to \infty$.

Next, to obtain a near-optimal sampling distribution of $(X_2,\dots,X_n)$ conditioned on X_1, the authors show that the set $\mathscr{E}_r(X_1) := \{(x_2,\dots,x_n) \,|\, \mathbf{x} \in \mathscr{E}_r, x_1 = X_1\}$, i.e., the cross section of $\mathscr{E}_r$ at $x_1 = X_1$, gets closer to $\{(x_2,\dots,x_n) \,|\, (x_2,\dots,x_s) \in \Delta(X_1 - r), (x_{s+1},\dots,x_n) \in \widetilde{\mathscr{D}}\}$ as $r \to \infty$, where $s = \operatorname{rank}(\mathbf{a}_1,\dots,\mathbf{a}_k)$,

$$\Delta(y) = \left\{ (x_2,\dots,x_s) \,\middle|\, a_{i,1}y + \sum_{j=2}^{s} a_{i,j}x_j \ge 0, \ \forall\, i : \mathbf{a}_i^\top \mathbf{x}^\star = b_i, a_{i,s+1} = \cdots = a_{i,n} = 0 \right\}, \tag{2}$$

and $\widetilde{\mathscr{D}}$ is some convex polyhedron in $\mathbb{R}^{n-s}$ independent of X_1. They further verify that $\Delta(y)$ is a convex bounded polyhedron with nonempty interior for $y > 0$ and satisfies $\Delta(0) = \{\mathbf{0}\}$, which implies that $x_2^2 + \cdots + x_s^2 \approx 0$ whenever $(x_2,\dots,x_s) \in \Delta(X_1 - r)$ as $r \to \infty$. Based on this result, in their method, the random vector $(X_2,\dots,X_s)$ is sampled from the uniform distribution over $\Delta(X_1 - x_1^\star)$ conditioned on X_1, which is easy to simulate in many cases, while the random vector $(X_{s+1},\dots,X_n)$ is sampled from the standard normal distribution.

Accordingly, the unbiased estimator for $\mathsf{E}[h(\mathbf{X})\mathbb{1}_{\mathscr{E}}(\mathbf{X})]$ based on the above method is written as

$$Z(\mathscr{E};h) := h(\mathbf{X})\mathbb{1}_{\mathscr{E}}(\mathbf{X})L(\mathbf{X}), \tag{3}$$

where $\mathbf{X}$ is generated as described above and the associated likelihood ratio function $L(\mathbf{x})$ is given by

$$L(\mathbf{x}) = \frac{\operatorname{vol}(\Delta(x_1 - x_1^\star))}{(2\pi)^{s/2} x_1^\star} e^{x_1^\star(x_1 - x_1^\star) - (x_1^2 + \cdots + x_s^2)/2}.$$

Note that $\operatorname{vol}(\Delta(y))$ means the Lebesgue measure of $\Delta(y)$. Below we restate Theorem 4 of Ahn and Zheng. (2023), which establishes the strong efficiency of $Z(\mathscr{E};h)$ under a mild condition on h.

Algorithm 1 Conditional Importance Sampling

Find $\mathbf{x}^\star = \operatorname{argmin}_{\mathbf{x}\in\mathscr{E}} \|\mathbf{x}\|^2$, its KKT multipliers $\{\eta_i\}_{i=1}^m$, and an integer k satisfying condition (A2)
Set $\mathbf{a}_i = -\nabla f_i(\mathbf{x}^\star)$, $b_i = f_i(\mathbf{x}^\star) - \nabla f_i(\mathbf{x}^\star)^\top \mathbf{x}^\star$, and $s = \operatorname{rank}(\mathbf{a}_1,\ldots,\mathbf{a}_k)$
Sample $Y = y$ from the standard exponential distribution and set $x_1 = x_1^\star + y/x_1^\star$
Set $L_1 = (2\pi)^{-1/2}(x_1^\star)^{-1} e^{-x_1^2/2 + x_1^\star(x_1 - x_1^\star)}$ and $L_2 = 1$
if $s \geq 2$ **then**
    Sample $(X_2,\ldots,X_s) = (x_2,\ldots,x_s)$ from the uniform distribution over $\Delta(x_1 - x_1^\star)$ defined in (2)
    Set $L_2 = (2\pi)^{-(s-1)/2}\operatorname{vol}(\Delta(x_1 - x_1^\star))\exp(-(x_2^2+\cdots+x_s^2)/2)$
end if
if $s < n$ **then**
    Sample $(X_{s+1},\ldots,X_n) = (x_{s+1},\ldots,x_n)$ from the standard normal distribution
end if
return $h(\mathbf{x})\mathbb{1}_{\mathscr{E}}(\mathbf{x})L_1L_2$.

Theorem 1 Suppose that $f_i(\mathbf{x}) = b_i - \mathbf{a}_i^\top \mathbf{x}$ for some $\mathbf{a}_i \in \mathbb{R}^n$ and $b_i \in \mathbb{R}$ satisfying conditions (A1) to (A3) for each i, and that there exist $C > 0$ and $\kappa \geq 0$ satisfying $h(\mathbf{x})/h(\mathbf{y}) \leq C\exp(\kappa\|\mathbf{x}-\mathbf{y}\|)$ for all $\mathbf{x},\mathbf{y} \in \mathbb{R}^n$. Let $s = \operatorname{rank}(\mathbf{a}_1,\ldots,\mathbf{a}_k)$. Then, there exist two constants $C_1, C_2 \in (0,\infty)$ such that

$$\liminf_{r\to\infty} h(r\mathbf{e}_1)^{-1} r^s e^{r^2/2}\mathsf{E}[h(\mathbf{X})\mathbb{1}_{\mathscr{E}_r}(\mathbf{X})] \geq C_1 \text{ and } \limsup_{r\to\infty} h(r\mathbf{e}_1)^{-2} r^{2s} e^{r^2}\mathsf{E}[Z(\mathscr{E}_r;h)^2] \leq C_2.$$

Consequently, $Z(\mathscr{E}_r;h)$ has bounded relative error as $r \to \infty$.

It is worth noting that the condition on h in Theorem 1 is satisfied by all polynomial functions and many log-concave functions including $h(\mathbf{x}) = \exp(\theta^\top \mathbf{x})$ for some $\theta \in \mathbb{R}^n$, which appear in various applications.

4 CONDITIONAL IMPORTANCE SAMPLING FOR CONVEX EVENT SETS

We now turn to the general setting with $\mathscr{E} = \{\mathbf{x} \,|\, f_i(\mathbf{x}) \leq 0, i = 1,\ldots,m\}$ where f_i is a possibly nonlinear function for each i. Somewhat surprisingly, a simple adaptation of the CIS discussed in the last section is sufficient to maintain its strong efficiency for general convex rare-event sets, which significantly improves the applicability of the method.

Specifically, using the first-order approximation of $f_i(\cdot)$ at $\mathbf{x}^\star$, we construct the polyhedral set $\mathscr{P} := \{\mathbf{x} \,|\, \mathbf{a}_i^\top \mathbf{x} \geq b_i, i = 1,\ldots,m\}$ with $\mathbf{a}_i = -\nabla f_i(\mathbf{x}^\star)$ and $b_i = f_i(\mathbf{x}^\star) - \nabla f_i(\mathbf{x}^\star)^\top \mathbf{x}^\star$. Due to the convexity of $\mathscr{E}$, $\mathscr{E} \subseteq \mathscr{P}$ and $\mathscr{P}$ is a close approximation of $\mathscr{E}$ particularly near $\mathbf{x}^\star$ where most masses are concentrated. Thanks to these properties, most samples belonging to $\mathscr{P}$ would be included in $\mathscr{E}$. Based on this intuition, we propose to sample $\mathbf{X}$ using the CIS method in Section 3, considering $\mathscr{P}$ as the target set. Note that $\mathscr{P}$ satisfies conditions (A1) to (A3). Finally, our estimator becomes

$$\widetilde{Z}(\mathscr{E};h) := Z(\mathscr{P};h)\mathbb{1}_{\mathscr{E}}(\mathbf{X}),$$

where $Z(\cdot;h)$ is defined in (3). This is clearly an unbiased estimator since $\mathscr{E} \subseteq \mathscr{P}$. The full simulation procedure is summarized in Algorithm 1.

Our estimator $\widetilde{Z}(\mathscr{E}_r;h)$ for $\mathsf{E}[h(\mathbf{X})\mathbb{1}_{\mathscr{E}_r}(\mathbf{X})]$ induces asymptotically negligible efficiency loss as $r \to \infty$, which is formalized in Theorem 2.

Theorem 2 Suppose that $f_1,\ldots,f_m$ are convex functions satisfying conditions (A1) to (A3) and that h satisfies the condition in Theorem 1. Then, $\widetilde{Z}(\mathscr{E}_r;h)$ achieves bounded relative error as $r \to \infty$.

Proof of Theorem 2. Define $\mathscr{P}_r := \{r\mathbf{x} \,|\, \mathbf{x} \in \mathscr{P}\}$. We have $\mathscr{E}_r \subseteq \mathscr{P}_r$ due to the convexity of $\mathscr{E}_r$. Therefore, using Theorem 1, there exists a positive constant C_2 such that the second moment of the estimator $\widetilde{Z}(\mathscr{E}_r;h)$ satisfies

$$\limsup_{r\to\infty} h(r\mathbf{e}_1)^{-2} r^{2s} e^{r^2}\mathsf{E}\big[\widetilde{Z}(\mathscr{E}_r;h)^2\big] \leq \limsup_{r\to\infty} h(r\mathbf{e}_1)^{-2} r^{2s} e^{r^2}\mathsf{E}\big[Z(\mathscr{P}_r;h)^2\big] \leq C_2. \tag{4}$$

Since the function h satisfies $h(\mathbf{y}) \geq C^{-1} h(\mathbf{x}) \exp(-\kappa \|\mathbf{x}-\mathbf{y}\|)$ for all points $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, we have

$$\mathsf{E}[h(\mathbf{X}) \mathbb{1}_{\mathscr{E}_r}(\mathbf{X})] \geq h(r\mathbf{e}_1) C^{-1} \mathsf{E}\left[e^{-\kappa \|\mathbf{X}-r\mathbf{e}_1\|} \mathbb{1}_{\mathscr{E}_r}(\mathbf{X})\right]. \tag{5}$$

For each $r > 0$, we use a change of variables: $y_1 = r(x_1 - r)$, $y_i = x_i/(x_1 - r)$ for $i = 2, \dots, s$, and $y_j = x_j$ for $j > s$, and observe that

$$\begin{aligned} \mathsf{E}\left[e^{-\kappa \|\mathbf{X}-r\mathbf{e}_1\|} \mathbb{1}_{\mathscr{E}_r}(\mathbf{X})\right] = (2\pi)^{-n/2} e^{-r^2/2} r^{-s} \int_{\mathscr{D}_r} & y_1^{s-1} e^{-\kappa (y_1^2/r^2 + y_1^2(1+y_2^2+\cdots+y_s^2)/r^2 + y_{s+1}^2+\cdots+y_n^2)^{1/2}} \\ & \times e^{-y_1 - y_1^2(y_2^2+\cdots+y_s^2)/(2r^2) - (y_{s+1}^2+\cdots+y_n^2)/2} \mathrm{d}\mathbf{y}, \end{aligned} \tag{6}$$

where $\mathscr{D}_r = \{\mathbf{y} \,|\, f_i(1 + y_1/r^2, y_1 y_2/r^2, \dots, y_1 y_s/r^2, y_{s+1}/r, \dots, y_n/r) \leq 0,\, i = 1, \dots, m\}$. From condition (A1) and the Taylor expansion of f_i at $\mathbf{x}^\star = \mathbf{e}_1$, the constraints of $\mathscr{D}_r$ can be recast as

$$f_i(\mathbf{x}^\star) + \sum_{j=s+1}^{n} \frac{\partial f_i(\mathbf{x}^\star)}{\partial x_j} \frac{y_j}{r} + \frac{\partial f_i(\mathbf{x}^\star)}{\partial x_1} \frac{y_1}{r^2} + \sum_{j=2}^{s} \frac{\partial f_i(\mathbf{x}^\star)}{\partial x_j} \frac{y_1 y_j}{r^2} + \sum_{j,j'>s} \frac{\partial^2 f_i(\mathbf{x}^\star)}{\partial x_j \partial x_{j'}} \frac{y_j y_{j'}}{r^2} + o(r^{-2}) \leq 0,\ \forall\, i. \tag{7}$$

Define $I = \{i \,|\, f_i(\mathbf{x}^\star) = 0\}$, $I_1 = \{i \in I \,|\, \partial f_i(\mathbf{x}^\star)/\partial x_j = 0, \forall\, j > s\}$, and $I_2 = I \setminus I_1$. Denote by $\mathscr{D}$ the set of points $\mathbf{y} \in \mathbb{R}^n$ satisfying the following inequalities:

$$\frac{\partial f_i(\mathbf{x}^\star)}{\partial x_1} y_1 + \sum_{j=2}^{s} \frac{\partial f_i(\mathbf{x}^\star)}{\partial x_j} y_1 y_j + \sum_{j,j'>s} \frac{\partial^2 f_i(\mathbf{x}^\star)}{\partial x_j \partial x_{j'}} y_j y_{j'} \leq 0,\ \forall\, i \in I_1, \tag{8}$$

$$\sum_{j=s+1}^{n} \frac{\partial f_i(\mathbf{x}^\star)}{\partial x_j} y_j \leq 0,\ \forall\, i \in I_2. \tag{9}$$

If $i' \notin I$, $f_{i'}(\mathbf{x}^\star) < 0$, and hence, for each $\mathbf{y} \in \mathbb{R}^n$, (7) is satisfied for all sufficiently large r. For all $i \in I_1$, since the first two terms of (7) are zeros, we have (8) by multiplying both sides of (7) by r^2 and letting $r \to \infty$. Similarly, for all $i \in I_2$, we obtain (9) by multiplying both sides of (7) by r and letting $r \to \infty$. These observations indicate that $\limsup_{r\to\infty} \mathscr{D}_r \subseteq \mathscr{D}$ (Rockafellar and Wets 2009, part (a) of Theorem 4.10). On the other hand, for any interior point of $\mathscr{D}$, it lies in the interior of $\mathscr{D}_r$ for sufficiently large r according to (7), which implies that $\liminf_{r\to\infty} \mathscr{D}_r \supseteq \mathscr{D}$ (Rockafellar and Wets 2009, part (b) of Theorem 4.10). In conclusion, $\lim_{r\to\infty} \mathscr{D}_r = \mathscr{D}$. By the dominated convergence theorem, the integral in (6) converges to $C_3 := \int_{\mathscr{D}} y_1^{s-1} \exp(-y_1 - (y_{s+1}^2 + \cdots + y_n^2)/2 - \kappa (y_{s+1}^2 + \cdots + y_n^2)^{1/2}) \mathrm{d}\mathbf{y}$ as $r \to \infty$. Thus, from (5) and (6), we find the asymptotic lower bound of the first moment as follows:

$$\liminf_{r\to\infty} h(r\mathbf{e}_1)^{-1} r^s e^{r^2/2} \mathsf{E}[h(\mathbf{X}) \mathbb{1}_{\mathscr{E}_r}(\mathbf{X})] \geq (2\pi)^{-n/2} C^{-1} C_3. \tag{10}$$

It remains to show that C_3 is positive and finite. With some abuse of notation, define

$$\Delta(y) = \left\{ (x_2, \dots, x_s) \,\middle|\, \frac{\partial f_i(\mathbf{x}^\star)}{\partial x_1} y + \sum_{j=2}^{s} \frac{\partial f_i(\mathbf{x}^\star)}{\partial x_j} x_j \leq 0,\ \forall\, i \in I_1 \right\}, \tag{11}$$

and let $\widetilde{\mathscr{D}}$ be the set of points $(y_{s+1}, \dots, y_n)$ satisfying (9) . By the Karush–Kuhn–Tucker theorem (Bertsekas 1999), the two optimization problems $\min_{\mathbf{x}\in\mathscr{E}} \|\mathbf{x}\|^2$ and $\min_{\mathbf{x}\in\mathscr{P}} \|\mathbf{x}\|^2$ share the same optimal solution $\mathbf{x}^\star = \mathbf{e}_1$ and the same Lagrangian multipliers. Hence, as noted in Section 3, $\Delta(y)$ is a bounded set with nonempty interior for all $y > 0$. Further, it can be shown that $\widetilde{\mathscr{D}}$ has an interior point using Farkas' lemma (Rockafellar 1970, Theorem 22.2) and the facts that $s = \operatorname{rank}(\nabla f_1(\mathbf{x}^\star), \dots, \nabla f_k(\mathbf{x}^\star))$ and k is the number of positive KKT multipliers of $\min_{\mathbf{x}\in\mathscr{E}_1}\|\mathbf{x}\|^2$. Let $(x_2',\ldots,x_s')$ be an interior point of $\Delta(1)$ and $(x_{s+1}',\ldots,x_n')$ be an interior point of $\widetilde{\mathscr{D}}$. For all $i\in I_1$, the convexity of f_i implies that $\sum_{j,j'>s}(\partial^2 f_i(\mathbf{x}^\star)/\partial x_j\partial x_{j'})x_j'x_{j'}'\geq 0$. If we choose

$$x_1' = 1+\max_{i\in I_1}\left\{\left|\frac{\partial f_i(\mathbf{x}^\star)}{\partial x_1}+\sum_{j=2}^{s}\frac{\partial f_i(\mathbf{x}^\star)}{\partial x_j}x_j'\right|^{-1}\sum_{j,j'>s}\frac{\partial^2 f_i(\mathbf{x}^\star)}{\partial x_j\partial x_{j'}}x_j'x_{j'}'\right\},$$

then $\mathbf{x}'$ is an interior point of $\mathscr{D}$. This implies that $\mathscr{D}$ has nonempty interior and, by definition, $C_3>0$. Note that $\mathscr{D}\subseteq\{\mathbf{y}\,|\,y_1\geq 0\}$ since $\mathscr{D}_r\subseteq\{\mathbf{y}\,|\,y_1\geq 0\}$ for each r and $\lim_{r\to\infty}\mathscr{D}_r=\mathscr{D}$. Accordingly, from (8) and (11), we observe that $(y_2,\ldots,y_s)\in\Delta(1)$ for any $\mathbf{y}\in\mathscr{D}$ satisfying $y_1>0$. Consequently, we have $C_3<\infty$ due to the boundedness of $\Delta(1)$. Combining (4) and (10), the proof is complete. □

We conclude this section by pointing out that Algorithm 1 can be extended to handle the case where the target set $\mathscr{E}$ is a union of convex sets $\mathscr{E}^1,\ldots,\mathscr{E}^p$ with a finite p. The idea is to sample $\mathbf{X}$ from the mixture $\sum_{i=1}^p w_i g_i(\mathbf{x})$, where $g_i(\cdot)$ is the CIS density associated with the set $\mathscr{E}^i$, and the weights satisfy $w_i>0$ and $\sum_{i=1}^p w_i=1$. It can be easily shown that the resulting estimator is unbiased and has bounded relative error. Moreover, additional variance reduction is possible by judiciously selecting w_i. The reader is referred to the appendix of Ahn and Zheng. (2023) for a more detailed discussion.

5 PROBABILITY OF RARE EVENTS UNDER DISTRIBUTIONAL UNCERTAINTY

In this section, we consider the problem of estimating the probability $\mathsf{P}(\mathbf{X}\in\mathscr{E})$ when the distribution of $\mathbf{X}$ is subject to some degree of uncertainty and the event $\{\mathbf{X}\in\mathscr{E}\}$ is rare. This problem often arises in risk analysis contexts where the distributional assumption cannot be fully justified due to the lack of data. To capture the additional hazard inflicted by distributional uncertainty, a useful approach is to quantify the effect of changes in probability laws (Glasserman and Xu 2014; Ghosh and Lam 2019). Among various quantification tools, we particularly focus on the optimal transport framework of Blanchet and Murthy (2019) to highlight the significance of our algorithm in this setting.

Distributionally robust rare-event probability. With some abuse of notation, let $\mathsf{P}(\cdot)$ and $\mathsf{Q}(\cdot)$ be the baseline distribution and an alternative distribution of $\mathbf{X}$, respectively, and denote by E the expectation under the distribution P. Define the collection of plausible distributions of $\mathbf{X}$ as $\mathscr{U}(\varepsilon):=\{\mathsf{Q}(\cdot):\inf_{\pi\in\Pi(\mathsf{Q},\mathsf{P})}\mathsf{E}_\pi[\|\mathbf{X}-\mathbf{Y}\|]\leq\varepsilon\}$, where $\Pi(\mathsf{Q},\mathsf{P})$ is the collection of couplings between $\mathsf{Q}(\cdot)$ and $\mathsf{P}(\cdot)$, and E_π means the expectation over the pair of random vectors $(\mathbf{X},\mathbf{Y})$ following the joint distribution π. The parameter $\varepsilon>0$ specifies the maximum degree of uncertainty in the distribution of $\mathbf{X}$. For fixed ε, the goal is to estimate

$$\sup_{\mathsf{Q}(\cdot)\in\mathscr{U}(\varepsilon)}\mathsf{Q}(\mathbf{X}\in\mathscr{E}), \tag{12}$$

which provides a sharp upper bound of our rare-event probability $\mathsf{P}(\mathbf{X}\in\mathscr{E})$ that is robust to distributional uncertainty. Note that the optimal choice of ε is problem-dependent and thus beyond the scope of this paper. It is shown in Blanchet and Murthy (2019) that (12) is equal to the following probability under the baseline distribution:

$$\alpha(\delta):=\mathsf{P}(\mathbf{X}\in N_\delta(\mathscr{E})),$$

where $\delta=\inf\{u\geq 0\,|\,\lambda(u)\geq\varepsilon\}$,

$$\lambda(u):=\mathsf{E}\left[\mathbb{1}_{N_u(\mathscr{E})}(\mathbf{X})\cdot\min_{\mathbf{y}\in\mathscr{E}}\|\mathbf{X}-\mathbf{y}\|\right],$$

and $N_u(\mathscr{E}):=\{\mathbf{x}\,|\,\inf_{\mathbf{y}\in\mathscr{E}}\|\mathbf{x}-\mathbf{y}\|\leq u\}$ for all $u\geq 0$. Since $\mathscr{E}$ is a rare event set, so is $N_u(\mathscr{E})$ for all sufficiently small u by definition. Thus, if $\mathscr{E}$ is convex, both quantities $\lambda(u)$ and $\alpha(\delta)$ are expectations over convex rare-event sets that can be efficiently computed by our CIS method. This observation can be further generalized to the case with a union of convex sets since for $\mathscr{E} = \bigcup_{i=1}^{p} \mathscr{E}^i$ with convex sets $\mathscr{E}^1, \dots, \mathscr{E}^p$, we have $N_\delta(\mathscr{E}) = \bigcup_{i=1}^{p} N_\delta(\mathscr{E}^i)$ and each $N_\delta(\mathscr{E}^i)$ is convex.

Estimation of $\alpha(\delta)$. As alluded to earlier, we can use our CIS method to estimate $\alpha(\delta)$ in two steps. In Step 1, we estimate δ by numerically calculating $\widehat{\delta} := \inf\{u \,|\, \widehat{\lambda}(u) \geq \varepsilon\}$ with the CIS estimate $\widehat{\lambda}(u)$ of $\lambda(u)$ for each $u \geq 0$. In Step 2, given $\widehat{\delta}$, we obtain the CIS estimate $\widehat{\alpha}(\widehat{\delta})$ of $\alpha(\widehat{\delta})$ and consider it as an estimate for $\alpha(\delta)$. Due to this two-step estimation procedure, the variance of $\widehat{\delta}$ in the first step has a significant impact on the variance of the final estimate $\widehat{\alpha}(\widehat{\delta})$ in the second step.

To quantify the overall error of the proposed procedure, we follow the standard analysis of simulation output with input uncertainty, see e.g., Glynn and Lam (2018). Denote the normal distribution with mean μ and variance σ^2 by $\mathscr{N}(\mu, \sigma^2)$. Suppose that we use N_1 samples to estimate δ and N_2 samples for estimating $\alpha(\widehat{\delta})$. The central limit theorem for M-estimates (Serfling 1980, Theorem 7.2.2A) yields $\widehat{\delta} \overset{d}{\approx} \mathscr{N}(\delta, N_1^{-1}\sigma^2(\delta)\lambda'(\delta)^{-2})$ as $N_1 \to \infty$ under some mild conditions, e.g., the differentiability of $\lambda(\cdot)$ at δ (Uryas'ev 1994, Theorem 2.1), where $\overset{d}{\approx}$ represents approximation in distribution and $\sigma^2(\cdot)$ is the variance of $\widehat{\lambda}(\cdot)$. Moreover, by the central limit theorem, we have $\widehat{\alpha}(\widehat{\delta}) \overset{d}{\approx} \mathscr{N}(\alpha(\widehat{\delta}), N_2^{-1}s^2(\widehat{\delta}))$ as $N_2 \to \infty$ given $\widehat{\delta}$, where $s^2(\cdot)$ is the variance of $\widehat{\alpha}(\cdot)$. Finally, note that $\alpha(\widehat{\delta}) \overset{d}{\approx} \mathscr{N}(\alpha(\delta), N_1^{-1}\alpha'(\delta)^2\lambda'(\delta)^{-2}\sigma^2(\delta))$ by the delta method (Serfling 1980, Theorem 3.1A), where the existence of $\alpha'(\delta)$ is also guaranteed by Theorem 2.1 of Uryas'ev (1994). Then, under the assumption that there exists a real number $\beta \in (0,1)$ such that $N_1/(N_1+N_2) \to \beta$ as $\min\{N_1, N_2\} \to \infty$, combining all these results gives

$$\lim_{\min\{N_1,N_2\}\to\infty} \sqrt{N_1+N_2}(\widehat{\alpha}(\widehat{\delta}) - \alpha(\delta)) = \mathscr{N}\left(0, \frac{\alpha'(\delta)^2}{\lambda'(\delta)^2}\frac{\sigma^2(\delta)}{\beta} + \frac{s^2(\delta)}{1-\beta}\right), \tag{13}$$

which provides a basis of forming a confidence interval for $\widehat{\alpha}(\widehat{\delta})$. From (13), one can clearly see that our method can contribute to reducing both $\sigma^2(\delta)$ and $s^2(\delta)$, thereby improving the simulation efficiency for estimating $\alpha(\delta)$. In Section 6, we will numerically validate the effectiveness of our approach in this problem.

6 NUMERICAL EXPERIMENTS

To illustrate the numerical performance of the CIS method compared with the existing methods, we report two performance indicators for each experiment conducted below: variance ratio (VR) and efficiency ratio (ER). In particular, for a Monte Carlo estimator Z^{MC} with simulation time τ^{MC} and another estimator Z with simulation time τ, we define $\text{VR} := \mathbb{V}ar(Z^{MC})/\mathbb{V}ar(Z)$ and $\text{ER} := \text{VR} \times \tau^{MC}/\tau$. We also report the relative error of an estimator Z at the 95% confidence level defined as $1.96\sqrt{\mathbb{V}ar(Z)}(\mathsf{E}[Z])^{-1}$. Although VR is commonly used for comparing simulation efficiencies, we emphasize that ER is a better efficiency indicator in terms of taking both the variance reduction and simulation time into account.

6.1 Asian Option Pricing

We consider the pricing of Asian put options in this section. Note that the same methodology can be applied to the pricing of basket options. For simplicity, consider a discretely monitored Asian put option whose payoff at maturity T is given by $\max\{0, K - n^{-1}\sum_{i=1}^{n} S(iT/n)\}$, where $S(t)$ is the time-t price of the underlying asset and $K > 0$ is the strike price. We assume the Black-Scholes model under the risk-neutral measure, i.e., the price dynamics can be written as $S(t) = S(0)\exp((r - \sigma^2/2)t + \sigma B(t))$, where $\{B(t) \,|\, t \geq 0\}$ is a standard Brownian motion, r is the risk-free rate, and σ is the volatility of the asset price. By the risk-neutral pricing theory (Shreve 2004), the price of the Asian put option at time 0 is given by

$$\mathsf{E}\left[e^{-rT}\left(K - \frac{1}{n}\sum_{i=1}^{n}\exp\left(\left(r - \frac{\sigma^2}{2}\right)\frac{iT}{n} + \sigma\sqrt{\frac{T}{n}}(X_1 + \cdots + X_i)\right)\right)\mathbb{1}_{\mathscr{E}}(\mathbf{X})\right], \tag{14}$$

Table 1: Estimates of Asian put option prices based on four different simulation methods.

Method	K	Estimate (95% rel. err.)	Time (sec)	VR	ER
MC	70	8.15×10^{-5} (5.20%)	67.12	–	–
	65	1.67×10^{-5} (10.58%)	67.26	–	–
	60	2.66×10^{-6} (24.38%)	66.96	–	–
	55	2.54×10^{-7} (73.69%)	67.32	–	–
CV	70	8.39×10^{-5} (3.93%)	86.54	1.7	1.3
	65	1.53×10^{-5} (9.65%)	84.80	1.4	1.1
	60	2.46×10^{-6} (22.53%)	85.14	1.4	1.1
	55	2.45×10^{-7} (80.17%)	85.17	0.9	0.7
GHS	70	8.16×10^{-5} (0.01%)	118.76	1.3×10^5	7.5×10^4
	65	1.58×10^{-5} (0.03%)	119.50	1.5×10^5	8.4×10^4
	60	2.28×10^{-6} (0.04%)	121.98	5.9×10^5	3.2×10^5
	55	2.26×10^{-7} (0.05%)	120.74	3.1×10^6	1.7×10^6
CIS	70	8.16×10^{-5} (0.02%)	335.08	7.8×10^4	1.6×10^4
	65	1.58×10^{-5} (0.02%)	337.74	3.3×10^5	6.7×10^4
	60	2.28×10^{-6} (0.02%)	329.67	2.0×10^6	4.0×10^5
	55	2.26×10^{-7} (0.02%)	329.18	1.5×10^7	3.2×10^6

where $\mathbf{X}=(X_1,\ldots,X_n)$ follows the standard normal distribution and

$$\mathscr{E}=\left\{\mathbf{x}\,\middle|\,\log\left[\sum_{j=1}^{n}S_0\exp\left(\left(r-\frac{\sigma^2}{2}\right)\frac{jT}{n}+\sigma\sqrt{\frac{T}{n}}\sum_{i=1}^{j}x_i\right)\right]\le\log(nK)\right\}.$$

The event $\{\mathbf{X}\in\mathscr{E}\}$ represents the case where the option yields a positive payoff, and we focus on deep-out-of-the money options for which K is extremely small so that $\{\mathbf{X}\in\mathscr{E}\}$ becomes a rare event. Furthermore, $\mathscr{E}$ is a convex set because the log-sum-exp function $f(x_1,\ldots,x_n)=\log(e^{x_1}+\cdots+e^{x_n})$ is convex. Therefore, (14) corresponds to a special case of (1) which can be efficiently estimated with Algorithm 1.

In the numerical experiments, we compare four different simulation algorithms: the crude Monte Carlo (MC), the control variate method in Kemna and Vorst (1990) (CV), the method in Glasserman et al. (1999) (GHS), and Algorithm 1 (CIS). CV uses the payoff of the discretely monitored geometric average Asian option as a control variate, which is feasible because its price admits the Black-Scholes formula. By combining the optimal exponential twisting with stratification along an optimal direction, GHS has been shown to have huge variance reduction in pricing path-dependent options. We use 10^8 samples for all methods and set $T=1$, $S_0=200$, $\sigma=50\%$, $r=6\%$, and $n=250$. The results are reported in Table 1.

Based on the VR and ER columns of Table 1, CV is almost as inefficient as MC for pricing deep-out-of-the-money options. While CIS underperforms GHS for relatively large K, the former completely dominates the latter under rare events, i.e., when K is small. In particular, CIS has higer values of ER than GHS for $K\le 60$, indicating its superior efficiency despite the longer simulation time. This result is remarkable since GHS is designed for pricing path-dependent options, while CIS is a general-purpose algorithm. Further, the variance of the GHS estimator can grow at a subexponential rate as the event $\{\mathbf{X}\in\mathscr{E}\}$ becomes rarer, whereas CIS attains bounded relative error in the same setting. Although the prices in Table 1 are perhaps too small to be of much practical interest, the significant variance reductions illustrate the potential of the CIS method.

6.2 Stochastic Network

We consider a stochastic network where the edge weights are random. The goal is to estimate the probability that the total weight along the shortest path between two nodes is within a critical threshold (Frank 1969;

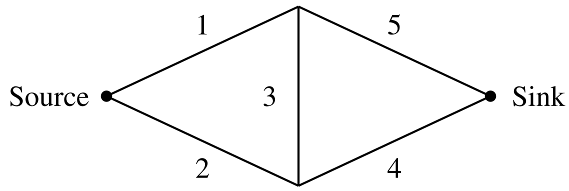

Figure 1: A bridge network.

Table 2: Estimates of probabilities of arriving on time in a bridge network based on three different simulation methods.

Method	γ	Estimate (95% rel. err.)	Time (sec)	VR	ER
MC	0.3	1.01×10^{-3} (0.62%)	6.36	–	–
	0.2	1.18×10^{-4} (1.81%)	6.47	–	–
	0.1	1.39×10^{-6} (16.62%)	6.43	–	–
ET	0.3	1.02×10^{-3} (0.05%)	41.86	177.3	26.96
	0.2	1.19×10^{-4} (0.05%)	40.93	1250.1	197.48
	0.1	1.48×10^{-6} (0.06%)	40.00	70135.9	11269.57
CIS	0.3	1.02×10^{-3} (0.02%)	66.38	1210.7	116.07
	0.2	1.19×10^{-4} (0.02%)	69.86	8320.1	770.12
	0.1	1.48×10^{-6} (0.02%)	69.29	582470.1	54028.63

Kulkarni 1986; Asghari et al. 2015). According to Fan et al. (2005) and Jaillet et al. (2016), this problem has been applied to transportation planning under uncertainty where punctuality of arrival is important; e.g., scheduling freights to deliver perishable goods. Another example of the problem is routing in certain computer systems where timely delivery of data is crucial (Acer et al. 2012).

For illustrative purposes, we focus on a bridge network with n edges and p paths from the source to the sink. Figure 1 describes an example of the network with 5 edges, $j = 1,\dots,5$, and 4 paths, denoted as $S_1 = \{1,5\}$, $S_2 = \{2,4\}$, $S_3 = \{1,3,4\}$, and $S_4 = \{2,3,5\}$, respectively. Suppose a traveler starting from the source wants to reach the sink before a deadline $\gamma > 0$ by choosing the shortest path. Let W_j be the random travel time along edge j. Following Acer et al. (2012), we assume that W_j follows a lognormal distribution with parameter μ_j and σ_j for $j = 1,\dots,n$ and that $W_1,\dots,W_n$ are mutually independent. Our objective is to estimate the probability of arriving on time, which is given by

$$\mathsf{P}\left(\min_{i=1,\dots,p}\sum_{j\in S_i} W_j < \gamma\right) = \mathsf{P}\left(\bigcup_{i=1,\dots,p}\left\{\log\left(\sum_{j\in S_i} e^{\mu_j+\sigma_j X_j}\right) < \log\gamma\right\}\right), \tag{15}$$

where $(X_1,\dots,X_n)$ is an n-dimensional standard normal random vector. When γ is small, the traveler has a tight schedule, and estimating (15) belongs to the realm of rare event simulation. Since the log-sum-exp function $f(\mathbf{x}) = \log(e^{x_1} + \cdots + e^{x_n})$ is convex, the target set of (15) is a union of convex sets, rendering Algorithm 1 feasible for estimation.

Numerical experiments under the basic setup. We use the example in Figure 1 to compare Algorithm 1 (CIS) with the crude Monte Carlo method (MC) and the traditional exponential twisting method (ET) in Section 5.2.2 of Bucklew (2004). We set $\mu_j = i/10$ and $\sigma_j = 1$ for $j = 1,\dots,5$ and run 10^8 iterations for each method. The results are reported in Table 2. As can be seen from the VR column of Table 2, CIS significantly outperforms all the other methods in terms of variance reduction. The variance reduction of ET is also notable but is much smaller than that of CIS for all γ. Based on the ER column, we conclude

Table 3: Estimates of δ and the robust bounds for the probability of arriving on time in a bridge network based on three different simulation methods.

Method	γ	$\widehat{\delta}$ (95% rel. err.)	$\widehat{\alpha}(\widehat{\delta})$ (95% rel. err.)	Time (sec)	VR	ER
MC-MC	0.1	1.48 (2.74%)	9.50×10^{-4} (35.09%)	554.0	–	–
	0.01	4.37 (1.14%)	3.90×10^{-4} (52.16%)	642.5	–	–
	0.001	7.49 (0.89%)	2.10×10^{-4} (78.11%)	720.2	–	–
MC-CIS	0.1	1.46 (3.77%)	9.01×10^{-4} (15.25%)	539.9	5.9	6.0
	0.01	4.33 (1.53%)	2.56×10^{-4} (13.23%)	625.9	36.2	37.2
	0.001	7.42 (2.81%)	1.46×10^{-4} (28.21%)	706.3	15.8	16.1
CIS-CIS	0.1	1.44 (0.10%)	8.43×10^{-4} (0.83%)	540.5	2288.3	2345.3
	0.01	4.32 (0.03%)	2.46×10^{-4} (0.56%)	627.3	21842.4	22371.2
	0.001	7.41 (0.03%)	1.39×10^{-4} (0.56%)	718.1	44615.2	44749.4

that despite the relatively longer simulation time, CIS is still significantly more efficient than the other two methods.

Numerical experiments under distributional uncertainty. We next consider the estimation of the robust bound for the probability of arriving on time in the same setting as above except that the distributions of travel times are now subject to uncertainty (Zhang et al. 2018). Following Section 5, we set $\varepsilon = 0.001$ and denote the robust bound by $\alpha(\delta)$.

To investigate the effectiveness of Algorithm 1, we compare three different cases: (1) replacing the CIS estimates in Steps 1 and 2, i.e., $\hat{\lambda}(\cdot)$ and $\hat{\alpha}(\cdot)$, with the associated crude Monte Carlo estimates (MC-MC), (2) replacing the CIS estimate $\hat{\lambda}(\cdot)$ in Step 1 with the associated crude Monte Carlo estimate and using the CIS estimate $\hat{\alpha}(\cdot)$ for Step 2 (MC-CIS), and (3) using the CIS estimates for both steps (CIS-CIS). For each case, we use 10^5 samples to construct each of the estimates $\widehat{\lambda}(\cdot)$ and $\widehat{\alpha}(\cdot)$. In the implementation, we estimate the derivatives $\lambda'(\delta)$ and $\alpha'(\delta)$ via the finite difference method (Fu 2006). To enhance the efficiency in estimating δ, we first use a pilot run to obtain a crude estimate of δ so as to update the target set $N_\delta(\mathscr{E})$, and then use all remaining samples to construct $\widehat{\delta}$ and its associated confidence interval. Note that this procedure shares the same vein with stochastic root-finding algorithms in the literature, e.g., cross entropy (Rubinstein and Kroese 2004) and adaptive importance sampling (He et al. 2023).

The corresponding results are reported in Table 3. From the VR column of Table 3, we observe that the overall variance reduction with the method of MC-CIS is relatively marginal, while the method of CIS-CIS attains significant variance reduction. This shows that reducing the estimation error in δ can have a critical impact on the overall error, and Algorithm 1 is effective in this regard. Comparing the computational times in Tables 2 and 3 reveals that all methods require much longer time in estimating $\alpha(\delta)$ than in estimating $\mathsf{P}(\mathbf{X}\in\mathscr{E})$ and that there is no significant difference in the time between the methods. This is mainly because evaluating whether the event $\{\mathbf{X}\in N_u(\mathscr{E})\}$ occurs requires solving the quadratic program $\min_{\mathbf{y}\in\mathscr{E}}\|\mathbf{X}-\mathbf{y}\|^2$ for each sample of $\mathbf{X}$.

7 CONCLUSION

We have proposed, analyzed, and tested a highly efficient conditional importance sampling method for estimating expectations defined on convex rare-event sets. It is theoretically proven that the algorithm has bounded relative error. This algorithm is widely applicable as convex target sets appear in various ways in the analysis of stochastic systems. Our numerical studies show that the proposed method significantly outperforms all methods in efficiency when the target event is extremely rare. Nonetheless, this paper is not without limitations and opens up several directions for further investigation. Firstly, relaxing the normality assumption would be an interesting direction for the development of rare-event simulation algorithms that are more generally applicable. Secondly, the performance of the proposed algorithm in gradient estimation is not investigated, which could be useful in stochastic optimization and in estimating option price sensitivities. Lastly, our method relies on the first-order approximation of the constraint functions. One may consider higher-order approximations and develop more sophisticated sampling methods for polynomially constrained target sets.

REFERENCES

Acer, U. G., P. Giaccone, D. Hay, G. Neglia, and S. Tarapiah. 2012. "Timely Data Delivery in a Realistic Bus Network". *IEEE Transactions on Vehicular Technology* 61(3):1251–1265.

Ahn, D., and K.-K. Kim. 2018. "Efficient Simulation for Expectations over the Union of Half-Spaces". *ACM Transactions on Modeling and Computer Simulation* 28(3):Article 23, 1–20.

Ahn, D., and L. Zheng. 2023. "Efficient Simulation of Polyhedral Expectations with Applications to Finance". Available at SSRN: https://ssrn.com/abstract=4447428, accessed 6th August 2023.

Asghari, M., T. Emrich, U. Demiryurek, and C. Shahabi. 2015. "Probabilistic Estimation of Link Travel Times in Dynamic Road Networks". In *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*, edited by M. Ali, Y. Huang, M. Gertz, M. Renz, and J. Sankaranarayanan, Article 47, 1–10. New York: Association for Computing Machinery.

Asmussen, S., and H. Albrecher. 2010. *Ruin Probabilities*. 2nd ed. Singapore: World Scientific.

Asmussen, S., and P. W. Glynn. 2007. *Stochastic Simulation: Algorithms and Analysis*. New York: Springer.

Bai, Y., S. He, H. Lam, G. Jiang, and M. C. Fu. 2022. "Importance Sampling for Rare-Event Gradient Estimation". In *Proceedings of the 2022 Winter Simulation Conference*, edited by B. Feng, G. Pedrielli, Y. Peng, S. Shashaani, E. Song, C. Corlu, L. Lee, E. Chew, T. Roeder, , and P. Lendermann, 3063–3074. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Bai, Y., Z. Huang, H. Lam, and D. Zhao. 2022. "Rare-Event Simulation for Neural Network and Random Forest Predictors". *ACM Transactions on Modeling and Computer Simulation* 32(3):Article 18, 1–33.

Bassamboo, A., S. Juneja, and A. Zeevi. 2008. "Portfolio Credit Risk with Extremal Dependence: Asymptotic Analysis and Efficient Simulation". *Operations Research* 56(3):593–606.

Bauer, D., A. Reuss, and D. Singer. 2012. "On the Calculation of the Solvency Capital Requirement Based on Nested Simulations". *ASTIN Bulletin: The Journal of the IAA* 42(2):453–499.

Bertsekas, D. P. 1999. *Nonlinear Programming*. 2nd ed. Belmont, Massachusetts: Athena Scientific.

Blanchet, J., and H. Lam. 2014. "Rare-Event Simulation for Many-Server Queues". *Mathematics of Operations Research* 39(4):1142–1178.

Blanchet, J., and K. Murthy. 2019. "Quantifying Distributional Model Risk via Optimal Transport". *Mathematics of Operations Research* 44(2):565–600.

Botev, Z., D. Mackinlay, and Y.-L. Chen. 2017. "Logarithmically Efficient Estimation of the Tail of the Multivariate Normal Distribution". In *Proceedings of the 2017 Winter Simulation Conference*, edited by W. K. V. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page, 1903–1913. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Botev, Z. I. 2017. "The Normal Law under Linear Restrictions: Simulation and Estimation via Minimax Tilting". *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 79(1):125–148.

Bucklew, J. A. 2004. *Introduction to Rare Event Simulation*. New York: Springer.

Bucklew, J. A. 2005. "Conditional Importance Sampling Estimators". *IEEE Transactions on Information Theory* 51(1):143–153.

Chen, Z., and P. Glasserman. 2008. "Fast Pricing of Basket Default Swaps". *Operations Research* 56(2):286–303.

Dupuis, P., A. D. Sezer, and H. Wang. 2007. "Dynamic Importance Sampling for Queueing Networks". *The Annals of Applied Probability* 17(4):1306–1346.

Dupuis, P., and H. Wang. 2004. "Importance Sampling, Large Deviations, and Differential Games". *Stochastics and Stochastics Reports* 76(6):481–508.

Fan, Y. Y., R. E. Kalaba, and J. E. Moore. 2005. "Arriving on Time". *Journal of Optimization Theory and Applications* 127(3):497–513.

Frank, H. 1969. "Shortest Paths in Probabilistic Graphs". *Operations Research* 17(4):583–599.

Fu, M. C. 2006. "Gradient Estimation". In *Handbooks in Operations Research and Management Science, Volume 13, Simulation*, edited by S. G. Henderson and B. L. Nelson, 575–616. Amsterdam: Elsevier.

Ghosh, S., and H. Lam. 2019. "Robust Analysis in Stochastic Simulation: Computation and Performance Guarantees". *Operations Research* 67(1):232–249.

Glasserman, P., P. Heidelberger, and P. Shahabuddin. 1999. "Asymptotically Optimal Importance Sampling and Stratification for Pricing Path-Dependent Options". *Mathematical Finance* 9(2):117–152.

Glasserman, P., P. Heidelberger, and P. Shahabuddin. 2000. "Variance Reduction Techniques for Estimating Value-at-Risk". *Management Science* 46(10):1349–1364.

Glasserman, P., and J. Li. 2005. "Importance Sampling for Portfolio Credit Risk". *Management Science* 51(11):1643–1656.

Glasserman, P., and X. Xu. 2014. "Robust Risk Measurement and Model Risk". *Quantitative Finance* 14(1):29–58.

Glynn, P. W., and D. L. Iglehart. 1989. "Importance Sampling for Stochastic Simulations". *Management Science* 35(11):1367–1392.

Glynn, P. W., and H. Lam. 2018. "Constructing Simulation Output Intervals under Input Uncertainty via Data Sectioning". In *Proceedings of the 2018 Winter Simulation Conference*, edited by M. Rabe, A. A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, 1551–1562. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Haraszti, Z., and J. K. Townsend. 1999. "Rare Event Simulation of Delay in Packet Switching Networks using DPR-Based Splitting". In *Proceedings of the 1999 Winter Simulation Conference*, edited by P. A. Farringtong, H. B. Nembhard, D. T. Sturrock, and G. W. Evans, 402–410. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

He, S., G. Jiang, H. Lam, and M. C. Fu. 2023. "Adaptive Importance Samping for Efficient Stochastic Root Finding and Quantile Estimation". *Operations Research* (Articles in Advance):1–19.

Heidelberger, P. 1995. "Fast Simulation of Rare Events in Queueing and Reliability Models". *ACM Transactions on Modeling and Computer Simulation* 5(1):43–85.

Huang, Z., H. Lam, and D. Zhao. 2018. "Rare-Event Simulation Without Structural Information: A Learning-Based Approach". In *Proceedings of the 2018 Winter Simulation Conference*, edited by M. Rabe, A. A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, 1826–1837. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Jaillet, P., J. Qi, and M. Sim. 2016. "Routing Optimization Under Uncertainty". *Operations Research* 64(1):186–200.

Juneja, S., and P. Shahabuddin. 2006. "Rare-Event Simulation Techniques: An Introduction and Recent Advances". In *Handbooks in Operations Research and Management Science, Volume 13, Simulation*, edited by S. G. Henderson and B. L. Nelson, 291–350. Amsterdam: Elsevier.

Kemna, A., and A. Vorst. 1990. "A Pricing Method for Options Based on Average Asset Values". *Journal of Banking & Finance* 14(1):113–129.

Kulkarni, V. G. 1986. "Shortest Paths in Networks with Exponentially Distributed Arc Lengths". *Networks* 16(3):255–274.

Lokshina, I., and M. R. Bartolacci. 2012. "Accelerated Rare Event Simulation with Markov Chain Modelling in Wireless Communication Networks". *International Journal of Mobile Network Design and Innovation* 4(4):185–191.

Ma, N., and W. Whitt. 2018. "A Rare-Event Simulation Algorithm for Periodic Single-Server Queues". *INFORMS Journal on Computing* 30(1):71–89.

O'Kelly, M., A. Sinha, H. Namkoong, J. Duchi, and R. Tedrake. 2018. "Scalable End-to-End Autonomous Vehicle Testing via Rare-Event Simulation". In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, edited by S. Bengio, H. M. Wallach, H. Larochelle, K. Grauman, and N. Cesa-Bianchi, 9849–9860. New York: Curran Associates Inc.

Owen, A. B., Y. Maximov, and M. Chertkov. 2019. "Importance Sampling the Union of Rare Events with an Application to Power Systems Analysis". *Electronic Journal of Statistics* 13(1):231–254.

Rockafellar, R. T. 1970. *Convex Analysis*. Princeton, New Jersey: Princeton University Press.

Rockafellar, R. T., and R. J.-B. Wets. 2009. *Variational Analysis*. 3rd ed. Berlin: Springer.

Rubinstein, R. Y., and D. P. Kroese. 2004. *The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation, and Machine Learning*. New York: Springer.

Serfling, R. J. 1980. *Approximation Theorems of Mathematical Statistics*. New York: John Wiley & Sons.

Shreve, S. E. 2004. *Stochastic Calculus for Finance II: Continuous-Time Models*. New York: Springer.

Siegmund, D. 1976. "Importance Sampling in the Monte Carlo Study of Sequential Tests". *The Annals of Statistics* 4(4):673–684.

Uryas'ev, S. 1994. "Derivatives of Probability Functions and Integrals Over Sets Given by Inequalities". *Journal of Computational and Applied Mathematics* 56:197–223.

Zhang, Y., S. Song, Z.-J. M. Shen, and C. Wu. 2018. "Robust Shortest Path Problem With Distributional Uncertainty". *IEEE Transactions on Intelligent Transportation Systems* 19(4):1080–1090.

AUTHOR BIOGRAPHIES

DOHYUN AHN is an assistant professor in the Department of Systems Engineering and Engineering Management at The Chinese University of Hong Kong. He received his B.S., M.S., and Ph.D. degrees in Industrial & Systems Engineering, all from KAIST. His research interests include, but are not limited to, quantitative risk management, stochastic simulation, decision making via uncertainty quantification, and network analysis in finance and operations. His email address is dohyun.ahn@cuhk.edu.hk.

LEWEN ZHENG is a Ph.D. candidate in the Department of Systems Engineering and Engineering Management at The Chinese University of Hong Kong. He received his B.S. degree from the School of Mathematics at Sun Yat-sen University in Guangzhou, China. His research interests include simulation methods and their applications to finance. His email address is lwzheng@se.cuhk.edu.hk.