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

CURSE OF DIMENSIONALITY IN RARE-EVENT SIMULATION

Yuanlu Bai
Antonius B. Dieker
Henry Lam

Industrial Engineering and
Operations Research
Columbia University
500 West 120th Street
New York, NY 10027, USA

ABSTRACT

In rare-event simulation, importance sampling (IS) is widely used to improve the efficiency of probability estimation. Asymptotic optimality is a common efficiency criterion, which requires that the relative error of the estimator only grows subexponentially in the rarity parameter. Most studies, however, consider low-dimensional problems and the effect of dimensionality is seldom analyzed. Motivated by recent AI-related applications, we take a first step towards high-dimensional rare-event simulation and demonstrate that for very simple examples, IS proposals that utilize exponential tilting, arguably the most common IS approach, can suffer from the "curse of dimensionality". That is, while the growth rate of the relative error is polynomial in the rarity parameter thus leading to asymptotic optimality, the degree of the polynomial depends on the problem dimensionality. Therefore, when the dimension is high, the relative error can be huge even in the rarity parameter regime where IS is conventionally believed to work well.

1 INTRODUCTION

Rare-event simulation aims to estimate quantities of interest associated with events with tiny probabilities. This problem is fundamental in risk assessment and management, and applies across a wide range of disciplines such as queueing systems (Dupuis et al. 2007; Dupuis and Wang 2009; Blanchet et al. 2009; Blanchet and Lam 2014; Kroese and Nicola 1999; Ridder 2009; Sadowsky 1991; Szechtman and Glynn 2002), reliability (Heidelberger 1995; Tuffin 2004; Nicola, Nakayama, Heidelberger, and Goyal 1993; Nicola, Shahabuddin, and Nakayama 2001), finance (Glasserman 2003; Glasserman and Li 2005; Glasserman et al. 2008) and insurance (Asmussen 1985; Asmussen and Albrecher 2010). The most straightforward method is to apply crude Monte Carlo (MC) simulation. That is, we simulate the original dynamics and use the frequency of hitting the target event to estimate the probability. However, by the very nature of rare events, the target event is seldom observed. Statistically, this results in a large estimation variance and in turn requires a large number of simulation replications to ensure acceptable estimation, thus rendering crude MC computationally expensive.

To address this challenge, various variance reduction techniques have been developed. These techniques alter or enhance the sampling scheme to reduce the variance per simulation run, and subsequently the required number of replications. Among these approaches, importance sampling (IS) (Juneja and Shahabuddin 2006; Rubino and Tuffin 2009) is arguably the most powerful for rare-event estimation. It applies a change of measure to the original distribution, which results in the IS distribution, to increase the frequency of hitting the target event. To maintain unbiasedness, the simulation output needs to be multiplied with the so-called likelihood ratio.

In the rare-event context, the efficiency of IS and variance reduction techniques more generally is measured by the relative error of the estimator, i.e., the ratio of the standard deviation (per simulation replication) to the true probability. Relative error is used because, as the target probability is very small, the estimation is meaningful only if the error is small relative to the magnitude of the target quantity. Via Chebyshev's inequality, it can be shown that the required number of replications to attain a prescribed level of overall relative error is proportional to the squared relative error, so the smaller the relative error, the more efficient the procedure. In particular, in typical large deviations settings, the target probability decays exponentially in a "rarity parameter". Correspondingly, crude MC can be readily shown to have an exponentially growing relative error. On the other hand, IS with suitably chosen IS distributions can result in the relative error growing subexponentially in the rarity parameter, thus providing a significant gain in efficiency. This type of IS achieves so-called *asymptotic optimality* (Juneja and Shahabuddin 2006), also known as asymptotic efficiency or logarithmic efficiency in the literature, which is a common efficiency criterion that requires the relative error per simulation replication to grow at most subexponentially.

While asymptotic optimality has been widely studied in the rare-event literature, most existing work focuses on low-dimensional problems without analyzing the effect of dimensionality. On the other hand, motivated by emerging AI-related applications in safety evaluation of intelligent systems and robustness quantification of machine learning models (Wang et al. 2021; Webb et al. 2018; Weng et al. 2019), rare-event simulation for high-dimensional problems becomes increasingly relevant (Bai et al. 2022; Bai et al. 2021). In this paper, we take a first step in this direction and investigate a simple setting to bring out the prevalence of the "curse of dimensionality" in this context. More precisely, we consider the probability of a Gaussian random vector hitting an infinite box in a finite-dimensional space. We show that IS distributions formed by exponential tilting (Siegmund 1976; Sadowsky and Bucklew 1990), arguably the commonest type of change of measure, give rise to a relative error that is polynomial in the rarity parameter, thus leading to asymptotic optimality. However, this polynomial has a degree that depends on the problem dimension. As a result, when the dimension is high, the relative error can be huge even for the rarity parameter regime where the IS is believed to work well. In fact, if we let the dimension grow with the rarity parameter, then the relative error could grow exponentially, thus violating the asymptotic optimality property commonly satisfied by exponential tilting. Such a behavior is in contrast with existing rare-event problems where the large rarity parameter is often the sole driver of the resulting tiny probability.

The rest of this paper is organized as follows. In Section 2, we review the background related to IS and exponential tilting. In Section 3, we demonstrate the curse of dimensionality in a Gaussian example using exponential tilting. In Section 4, we briefly summarize the key findings of our investigation and discuss some future directions.

2 BACKGROUND

2.1 Challenge of Crude MC

Suppose that we are interested in estimating $p = P(A)$ where A is a rare event, i.e., p is tiny. The crude MC estimator is defined by $\mathbb{I}(\omega \in A)$ where ω is sampled under the original measure P and $\mathbb{I}$ denotes the indicator function. If we generate n i.i.d. samples $\omega_1, \dots, \omega_n$, then we estimate p with $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(\omega_i \in A)$. Since p is tiny, we would like to ensure that the discrepancy between $\hat{p}$ and p is small relative to p itself. By Chebyshev's inequality, for any $\delta > 0$, we have that

$$P(|\hat{p} - p| > \delta p) \leq \frac{\mathrm{var}(\hat{p})}{\delta^2 p^2} = \frac{\mathrm{var}(\mathbb{I}(\omega \in A))}{n\delta^2 p^2}. \tag{1}$$

We define the relative error of an unbiased estimator as the ratio of the standard deviation per simulation run and the target quantity. Thus, in the last term, $\mathrm{var}(\mathbb{I}(\omega \in A))/p^2$ is the squared relative error for $\hat{p}$. Moreover, for any $\varepsilon > 0$, $n \geq \mathrm{var}(\mathbb{I}(\omega \in A))/(\varepsilon\delta^2 p^2)$ is a sufficient condition to guarantee that $P(|\hat{p} - p| > \delta p) \leq \varepsilon$, or in other words the required number of replications n to attain $P(|\hat{p} - p| > \delta p) \leq \varepsilon$ is proportional to the squared relative error. As $\mathbb{I}(\omega_1 \in A) \sim Bernoulli(p)$, we know that $\mathrm{var}(\mathbb{I}(\omega \in A)) = p(1-p)$, and hence $\mathrm{var}(\mathbb{I}(\omega \in A))/p^2 = (1-p)/p \approx 1/p$. This further implies that the required number of replications n to guarantee a high-confidence relative closeness between $\hat{p}$ and p is reciprocal in p. In this regard, the crude MC estimator is well known to be inefficient in the rare-event setting when p is tiny.

2.2 Importance Sampling and Asymptotic Optimality

As mentioned in the introduction, IS is a common approach to speed up the simulation. The key idea is to choose an alternative measure $\tilde{P}$ and use the IS estimator $Z = \mathbb{I}(\omega \in A)\frac{\mathrm{d}P}{\mathrm{d}\tilde{P}}(\omega)$ where $\frac{\mathrm{d}P}{\mathrm{d}\tilde{P}}$ is the Radon-Nikodym derivative, or likelihood ratio, between P and $\tilde{P}$. By this definition, the estimator Z is guaranteed to be unbiased regardless of the choice of $\tilde{P}$, as long as the estimator is well-defined in the sense of satisfying the requisite absolute continuity requirement. Note that, while unbiasedness is easy to obtain, low variance is not, and we still need to choose $\tilde{P}$ carefully to ensure a small variance or relative error of Z.

Now we suppose that the rare event A is parameterized by a rarity parameter γ, such that as $\gamma \to \infty$, A gets rarer and $p \to 0$. We will show later that in the classic large deviations setting, p decays exponentially in γ, which implies that the relative error of the crude MC estimator grows exponentially in γ. In contrast, we would like to find an IS estimator whose relative error only grows subexponentially in γ, which implies that the required simulation size is only subexponential following (1).

Formally, asymptotic optimality is defined as follows:

Definition 1 (Asymptotic Optimality) The IS estimator $Z = \mathbb{I}(\omega \in A)\frac{\mathrm{d}P}{\mathrm{d}\tilde{P}}(\omega)$ is said to be asymptotically optimal if $\lim_{\gamma\to\infty} \frac{\log(\tilde{E}(Z^2))}{\log p} = 2$ where $\tilde{E}(\cdot)$ denotes the expectation under $\tilde{P}$.

We note that $\tilde{E}(Z^2) \geq p^2$, so the limit condition $\lim_{\gamma\to\infty} \frac{\log(\tilde{E}(Z^2))}{\log p} = 2$ in Definition 1 cannot be improved. Intuitively, the condition implies that $\tilde{E}(Z^2)$ and p^2 have the same exponential rate of decay. In particular, when p is exponentially decaying in γ, the criterion of asymptotic optimality is equivalent to $\tilde{E}(Z^2)/p^2$, or the relative error of Z, growing at most subexponentially in γ.

2.3 Exponential Tilting

To achieve asymptotic optimality, exponential tilting is often employed. Here we explain the methodology with a standard Gartner-Ellis regime (Dembo and Zeitouni 2009) in the large deviations theory. Suppose that the rare event A is formulated as $A = \{\frac{1}{\gamma}X_\gamma \in \mathscr{E}\}$ where $\{X_\gamma\}_\gamma \subset \mathbb{R}^d$ are random vectors indexed by γ and $\mathscr{E} \subset \mathbb{R}^d$ is a fixed Borel set. Define $\mu_\gamma(x) = \frac{1}{\gamma}\log(e^{x^\top X_\gamma}), x \in \mathbb{R}^d$, which is the scaled cumulant generating function. We assume that $\mu(x) = \lim_{\gamma\to\infty}\mu_\gamma(x)$ exists for any $x \in \mathbb{R}^d$, where we allow ∞ both as a limit value and as an element of the sequence $\{\mu_\gamma(x)\}_\gamma$. Then we define the rate function as its Legendre transform, i.e., $I(y) = \sup_{x\in\mathbb{R}^d}\{x^\top y - \mu(x)\}, y \in \mathbb{R}^d$. Some general assumptions on the properties of μ, I and $\mathscr{E}$ are needed, but we do not list them in detail here as they are standard in the literature. Under the assumptions, the Gartner-Ellis theorem states that

$$-\lim_{\gamma\to\infty} \frac{1}{\gamma} P\left(\frac{1}{\gamma}X_\gamma \in \mathscr{E}\right) = \inf_{y\in\mathscr{E}} I(y) =: I(\mathscr{E}),$$

which implies that the true probability $p = P(A)$ decays exponentially in γ as long as $0 < I(\mathscr{E}) < \infty$. Intuitively, the rate function I measures the likelihood of hitting each point on the negative logarithm scale, and the exponential rate of decay of p is determined by the minimum rate (or equivalently, maximum likelihood) point over $\mathscr{E}$. Following this intuition, shifting the mean of the distribution to this minimum rate point potentially improves the efficiency, which is implemented via exponential tilting.

More specifically, let a denote the minimum rate point in $\mathscr{E}$, i.e., $a \in \overline{\mathscr{E}}$ and $I(a) = \inf_{y\in\mathscr{E}} I(y)$ where $\overline{\mathscr{E}}$ is the closure of $\mathscr{E}$, and assume that there exists a unique $s_a \in \mathbb{R}^d$ such that $\nabla\mu(s_a) = a$. Then the exponentially tilted IS measure $\tilde{P}$ is defined by

$$\frac{\mathrm{d}\tilde{P}}{\mathrm{d}P} = \exp\{s_a^\top X_\gamma - \gamma\mu_\gamma(s_a)\}. \tag{2}$$

It is well known that under certain conditions the IS estimator with this choice of $\tilde{P}$ is asymptotically optimal, and under more general conditions, mixing the distributions which are exponentially tilted to a set of points is asymptotically optimal (Sadowsky and Bucklew 1990; Dieker and Mandjes 2005). For simplicity, in this paper we will only consider the case where tilting to the minimum rate point is asymptotically optimal.

3 EXPONENTIAL TILTING IN HIGH DIMENSIONS

In the above discussion and also in most previous works, the effect of the dimension d is not analyzed. However, as high-dimensional problems appear in emerging AI-driven applications, it is increasingly important to understand whether the IS estimator is still "sufficiently efficient" in high dimensions. First, we consider the scenario where d is fixed but high. We still have that the growth rate of the relative error in γ is subexponential, but how would this growth rate depend on d? Say, if the relative error is polynomial in γ with a high degree when d is large, then in the finite-sample applications, the required simulation size can be prohibitively large, even though the relative error appears to be controlled in terms of γ. Second, we consider the scenario where the problem also scales with d. That is, we can view d as another rarity parameter. In this case, would the relative error still grow subexponentially in d? We note that in this case, as we allow the dimension d to change, the result inevitably depends on how exactly the problem evolves with d. While developing general results can be hard, it is easier to enforce a regime on the problem evolution.

3.1 Fixed but High Dimensionality

We analyze a specific example to demonstrate the curse of dimensionality for exponentially tilted IS. Suppose that $X_\gamma \sim N(\gamma\lambda, \gamma\Sigma)$ where $\mu \in \mathbb{R}^d$ and $\Sigma \in \mathbb{R}^{d\times d}$ is a positive definite matrix. Thus, $\frac{1}{\gamma}X_\gamma \sim N(\lambda, \frac{1}{\gamma}\Sigma)$, which becomes more concentrated around the fixed mean λ as γ grows. To ensure that the target event is rare, we assume that $\lambda \notin \overline{\mathscr{E}}$. For this Gaussian case, $\mu_\gamma(x) = \mu(x) = x^\top\lambda + \frac{1}{2}x^\top\Sigma x$ and $I(y) = \frac{1}{2}(y-\lambda)^\top\Sigma^{-1}(y-\lambda)$. We see a natural link between the rate function I and the density function of $N(\lambda, \frac{1}{\gamma}\Sigma)$. Besides, for any $a \in \mathbb{R}^d$, $s_a := \Sigma^{-1}(a-\lambda)$ is the unique solution to $\nabla\mu(s_a) = a$, and also the gradient of $I(y)$ at $y = a$. Following (2), under the exponentially tilted measure to point a, we have $\frac{1}{\gamma}X_\gamma \sim N(a, \frac{1}{\gamma}\Sigma)$. That is, we keep the covariance matrix and shift the mean from λ to a.

First, we derive a formula for the second moment of the IS estimator:

Lemma 1 Suppose we want to estimate $p = P(\frac{1}{\gamma}X_\gamma \in \mathscr{E})$. Here, $\gamma \in \mathbb{R}$ is the rarity parameter, $X_\gamma \sim N(\gamma\lambda, \gamma\Sigma)$, $\lambda \in \mathbb{R}^d$, $\Sigma \in \mathbb{R}^{d\times d}$ is positive definite and $\lambda \notin \overline{\mathscr{E}} \subset \mathbb{R}^d$. Consider IS estimator $Z = \mathbb{I}(\frac{1}{\gamma}X_\gamma \in \mathscr{E})\frac{dP}{d\tilde{P}}$ where $X_\gamma \sim N(\gamma a, \gamma\Sigma)$ under $\tilde{P}$ for some $a \in \mathbb{R}^d$. Then

$$\tilde{E}(Z^2) = \exp\{\gamma(a-\lambda)^\top\Sigma^{-1}(a-\lambda)\}\overline{P}\left(\frac{1}{\gamma}X_\gamma \in \mathscr{E}\right)$$

where $X_\gamma \sim N(\gamma(2\lambda - a), \gamma\Sigma)$ under $\overline{P}$.

Proof. We have that

$$Z = \mathbb{I}\left(\frac{1}{\gamma}X_\gamma \in \mathscr{E}\right)\frac{dP}{d\tilde{P}} = \mathbb{I}\left(\frac{1}{\gamma}X_\gamma \in \mathscr{E}\right)\exp\left\{-(X_\gamma - \gamma\lambda)^\top\Sigma^{-1}(a-\lambda) + \frac{\gamma}{2}(a-\lambda)^\top\Sigma^{-1}(a-\lambda)\right\}.$$

We thus find that

$$\begin{aligned}\tilde{E}(Z^2) =& \int_{x/\gamma\in\mathscr{E}} \exp\{-2(x-\gamma\lambda)^\top\Sigma^{-1}(a-\lambda)+\gamma(a-\lambda)^\top\Sigma^{-1}(a-\lambda)\}\times \\ & (2\pi)^{-d/2}|\gamma\Sigma|^{-1/2}\exp\left\{-\frac{1}{2\gamma}(x-\gamma a)^\top\Sigma^{-1}(x-\gamma a)\right\}\mathrm{d}x \\ =& \exp\{\gamma(a-\lambda)^\top\Sigma^{-1}(a-\lambda)\}\times \\ & \int_{x/\gamma\in\mathscr{E}}(2\pi)^{-d/2}|\gamma\Sigma|^{-1/2}\exp\left\{-\frac{1}{2\gamma}(x-\gamma(2\lambda-a))^\top\Sigma^{-1}(x-\gamma(2\lambda-a))\right\}\mathrm{d}x \\ =& \exp\{\gamma(a-\lambda)^\top\Sigma^{-1}(a-\lambda)\}\overline{P}\left(\frac{1}{\gamma}X_\gamma\in\mathscr{E}\right),\end{aligned}$$

as claimed. □

For simplicity, from now on we consider a special case where $\mathscr{E}=\{x\in\mathbb{R}^d : x\geq a\}$ for some $a>\lambda$. Here, $\geq$ and $>$ denote elementwise comparison. Clearly, a is the minimum rate point or maximum density point in $\mathscr{E}$, so we consider choosing the IS measure as the exponential tilting to the point a. As shown later in Theorem 1, the resulting IS estimator is asymptotically optimal when d is fixed. We note that this special case can be easily generalized to $\{x\in\mathbb{R}^d : Bx\geq b\}$ where B is a matrix (not necessarily square), $B\Sigma B^\top$ is positive definite and $b>B\lambda$. In the following lemma, we derive an expression for the probability that a Gaussian vector $\frac{1}{\gamma}X_\gamma$ falls in the set $\mathscr{E}$.

Lemma 2 Suppose we want to estimate $p=P(\frac{1}{\gamma}X_\gamma\geq a)$. Here, $\gamma\in\mathbb{R}$ is the rarity parameter, $X_\gamma\sim N(\gamma\lambda,\gamma\Sigma)$, $\lambda\in\mathbb{R}^d$, $\Sigma\in\mathbb{R}^{d\times d}$ is positive definite, $a\in\mathbb{R}^d$ and $a>\lambda$. Denote $s_i=e_i^\top\Sigma^{-1}(a-\lambda)$ where e_i denotes the i-th column of the $d\times d$ identity matrix. Then we have that

$$\begin{aligned}& P\left(\frac{1}{\gamma}X_\gamma\geq a\right) \\ =&(2\pi)^{-d/2}|\Sigma|^{-1/2}\gamma^{-d/2}\left(\prod_{i=1}^d s_i\right)^{-1}\exp\left\{-\frac{\gamma}{2}(a-\lambda)^\top\Sigma^{-1}(a-\lambda)\right\}\int_{x\geq 0}\exp\left\{-\frac{1}{2}\tilde{x}^\top\Sigma^{-1}\tilde{x}-x^\top\mathbf{1}\right\}\mathrm{d}x\end{aligned}$$

where $\tilde{x}$ denotes the vector such that $\tilde{x}_i=(\sqrt{\gamma}s_i)^{-1}x_i$ for any x and $\mathbf{1}$ denotes d-dimensional all-ones vector. Note that s_i is the i-th element of s_a.

Proof. Using the proof of Lemma 7.1 in Bai et al. (2022) (with $Y:=\frac{1}{\sqrt{\gamma}}(X_\gamma-\gamma\lambda), \tilde{\Sigma}:=\Sigma, \tilde{s}:=\sqrt{\gamma}(a-\lambda), y^*:=\sqrt{\gamma}(a-\lambda)$, $e_i^\top\Sigma^{-1}(a-\lambda)>0$ for any $i=1,\ldots,d$ and hence $I=I_1=\{1,\ldots,d\}$), we get this result. □

We note that although Lemma 2 is presented with $X_\gamma\sim N(\gamma\lambda,\gamma\Sigma)$, it can also be easily applied to the measure $\overline{P}$ in Lemma 1. Then by combining these results, we can analyze the relative error of the exponentially tilted IS estimator as in the following theorem:

Theorem 1 Suppose we want to estimate $p=P(\frac{1}{\gamma}X_\gamma\geq a)$. Here, $\gamma\in\mathbb{R}$ is the rarity parameter, $X_\gamma\sim N(\gamma\lambda,\gamma\Sigma)$, $\lambda\in\mathbb{R}^d$, $\Sigma\in\mathbb{R}^{d\times d}$ is positive definite, $a\in\mathbb{R}^d$ and $a>\lambda$. Denote $s_i=e_i^\top\Sigma^{-1}(a-\lambda)$ where e_i denotes the i-th column of the $d\times d$ identity matrix. Consider IS estimator $Z=\mathbb{I}(\frac{1}{\gamma}X_\gamma\geq a)\frac{dP}{d\tilde{P}}$ where $X_\gamma\sim N(\gamma a,\gamma\Sigma)$ under $\tilde{P}$. Then we have that

$$\tilde{E}(Z^2)/p^2=\gamma^{d/2}(\sqrt{2\pi}/2)^d|\Sigma|^{1/2}\left(\prod_{i=1}^d s_i\right)\frac{\int_{x\geq 0}\exp\{-\frac{1}{8}\tilde{x}^\top\Sigma^{-1}\tilde{x}-x^\top\mathbf{1}\}\mathrm{d}x}{(\int_{x\geq 0}\exp\{-\frac{1}{2}\tilde{x}^\top\Sigma^{-1}\tilde{x}-x^\top\mathbf{1}\}\mathrm{d}x)^2}$$

where $\tilde{x}$ denotes the vector such that $\tilde{x}_i = (\sqrt{\gamma} s_i)^{-1} x_i$ for any x and $\mathbf{1}$ denotes d-dimensional all-ones vector. Suppose λ_{max} and λ_{min} are the maximum and minimum eigenvalues of Σ^{-1}. Then we have that

$$\tilde{E}(Z^2)/p^2 \le \gamma^{d/2}(\sqrt{2\pi}/2)^d |\Sigma|^{1/2} \left(\prod_{i=1}^d s_i\right) \left(\prod_{i=1}^d \frac{\gamma s_i^2}{\gamma s_i^2 + \lambda_{max}}\right)^{-2}$$

and

$$\tilde{E}(Z^2)/p^2 \ge \gamma^{d/2}(\sqrt{2\pi}/2)^d |\Sigma|^{1/2} \left(\prod_{i=1}^d s_i\right) \left(\prod_{i=1}^d \frac{4\gamma s_i^2}{4\gamma s_i^2 + \lambda_{max}}\right).$$

Proof. From Lemma 1, we have that

$$\tilde{E}(Z^2)/p^2 = \exp\{\gamma(a-\lambda)^\top \Sigma^{-1}(a-\lambda)\}\overline{P}\left(\frac{1}{\gamma}X_\gamma \ge a\right)/P^2\left(\frac{1}{\gamma}X_\gamma \ge a\right).$$

Using lemma 2, we have

$$\begin{aligned} &P\left(\frac{1}{\gamma}X_\gamma \ge a\right) \\ =&(2\pi)^{-d/2}|\Sigma|^{-1/2}\gamma^{-d/2}\left(\prod_{i=1}^d s_i\right)^{-1} \exp\left\{-\frac{\gamma}{2}(a-\lambda)^\top\Sigma^{-1}(a-\lambda)\right\}\int_{x\ge 0}\exp\left\{-\frac{1}{2}\tilde{x}^\top\Sigma^{-1}\tilde{x} - x^\top\mathbf{1}\right\}\mathrm{d}x. \end{aligned}$$

Since $a > \lambda$, we also have $a > 2\lambda - a$. Thus, by applying Lemma 2 similarly (replacing λ with $2\lambda - a$) we get that

$$\begin{aligned} &\overline{P}\left(\frac{1}{\gamma}X_\gamma \ge a\right) \\ =&(2\pi)^{-d/2}|\Sigma|^{-1/2}\gamma^{-d/2}\left(\prod_{i=1}^d 2s_i\right)^{-1} \exp\left\{-2\gamma(a-\lambda)^\top\Sigma^{-1}(a-\lambda)\right\}\int_{x\ge 0}\exp\left\{-\frac{1}{8}\tilde{x}^\top\Sigma^{-1}\tilde{x} - x^\top\mathbf{1}\right\}\mathrm{d}x. \end{aligned}$$

Thus, we get that

$$\tilde{E}(Z^2)/p^2 = \gamma^{d/2}(\sqrt{2\pi}/2)^d|\Sigma|^{1/2}\left(\prod_{i=1}^d s_i\right)\frac{\int_{x\ge 0}\exp\{-\frac{1}{8}\tilde{x}^\top\Sigma^{-1}\tilde{x} - x^\top\mathbf{1}\}\mathrm{d}x}{(\int_{x\ge 0}\exp\{-\frac{1}{2}\tilde{x}^\top\Sigma^{-1}\tilde{x} - x^\top\mathbf{1}\}\mathrm{d}x)^2}.$$

First, we know that

$$\begin{aligned}
\int_{x\geq 0}\exp\left\{-\frac{1}{2}\tilde{x}^\top\Sigma^{-1}\tilde{x}-x^\top\mathbf{1}\right\}\mathrm{d}x &\leq \int_{x\geq 0}\exp\left\{-\frac{1}{2}\lambda_{min}\tilde{x}^\top\tilde{x}-x^\top\mathbf{1}\right\}\mathrm{d}x\\
&= \int_{x\geq 0}\exp\left\{\sum_{i=1}^{d}\left(-\frac{\lambda_{min}}{2\gamma s_i^2}x_i^2-x_i\right)\right\}\mathrm{d}x\\
&= \prod_{i=1}^{d}\int_0^\infty \exp\left\{-\frac{\lambda_{min}}{2\gamma s_i^2}\left(x_i+\frac{\gamma s_i^2}{\lambda_{min}}\right)^2+\frac{\gamma s_i^2}{2\lambda_{min}}\right\}\mathrm{d}x_i\\
&= \prod_{i=1}^{d}\exp\left\{\frac{\gamma s_i^2}{2\lambda_{min}}\right\}\sqrt{\frac{2\pi\gamma s_i^2}{\lambda_{min}}}\bar{\Phi}\left(\sqrt{\frac{\gamma s_i^2}{\lambda_{min}}}\right)\\
&\leq \prod_{i=1}^{d}\exp\left\{\frac{\gamma s_i^2}{2\lambda_{min}}\right\}\sqrt{\frac{2\pi\gamma s_i^2}{\lambda_{min}}}\frac{1}{\sqrt{\frac{\gamma s_i^2}{\lambda_{min}}}}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{\gamma s_i^2}{2\lambda_{min}}\right\}\\
&= 1,
\end{aligned}$$

where $\bar{\Phi}$ denotes the tail distribution function of the standard normal distribution. Next, we know that

$$\begin{aligned}
\int_{x\geq 0}\exp\left\{-\frac{1}{2}\tilde{x}^\top\Sigma^{-1}\tilde{x}-x^\top\mathbf{1}\right\}\mathrm{d}x &\geq \int_{x\geq 0}\exp\left\{-\frac{1}{2}\lambda_{max}\tilde{x}^\top\tilde{x}-x^\top\mathbf{1}\right\}\mathrm{d}x\\
&= \int_{x\geq 0}\exp\left\{\sum_{i=1}^{d}\left(-\frac{\lambda_{max}}{2\gamma s_i^2}x_i^2-x_i\right)\right\}\mathrm{d}x\\
&= \prod_{i=1}^{d}\int_0^\infty \exp\left\{-\frac{\lambda_{max}}{2\gamma s_i^2}\left(x_i+\frac{\gamma s_i^2}{\lambda_{max}}\right)^2+\frac{\gamma s_i^2}{2\lambda_{max}}\right\}\mathrm{d}x_i\\
&= \prod_{i=1}^{d}\exp\left\{\frac{\gamma s_i^2}{2\lambda_{max}}\right\}\sqrt{\frac{2\pi\gamma s_i^2}{\lambda_{max}}}\bar{\Phi}\left(\sqrt{\frac{\gamma s_i^2}{\lambda_{max}}}\right)\\
&\geq \prod_{i=1}^{d}\exp\left\{\frac{\gamma s_i^2}{2\lambda_{max}}\right\}\sqrt{\frac{2\pi\gamma s_i^2}{\lambda_{max}}}\frac{\sqrt{\frac{\gamma s_i^2}{\lambda_{max}}}}{\frac{\gamma s_i^2}{\lambda_{max}}+1}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{\gamma s_i^2}{2\lambda_{max}}\right\}\\
&= \prod_{i=1}^{d}\frac{\gamma s_i^2}{\gamma s_i^2+\lambda_{max}}.
\end{aligned}$$

Similarly, we get that

$$\prod_{i=1}^{d}\frac{4\gamma s_i^2}{4\gamma s_i^2+\lambda_{max}} \leq \int_{x\geq 0}\exp\left\{-\frac{1}{8}\tilde{x}^\top\Sigma^{-1}\tilde{x}-x^\top\mathbf{1}\right\}\mathrm{d}x \leq 1.$$

Therefore we derive the bounds in the theorem. □

When d is fixed and $\gamma\to\infty$, in Theorem 1, we note that $\lambda,\Sigma,a,s_i,\lambda_{max},\lambda_{min}$ are all constant values. First, from the proof, we get that the large deviations probability p itself satisfies

$$p \leq (2\pi)^{-d/2}|\Sigma|^{-1/2}\gamma^{-d/2}\left(\prod_{i=1}^{d}s_i\right)^{-1}\exp\left\{-\frac{\gamma}{2}(a-\lambda)^\top\Sigma^{-1}(a-\lambda)\right\}$$

and

$$p \geq (2\pi)^{-d/2}|\Sigma|^{-1/2}\gamma^{-d/2}\left(\prod_{i=1}^{d} s_i\right)^{-1} \exp\left\{-\frac{\gamma}{2}(a-\lambda)^\top\Sigma^{-1}(a-\lambda)\right\}\prod_{i=1}^{d}\frac{\gamma s_i^2}{\gamma s_i^2+\lambda_{max}}.$$

As d is fixed and $\gamma \to \infty$, we see that p decays exponentially in γ as stated in the Gartner-Ellis theorem, but in addition to the exponential term, there is also a power rate $\gamma^{-d/2}$, so it decays even faster when d is larger. Moreover, the lower and upper bounds of $\tilde{E}(Z^2)/p^2$ both grow in the order of $\gamma^{d/2}$, which is polynomial in γ. When d is large, although this growth rate is still subexponential, it could be fast in terms of γ.

3.2 When Dimension Grows with the Rarity Parameter

Furthermore, if we allow the dimension d to change, the situation would be more complicated to handle, as the relationship between the relative error and the dimension d also depends on how the problem parameters λ, Σ, a change with d. To obtain some insights, we consider the following simple example:

Example 1 Consider the setting in Theorem 1 with $\lambda = \mathbf{0}$, Σ is the identity matrix, and $a = x\mathbf{1}$ for some $x > 0$. In this case, we actually have that

$$p = \bar{\Phi}^d(\sqrt{\gamma x^2})$$

and by applying Lemma 1, we have that

$$\tilde{E}(Z^2) = \exp\{\gamma a^\top a\}P^d(N(-x,1/\gamma) \geq x) = \left(\exp\{\gamma x^2\}\bar{\Phi}(2\sqrt{\gamma x^2})\right)^d.$$

Therefore, we get that

$$\tilde{E}(Z^2)/p^2 = \left(\frac{\exp\{\gamma x^2\}\bar{\Phi}(2\sqrt{\gamma x^2})}{\bar{\Phi}^2(\sqrt{\gamma x^2})}\right)^d.$$

As mentioned before, when d is fixed and $\gamma \to \infty$, we have that

$$p = (2\pi)^{-d/2}\gamma^{-d/2}x^{-d}e^{-dx^2\gamma/2}(1+o(1))$$

and that

$$\tilde{E}(Z^2)/p^2 = \left(\frac{\sqrt{2\pi\gamma}x}{2}\right)^d(1+o(1)).$$

That is, the decay rate of p in γ also depends on d, and the growth rate of $\tilde{E}(Z^2)/p^2$ is $\gamma^{d/2}$, which is fast when d is large even though it is not exponential. When γ is fixed and $d \to \infty$, p decays exponentially in d. Moreover, the relative error is also exponential in d, and the growth rate depends on the values of γ and x. Even when γ and x are very close to 0, we still have that

$$\tilde{E}(Z^2)/p^2 = \left(\frac{\exp\{\gamma x^2\}\bar{\Phi}(2\sqrt{\gamma x^2})}{\bar{\Phi}^2(\sqrt{\gamma x^2})}\right)^d \approx \left(\frac{\exp(0)\bar{\Phi}(0)}{\bar{\Phi}^2(0)}\right)^d = 2^d.$$

Specifically, we have that

$$\lim_{d\to\infty}\frac{\log\tilde{E}(Z^2)}{\log p} = \lim_{d\to\infty}\frac{d\gamma x^2 + d\log\bar{\Phi}(2\sqrt{\gamma x^2})}{d\log\bar{\Phi}(\sqrt{\gamma x^2})} = \frac{\gamma x^2+\log\bar{\Phi}(2\sqrt{\gamma x^2})}{\log\bar{\Phi}(\sqrt{\gamma x^2})},$$

which can be strictly less than 2 (e.g., consider $\gamma, x \approx 0$) and violate the "asymptotic optimality" condition. In fact, under the IS measure, the probability of hitting the rare-event set is 2^{-d}, which is small. In this sense, the curse of dimensionality arises in this example.

4 CONCLUSION AND DISCUSSION

In this paper, we mainly analyze a specific example to understand how the problem dimensionality affects the efficiency of the "asymptotically optimal" exponentially tilted IS estimator. In the literature, the estimator is often considered efficient if the relative error grows only subexponentially in the rarity parameter. However, we demonstrate that in the high-dimensional case, even though the growth rate is polynomial in the rarity parameter, this polynomial has a high degree, so the resulting relative error can still be large and the IS is inefficient. Moreover, even if we let the problem scale with the dimension in a reasonable way, the relative error potentially grows exponentially in the dimension.

We plan to pursue several future directions. First, we would like to extend this investigation to other setups than box-type rare-event sets to understand how general are the insights gained from this paper. Second, while AI-driven problems serve as a general motivation for our investigation, we aim to find more concrete applications where high dimensionality as discussed in this paper is indeed the driving rarity factor. Third, we will study whether other types of estimators, such as strongly efficient estimators, can remain efficient in high dimensions. Fourth, given that dimensionality can now demonstratably deteriorate efficiency in addition to the rarity parameter, it would be interesting to obtain estimators that can balance both in some suitable sense. That is, a future goal is to obtain efficient rare-event estimators that can remedy the curse of dimensionality discovered in this paper.

ACKNOWLEDGMENTS

We gratefully acknowledge support from the InnoHK initiative, the Government of the HKSAR, Laboratory for AI-Powered Financial Technologies, and the Columbia SEAS Innovation Hub Research Grant.

.

REFERENCES

Asmussen, S. 1985. "Conjugate Processes and the Simulation of Ruin Problems". *Stochastic Processes and Their Applications* 20(2):213–229.

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

Bai, Y., Z. Huang, H. Lam, and D. Zhao. 2021. "Over-Conservativeness of Variance-Based Efficiency Criteria and Probabilistic Efficiency in Rare-Event Simulation". To appear in *Management Science*, arXiv preprint arXiv:2110.12573, accessed 20th April.

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):1–33.

Blanchet, J., P. Glynn, and H. Lam. 2009. "Rare Event Simulation for a Slotted Time $M/G/s$ Model". *Queueing Systems* 63(1):33–57.

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

Dembo, A., and O. Zeitouni. 2009. *Large Deviations Techniques and Applications*. 2nd ed. Germany: Springer Science & Business Media.

Dieker, A., and M. Mandjes. 2005. "On Asymptotically Efficient Simulation of Large Deviation Probabilities". *Advances in Applied Probability* 37(2):539–552.

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. 2009. "Importance Sampling for Jackson Networks". *Queueing Systems* 62(1):113–157.

Glasserman, P. 2003. *Monte Carlo Methods in Financial Engineering*. New York: Springer.

Glasserman, P., W. Kang, and P. Shahabuddin. 2008. "Fast Simulation of Multifactor Portfolio Credit Risk". *Operations Research* 56(5):1200–1217.

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

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

Juneja, S., and P. Shahabuddin. 2006. "Rare-Event Simulation Techniques: An Introduction and Recent Advances". *Handbooks in Operations Research and Management Science* 13:291–350.

Kroese, D. P., and V. F. Nicola. 1999. "Efficient Estimation of Overflow Probabilities in Queues with Breakdowns". *Performance Evaluation* 36:471–484.

Nicola, V. F., M. K. Nakayama, P. Heidelberger, and A. Goyal. 1993. "Fast Simulation of Highly Dependable Systems with General Failure and Repair Processes". *IEEE Transactions on Computers* 42(12):1440–1452.

Nicola, V. F., P. Shahabuddin, and M. K. Nakayama. 2001. "Techniques for Fast Simulation of Models of Highly Dependable Systems". *IEEE Transactions on Reliability* 50(3):246–264.

Ridder, A. 2009. "Importance Sampling Algorithms for First Passage Time Probabilities in the Infinite Server Queue". *European Journal of Operational Research* 199(1):176–186.

Rubino, G., and B. Tuffin. 2009. *Rare Event Simulation Using Monte Carlo Methods*. United Kingdom: John Wiley & Sons.

Sadowsky, J. S. 1991. "Large Deviations Theory and Efficient Simulation of Excessive Backlogs in a $GI/GI/m$ Queue". *IEEE Transactions on Automatic Control* 36(12):1383–1394.

Sadowsky, J. S., and J. A. Bucklew. 1990. "On Large Deviations Theory and Asymptotically Efficient Monte Carlo Estimation". *IEEE transactions on Information Theory* 36(3):579–588.

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

Szechtman, R., and P. W. Glynn. 2002. "Rare-Event Simulation for Infinite Server Queues". In *Proceedings of the 2002 Winter Simulation Conference*, edited by E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes, 416–423. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Tuffin, B. 2004. "On Numerical Problems in Simulations of Highly Reliable Markovian Systems". In *Proceedings of First International Conference on the Quantitative Evaluation of Systems, 2004.*, 156–164. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Wang, B., S. Webb, and T. Rainforth. 2021. "Statistically Robust Neural Network Classification". In *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence*, edited by C. de Campos and M. H. Maathuis, 1735–1745. Cambridge, Massachusetts: Proceedings of Machine Learning Research.

Webb, S., T. Rainforth, Y. W. Teh, and M. P. Kumar. 2018. "A Statistical Approach to Assessing Neural Network Robustness". arXiv preprint arXiv:1811.07209, accessed on 20$^{\text{th}}$ April.

Weng, L., P.-Y. Chen, L. Nguyen, M. Squillante, A. Boopathy, I. Oseledets, and L. Daniel. 2019. "PROVEN: Verifying Robustness of Neural Networks with a Probabilistic Approach". In *Proceedings of the 36th International Conference on Machine Learning*, edited by K. Chaudhuri and R. Salakhutdinov, 6727–6736. Cambridge, Massachusetts: Proceedings of Machine Learning Research.

AUTHOR BIOGRAPHIES

YUANLU BAI is a PhD student in the Department of IEOR at Columbia University. She received an M.S. in operations research from Columbia University, a B.S. in statistics and a B. Ec. in economics from Peking University. Her research interests are uncertainty quantification and rare-event simulation. Her email address is yb2436@columbia.edu.

ANTONIUS B. DIEKER is an Associate Professor in the Department of Industrial Engineering and Operations Research at Columbia University. His research interests include Monte Carlo methods and computational tools for stochastic modeling. His email address is ton.dieker@ieor.columbia.edu and his website is http://www.columbia.edu/~ad3217/.

HENRY LAM is an Associate Professor in the Department of Industrial Engineering and Operations Research at Columbia University. His research interests include Monte Carlo methods, uncertainty quantification, risk analysis and data-driven optimization. He serves as the area editor for Stochastic Models and Data Science in Operations Research Letters, and on the editorial boards of Operations Research, INFORMS Journal on Computing, Applied Probability Journals, Stochastic Models, Manufacturing and Service Operations Management, and Queueing Systems. His email address is henry.lam@columbia.edu and his website is http://www.columbia.edu/~khl2114/.