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

GENERALIZED IMPORTANCE SAMPLING FOR NESTED SIMULATION

Qingyuan Amber Chen

Department of Operations Research and
Information Engineering
Cornell University
206 Hoy Rd
Ithaca, NY 14853, USA

Mingbin Ben Feng

Department of Statistics and Actuarial Science
University of Waterloo
200 University Ave W
Waterloo, ON N2L 3G1, CANADA

ABSTRACT

Importance sampling (IS) is a classical variance reduction technique. Under mild conditions, an IS estimator is unbiased, so one often seeks the variance-minimizing optimal sampling distribution. IS has remarkable success in many applications, such as engineering, operations research, and finance. In some applications, such as enterprise risk management and input uncertainty quantification, complex simulation designs, such as nested simulation, arise naturally. The outer-level simulation generates a set of risk factors, i.e., the scenarios, which are used as inputs for inner-level simulations. Nested simulation leads to wasteful use of computations, as inner simulation outputs in each scenario are isolated from other scenarios. In this study, we propose, analyze, and test a generalized importance sampling technique for nested simulation. Our generalized IS approach reuses one set of inner simulation outputs across different outer scenarios. Numerical experiments show that our proposal is orders of magnitude more efficient than the standard procedure.

1 INTRODUCTION

In this paper, we develop a generalization of the classical variance reduction technique, importance sampling (IS). As an application of this generalization, we propose an efficient nested simulation procedure for estimating risk measures of the form:

$$\rho = \rho\left(\mu(\theta)\right), \quad \mu(\theta) = \mathbb{E}\left[g(X)|\theta\right], \tag{1}$$

where θ is a random variable that characterizes the conditional distribution of $X|\theta$. We refer to θ and $X|\theta$ as the outer scenarios and inner random variables given scenario θ, respectively. Additionally, $g(x)$ is a given function representing a simulation logic that transforms a sample x to a simulation output. Finally, ρ is a functional that maps the random variable $\mu(\theta)$ to a real value. Note that $\mu(\theta)$ is a random variable because θ is random. Problems with the above structure are known as *nested estimation* (Hong et al. 2017), which is common in financial and operations research applications such as enterprise risk management (Lee 1998; Gordy and Juneja 2010) and input uncertainty quantification (Cheng and Holland 1997; Barton 2012; Zhu et al. 2020).

Designing computationally efficient nested simulation procedures has attracted much research attention in the last two decades. Lee (1998), Lee and Glynn (2003), Gordy and Juneja (2010) analyzed how to minimize the mean square error (MSE) of the nested risk estimator by strategic budget allocation in standard nested simulation; the resulting MSE vanishes at an optimal rate of $\mathcal{O}(\Gamma^{-2/3})$, where Γ is the total simulation budget in a standard nested simulation procedure. Broadie et al. (2011) proposed a sequential nested simulation algorithm whose MSE converges at the rate $\mathcal{O}(\Gamma^{-4/5+\varepsilon})$ for any $\varepsilon > 0$. Broadie et al. (2015) proposed a regression-based approach with MSE converges at $\mathcal{O}(\Gamma^{-1+\delta})$ for any $\delta > 0$ to

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

a non-zero bias level. Hong et al. (2017) proposed a kernel smoothing approach with MSE converges at $\mathcal{O}(\Gamma^{-\min\{1,4/(d+2)\}})$. Recently, Zhang et al. (2022) proposed a sample recycling nested simulation procedure where all scenarios reuse the same set of inner replications. Specifically, the proposal therein is that the user should simulate *one* set of sample paths from a common sampling distribution, then reuse the same sample paths via the likelihood ratio method when estimating the conditional expectations of all scenarios. The likelihood ratio method is well-known in the literature, see Beckman and McKay (1987), L'Ecuyer (1990), and Rubinstein and Shapiro (1993), for example. Recently it is used for repeated simulation experiments and termed green simulation by Feng and Staum (2015) and Feng and Staum (2017). Under some assumptions, they show that the resulting nested risk measure's MSE converges at $\mathcal{O}(\Gamma^{-1})$, which is an impressively fast convergence rate matching that of a non-nested Monte Carlo simulation. While Zhang et al. (2022) showed that sample recycling via the likelihood ratio method is a promising approach for designing efficient nested simulation procedures, they did not specifically address how to select a good sampling distribution. We propose one solution to fill this gap in this paper.

Importance sampling (IS) is a classical variance reduction method, see Chapter 9 in Owen (2013) and Chapter 4.6 in Glasserman (2013), for example. In a Monte Carlo simulation experiment where one ought to generate random samples from a target distribution (also known as the nominal distribution), IS attempts to reduce variance by identifying an alternate sampling distribution (also known as the proposal distribution) from which random samples are generated. Under mild conditions, such as absolute continuity of the target distribution with respect to the sampling distribution, the IS estimator is unbiased. An optimal sampling distribution is identified by minimizing the variance of the IS estimator.

In this paper, we generalize the classical IS to identify an optimal sampling distribution, whose random samples are reused in different scenarios in a nested simulation procedure. In contrast to classical IS, which has *one* target distribution, a nested estimation problem has *many* target distributions, e.g., one for each outer scenario. We propose to identify a common sampling distribution for all the scenarios by minimizing the average variance of m given scenarios (or the integrated MSE for given outer scenario distribution). The optimal sampling distribution is derived in closed-form, but it has limited practical use. We further propose a weighted mixture approximation method based on non-negative least squares (NNLS) regression. Our numerical experiments show that the proposed procedure is more efficient and more accurate than the standard nested simulation procedure.

The rest of this paper is organized as follows. Section 2 reviews the nested simulation problem and classical importance sampling. In Section 3 we derive the generalized importance sampling and its application in a sample-recycling nested simulation procedure. Section 4 shows two practical implementations to generate random samples from the optimal sampling distribution or from its approximation. We perform numerical experiments and summarize the results and findings in Section 5. Section 6 concludes.

2 PROBLEM SETTINGS

2.1 Standard Nested Simulation

In this section we elaborate on the *nested estimation* problem (1) and the standard nested simulation procedure. For notational convenience, we define $\pi(\theta)$ as the outer scenario's distribution with support Θ. For any scenario $\theta \in \Theta$, we denote the conditional distribution of $X|\theta$ by $p(x|\theta)$ with corresponding support $\mathscr{X}_\theta$. We are interested in estimating

$$\rho_\pi = \rho\left(\mu(\theta)\right) = \rho\left(\mathbb{E}[g(X)|\theta]\right), \tag{2}$$

where $\theta \sim \pi(\theta)$. Given a scenario θ, we define the conditional expectation $\mu(\theta) = \mathbb{E}[g(X)|\theta] = \int_{\mathscr{X}} g(x)p(x|\theta)dx$ for that scenario. The support $\mathscr{X}$ is defined in Assumption 1, which we assume holds in this study for ease of exposure.

Assumption 1 The supports Θ and $\mathscr{X}_\theta$ and the conditional distribution $p(x|\theta)$ satisfy the following.

(i) For any $\theta \in \Theta$, $p(x|\theta)$ is a well-defined probability density function (pdf).
(ii) For all $\theta \in \Theta$, $p(x|\theta)$ has a common support $\mathscr{X}$, i.e., $\mathscr{X}_\theta = \mathscr{X}, \forall \theta \in \Theta$.
(iii) For any $\theta \in \Theta$, the user can calculate $p(x|\theta)$ for all $x \in \mathscr{X}$ and can generate samples from $p(x|\theta)$.

All three assumptions are necessary for the likelihood ratio method to be applicable in our problem setting. Admittedly, Assumption (iii), particularly its first half, is arguably strong thus limits the applicability of the proposed method. Nonetheless, in some applications, such as dynamic asset models in finance and queueing models in manufacturing, the input distributions of the simulation models are known to users. We focus of such applications in this paper and will explore possible relaxations in future research.

As a motivating example, consider a portfolio of complex financial derivatives whose values are affected by underlying risk factors such as stock prices, bond prices, interest rates, exchange rates, etc. Let T represent the portfolio's maturity, which is the longest maturity among its constituent assets. A portfolio manager may want to measure the risk associated with changes in the portfolio's value due to market fluctuations, i.e., the portfolio's profit and loss, between time 0 (now) and a future time $\tau > 0$ called the risk horizon (e.g., 3 months from now). Such a risk measurement problem is naturally formulated as a nested estimation problem. At the outer level, m independent and identically distributed (i.i.d.) scenarios $\theta_1, \ldots, \theta_m$ are simulated to represent the plausible evolution of the underlying risk factors from now to the risk horizon τ. These scenarios are simulated under the real-world probability measure, and their uncertainties propagate to the corresponding portfolio values. Under some assumptions, such as a complete market, the portfolio values in each scenario can be calculated as the (conditional) expected discounted payoff under the risk-neutral probability measure, i.e., $\mathbb{E}[g(X)|\theta_i], i = 1, \ldots, m$. In this setting, θ represents the evolution of the underlying risk factor from the risk horizon τ to the portfolio maturity T, and $g(X)$ is the portfolio's discounted payoff given the risk factors. When the portfolio payoff is complex, this conditional expectation cannot be calculated analytically, so an inner-level simulation is required to estimate it. Specifically, for each scenario θ_i, n i.i.d. inner replications $X_{i1}, \ldots, X_{in} \sim p(x|\theta_i)$ are simulated to estimate $\mathbb{E}[g(X)|\theta = \theta_i]$ by some estimator $\widehat{\mu}_i$, $i = 1, \ldots, m$. Finally, the risk measure ρ, such as variance, Value-at-Risk (VaR), and Conditional VaR (CVaR), etc., can be estimated based on the empirical distribution of $\widehat{\mu}_1, \widehat{\mu}_2, \ldots, \widehat{\mu}_m$.

Standard nested simulation is a natural approach for solving the nested estimation problem. Specifically, one first simulates m i.i.d. outer scenarios $\theta_1, \ldots, \theta_m \sim \pi(\theta)$. For each scenario θ_i, $i = 1, \ldots, m$, $\mu_i = \mu(\theta_i)$ is estimated by the standard MC estimator based on n i.i.d. inner replications. That is,

$$\widehat{\mu}_i = \widehat{\mu}(\theta_i) = \frac{1}{n} \sum_{j=1}^{n} g(X_{ij}), \quad X_{ij} \overset{i.i.d.}{\sim} p(x|\theta_i), \forall j = 1, \ldots, n.$$

The risk measure ρ in (2) can then be estimated based on the empirical distribution of $\widehat{\mu}_1, \ldots, \widehat{\mu}_m$. For example, the unconditional expectation $\mathbb{E}[\mu(\theta)]$ and variance $\mathbb{V}ar[\mu(\theta)]$ can be estimated by the sample average and sample variance, respectively. Also, one can sort the i.i.d. sample to obtained the order statistics $\widehat{\mu}_{(1)} \leq \widehat{\mu}_{(2)} \leq \cdots \leq \widehat{\mu}_{(m)}$. For a give confidence level $0 < \alpha < 100\%$, say $\alpha = 95\%$, and suppose αm is an integer for simplicity. Then the α-VaR of $\mu(\theta)$ can be estimated by the sample quantile $\widehat{\rho} = \widehat{\mu}_{(\alpha m)}$ and the α-CVaR can be estimated by the tail average $\widehat{\rho} = \frac{1}{(1-\alpha)m} \sum_{i=\alpha m+1}^{m} \widehat{\mu}_{(i)}$.

A major criticism of the standard nested simulation procedure is the high computational cost. Clearly, a standard nested simulation experiment requires a *simulation budget* of $\Gamma = n \times m$ inner replications. That is, there are a total of Γ inner replications, n times in each of the m outer scenarios, and the inner simulation model $g(x)$ is evaluated Γ times. To estimate ρ accurately, both m and n need to be sufficiently large: A large m captures the outer scenarios' stochasticity, which is crucial for representing the variability in $\mu(\theta)$. Also, a large n is needed for *every* outer scenario for an accurate estimate $\widehat{\mu}_i$ of μ_i. Practical applications may require complex stochastic models for outer scenarios θ and inner replications $X|\theta$ as well as complicated inner simulation model $g(x)$. In risk management applications, for instance, compromising accuracy to reduce the computational burden is not an acceptable solution due to the severe financial consequences.

As a result, the standard nested simulation procedure can be unbearably slow. In Section 3, we propose a more efficient nested simulation procedure that recycles and reuses simulation outputs.

2.2 Classical Importance Sampling

In this section we review the classical importance sampling. Specifically, we provide a constructive review to show how the optimal sampling distribution is derived using calculus of variation.

Consider a stochastic simulation experiment that aims to estimate $\mu = \mathbb{E}_p[g(X)] = \int_{\mathscr{X}} g(x)p(x)dx$ for some simulation model $g(x)$ and target pdf $p(x)$. The standard MC estimator for μ is $\widehat{\mu} = \frac{1}{n}\sum_{j=1}^{n} g(X_j)$ where $X_j \sim p(x)$, $\forall j = 1,\ldots,n$. Let $q(x)$ be any other pdf that satisfies $q(x) > 0$ whenever $g(x)p(x) \neq 0$. Note that this condition is trivially satisfied if $p(x)$ and $q(x)$ have the same support. Then, it is clear that

$$\mu = \int_{\mathscr{X}} g(x)p(x)dx = \int_{\mathscr{X}} g(x)\frac{p(x)}{q(x)}q(x)dx = \mathbb{E}_q\left[g(X)\frac{p(X)}{q(X)}\right]. \tag{3}$$

Equation (3) suggests that one can estimate μ via the importance sampling (IS) estimator

$$\widetilde{\mu} = \frac{1}{n}\sum_{j=1}^{n} g(\widetilde{X}_j)\cdot\frac{p(\widetilde{X}_j)}{q(\widetilde{X}_j)}, \quad \widetilde{X}_j \overset{i.i.d.}{\sim} q(x), \forall j = 1,\ldots,n. \tag{4}$$

As one draws random samples from $q(x)$ to estimate an expectation with respect to $p(x)$, we call $p(x)$ and $q(x)$ the *target distribution* and the *sampling distribution*, respectively. It is well-known that $\widetilde{\mu}$ is unbiased and its variance is $\mathbb{V}ar_q[\widetilde{\mu}] = \frac{1}{n}\widetilde{\sigma}_q^2$ where

$$\widetilde{\sigma}_q^2 = \mathbb{V}ar_q\left[g(X)\frac{p(X)}{q(X)}\right] = \mathbb{E}_q\left[\left(g(X)\frac{p(X)}{q(X)}\right)^2\right] - \mu^2. \tag{5}$$

The classical IS identifies the variance-minimizing sampling distribution, i.e., $q^*(x)$, such that the resulting IS estimator variance, i.e., $\widetilde{\sigma}_{q^*}^2$ in (5), is minimized. Note that, because the IS estimator is unbiased, a variance-minimizing sampling density is also MSE-minimizing.

One way to identify $q^*(x)$ is by *calculus of variation* for optimizing functionals. A complete and formal development of calculus of variation and functional derivatives is beyond the scope of this study, interested readers are referred to Gelfand and Fomin (2000). Relevant definitions and optimality conditions are summarized in Lemma 1.

Lemma 1 Let $q(x)$ be a real-valued function and let $L(y,z)$ be a differentiable function. Consider the following optimization problem

$$\min_{q(x)} \qquad J[q] = \int_{\mathscr{X}} L(x,q(x))dx \tag{6a}$$

$$\text{subject to} \qquad \int_{\mathscr{X}} q(x)dx = 1. \tag{6b}$$

A necessary condition for $q^*(x)$ to be a minimizer for (6) is that there exists $\lambda \in \mathbb{R}$ such that

$$\left.\frac{\partial L(y,z)}{\partial z}\right|_{(y,z)=(x,q^*(x))} + \lambda = 0. \tag{7}$$

A sufficient condition for $q^*(x)$ to be a minimizer for (6) is that $\left.\frac{\partial^2 L(y,z)}{\partial z^2}\right|_{(y,z)=(x,q^*(x))} \geq 0, \forall x \in \mathscr{X}$.

We note that (7) is the well-known Euler-Lagrange equation for constrained functional optimization problems. One can use Lemma 1 to construct an optimal sampling distribution $q^*(x)$ that minimizes the IS estimator variance $\widetilde{\sigma}_q^2$ in (5), as summarized in Lemma 2.

Lemma 2 An optimal sampling distribution $q^*(x)$ that minimizes the variance and the MSE of the IS estimator (4) is $q^*(x) = \lambda^*|g(x)|p(x)$, where $\lambda^* = (\int_{\mathscr{X}} |g(x)|p(x)dx)^{-1}$. The minimized variance is $\widetilde{\sigma}^2_{q^*} = (\mathbb{E}_p[|g(X)|])^2 - \mu^2$. Moreover, the variance reduction of $q^*(x)$ compared to any other sampling distribution $q(x)$ is $\widetilde{\sigma}^2_q - \widetilde{\sigma}^2_{q^*} = \mathbb{V}ar_q\left[|g(X)|\frac{p(X)}{q(X)}\right] \geq 0$.

Notably, when $g(x) \geq 0$ or $g(x) \leq 0$ for all $x \in \mathscr{X}$, the optimal IS estimator has zero variance. But this is difficult to achieve because calculating the normalizing constant λ^* is as difficult as calculating μ. Despite this difficulty, importance sampling remains a popular and useful variance reduction technique.

The optimal sampling distribution $q^*(x) \propto |g(x)|p(x)$ is a well-known conclusion in importance sampling. We nonetheless present Lemmas 1 and 2 for the constructive mathematical derivation of $q^*(x)$. In Section 3, we generalize the classical IS to cases with many target distributions and derive analogous results to Lemma 2.

3 GENERALIZED IMPORTANCE SAMPLING FOR NESTED SIMULATION

In this section we generalize classical IS multiple target distributions then use it to develop an efficient nested simulation procedure.

Before proceeding to further discussions, we distinguish our research focus from those of Feng and Staum (2015) and Zhang et al. (2022) despite methodological similarities, i.e., sample reuse via the likelihood ratio method. Green simulation (Feng and Staum 2015) aims to recycle and reuse simulation outputs in repeated simulations. The reused simulation outputs were generated in the past so the user does not choose the sampling distribution. Zhang et al. (2022) analyzed sample recycling in nested simulation and showed that, under some assumptions, the resulting risk estimator's MSE has fast convergence rate. They stated that "in practical applications usually there is a natural choice of sampling distribution" but did not consider how to find a good sampling distribution in general. In contrast, this study specifically considers the construction of provably optimal sampling distribution in the context of nested simulation.

For ease of exposure, we consider a setting where m outer scenarios, $\theta_1, \ldots, \theta_m$ had been simulated and we need to estimate the corresponding conditional expectations $\mu_i = \mu(\theta_i) = \mathbb{E}[g(X)|\theta_i]$, $i = 1, \ldots, m$. Recall that standard nested simulation only uses a small fraction, e.g., $\frac{1}{m}$, of the total simulation budget to estimate each conditional expectation $\mu(\theta_i)$ as it ignores simulation outputs from other scenarios $\theta_j \neq \theta_i$. We believe one can significantly improve the estimation efficiency and accuracy by recycling and reusing all the inner simulation outputs to estimate every conditional expectation $\mu(\theta_i)$.

Specifically, we propose to generate one set of random samples from a common sampling distribution, say $\widetilde{X}_j \sim q(x)$, $\forall j = 1, \ldots, \Gamma$ then estimate every scenario's conditional expectation $\mu(\theta_i)$ by the corresponding IS estimator

$$\widetilde{\mu}_i = \frac{1}{\Gamma}\sum_{j=1}^{\Gamma} g(\widetilde{X}_j) \cdot \frac{p(\widetilde{X}_j|\theta_i)}{q(\widetilde{X}_j)}, \quad \forall i = 1, \ldots, m. \tag{8}$$

To reiterate, the innovation in (8) is that, for different outer scenarios θ_i, $i = 1, \ldots, m$, we propose to use reuse *one set of random samples* generated form a *common sampling distribution* $q(x)$ to estimate μ_i by the corresponding IS estimator $\widetilde{\mu}_i$. We refer to the above sample recycling nested simulation procedure as the generalized importance sampling (GIS) and refer to $\widetilde{\mu}_i$'s as the GIS estimators.

The remaining question is then: *How to identify a good sampling distribution* $q(x)$? In classical IS, one assesses the goodness of a sampling distribution by the variance or the mean squared error (MSE) of the IS estimator. Inspired by classical IS, with m target distributions, we aim to minimize the average MSE (AMES) (or equivalently the average variances) of the m GIS estimators $\widetilde{\mu}_1, \ldots, \widetilde{\mu}_m$. Based on (5), the MSE for $\widetilde{\mu}_i$ is given by $MSE[\widetilde{\mu}_i] = \frac{1}{n}\widetilde{\sigma}^2_{i,q}$ where $\widetilde{\sigma}^2_{i,q} = \mathbb{E}_q\left[\left(g(X)\frac{p(X|\theta_i)}{q(X)}\right)^2\right] - \mu_i^2$. Therefore, if one generates Γ inner replications from a sampling distribution q, the AMSE of the m GIS estimators is given

by

$$AMSE_q = \frac{1}{m}\sum_{i=1}^{m}\frac{\widetilde{\sigma}_{i,q}^2}{\Gamma} = \frac{1}{m\Gamma}\sum_{i=1}^{m}\int_{\mathscr{X}}\frac{g^2(x)p^2(x|\theta_i)}{q(x)}dx - \frac{1}{m\Gamma}\sum_{i=1}^{m}\mu_i^2. \tag{9}$$

We note that $\frac{1}{m\Gamma}\sum_{i=1}^{m}\mu_i^2$ does not depend on the sampling distribution $q(x)$. So we can minimize the AMSE by minimizing $\frac{1}{m\Gamma}\sum_{i=1}^{m}\int_{\mathscr{X}}\frac{g^2(x)p^2(x|\theta_i)}{q(x)}dx = \frac{1}{m\Gamma}\int_{\mathscr{X}}\frac{g^2(x)\left[\sum_{i=1}^{m}p^2(x|\theta_i)\right]}{q(x)}dx$. Specifically, we formalize an optimization problem as follows:

$$\min_{q(x)} \quad J[q] = \int_{\mathscr{X}} L(x,q(x))dx = \int_{\mathscr{X}}\frac{g^2(x)\left[\sum_{i=1}^{m}p^2(x|\theta_i)\right]}{q(x)}dx \tag{10a}$$

$$\text{subject to} \quad \int_{\mathscr{X}} q(x)dx = 1. \tag{10b}$$

Note that, for problem (10), $\left.\frac{\partial L(y,z)}{\partial z}\right|_{(y,z)=(x,q^*(x))} = -\frac{g^2(x)\sum_{i=1}^{m}p^2(x|\theta_i)}{(q^*(x))^2}$. So, by Lemma 1, the optimal sampling distribution $q^*(x)$ for (10) must satisfy $-\frac{g^2(x)\sum_{i=1}^{m}p^2(x|\theta_i)}{(q^*(x))^2}+\lambda = 0 \Rightarrow q^*(x) \propto |g(x)|\sqrt{\sum_{i=1}^{m}p^2(x|\theta_i)}$ for some $\lambda = \lambda^*$. Respecting the constraint $\int_{\mathscr{X}} q(x)dx = 1$, we find that the normalizing constant is given by $\lambda^* = \left(\int_{\mathscr{X}}|g(x)|\sqrt{\sum_{i=1}^{m}p^2(x|\theta_i)}dx\right)^{-1}$. Checking the sufficient optimality condition, i.e., the second partial $\frac{\partial^2 L(x,q^*(x))}{\partial z^2} = \frac{2g^2(x)\sum_{i=1}^{m}p^2(x|\theta_i)}{(q^*(x))^3} \geq 0$ for all $x \in \mathscr{X}$ so $q^*(x)$ is indeed an optimal sampling distribution for (10).

Proposition 1 summarizes the optimal sampling distribution $q^*(x)$, the minimized AMSE, and the reduction in AMSE compared to other GIS estimators. Note that, in addition to Assumption 1, some technical assumptions are needed so that the normalizing constant λ^* is well-defined. For ease of reference, we refer to the GIS estimators that use the optimal sampling distribution $q^*(x)$ the OIS estimators.

Proposition 1 Consider given scenarios $\theta_1,\ldots,\theta_m$. Suppose Assumption 1 holds, $0 < \mathbb{E}[|g(X)||\theta_k]$ for some $k = 1,\ldots,m$, and $\mathbb{E}[|g(X)||\theta_i] < \infty$ for all $i = 1,\ldots,m$. Then the optimal sampling distribution $q^*(x)$ that minimizes the AMSE in (9) is

$$q^*(x) = \lambda^* \cdot |g(x)|\sqrt{\sum_{i=1}^{m}p^2(x|\theta_i)}, \tag{11}$$

where $\lambda^* = \left(\int_{\mathscr{X}}|g(x)|\sqrt{\sum_{i=1}^{m}p^2(x|\theta_i)}dx\right)^{-1}$. The minimized AMSE is

$$AMSE_{OIS} = \frac{1}{m\Gamma}\left[\left(\int_{\mathscr{X}}|g(x)|\sqrt{\sum_{i=1}^{m}p^2(x|\theta_i)}dx\right)^2 - \sum_{i=1}^{m}\mu_i^2\right].$$

Compared to any other sampling distribution $q(x)$, the optimal $q^*(x)$ reduces the AMSE by

$$AMSE_{OIS} = AMSE_q - \frac{1}{\Gamma}\mathbb{V}ar_q\left[\frac{|g(x)|\sqrt{\frac{1}{m}\sum_{i=1}^{m}p^2(x|\theta_i)}}{q(x)}\right]. \tag{12}$$

We first observe that Proposition 1 is indeed a generalization of Lemma 2, as the former coincides with the latter when $m = 1$. Secondly, the variance reduction (12) compares all possible sampling distributions applied in a sample recycling nested simulation procedure to the OIS estimators. This includes the so-called

mixture likelihood ratio (MLR) estimators (Feng and Staum 2017), which is also known as the balance heuristic (Veach 1997). In this article's terminology, the MLR estimators are GIS estimators whose proposal distribution is the mixture distribution $\bar{p}(x)$ given by

$$\bar{p}(x) = \frac{1}{m}\sum_{i=1}^{m} p(x|\theta_i). \tag{13}$$

The MLR estimators have been numerically demonstrated to good performances in different applications such as financial risk management and Monte Carlo rendering. Proposition 1, specifically (12), shows that the OIS estimator's AMSE is no greater than that of the MLR estimator's. We note that both OIS and MLR use a simulation budget of Γ replications from the respective sampling distributions.

Next, we compare the OIS estimators to the MC estimators used in standard nested simulation. One can show that the AMSE for the standard nested simulation estimators is

$$AMSE_{STD} = \frac{1}{m}\sum_{i=1}^{m}\frac{\mathbb{E}[g^2(X)|\theta_i]-\mu_i^2}{n} = \frac{1}{n}\int_{\mathscr{X}} g^2(x)\bar{p}(x)dx - \frac{1}{mn}\sum_{i=1}^{m}\mu_i^2.$$

Proposition 2 Consider any given scenarios $\theta_1,\ldots,\theta_m \in \theta$. If the conditions in Proposition 1 hold then

$$AMSE_{OIS} \le AMSE_{STD} + \frac{1}{n}\left((m-1)\left(E_{\bar{p}}[|g(x)|]\right)^2 - \mathbb{V}ar_{\bar{p}}\left[|g(x)|\right]\right). \tag{14}$$

In this comparison, the simulation budgets for $AMSE_{OIS}$ and $AMSE_{STD}$ are $\Gamma_{OIS} = n$ and $\Gamma_{STD} = mn$, respectively. Despite significant difference in simulation budgets, the difference between the two *AMSEs* are small. Specifically, in (14), the terms inside the brackets can be negative thus it is not a guaranteed variance reduction. Nonetheless, for any fixed m (however large m is), the difference between $AMSE_{OIS}$ and $AMSE_{STD}$ converges to 0 as $n \to \infty$ even though the former's simulation budget is only $1/m$ of the latter's. This showcases the high simulation efficiency of OIS estimators.

4 PRACTICAL IMPLEMENTATION: NON-NEGATIVE LEAST SQUARE (NNLS) HEURISTIC

Like classical IS, GIS is not practical because it is difficult to generate samples from $q^*(x)$. For instance, if one wishes to use the inverse-cdf method to generate random samples, one needs to integrate $q^*(x)$, which can be difficult due to its dependence on the simulation model $g(x)$. Though we know $q^*(x) \propto |g(x)|\sqrt{\frac{1}{m}\sum_{i=1}^{m} p(x|\theta_i)}$ from (11), calculating the normalizing constant $\lambda^* = (\int_{\mathscr{X}} |g(x)|\sqrt{\frac{1}{m}\sum_{i=1}^{m} p(x|\theta_i)}dx)^{-1}$ is more difficult than calculating $\mu_i = \int_{\mathscr{X}} g(x)\frac{1}{m}\sum_{i=1}^{m} p(x|\theta_i)dx$ directly.

In this section, we propose a practical implementation to generate random samples from a weighted mixture distribution that is similar to $q^*(x)$; similarity will be defined later.

Consider a mixture distribution of the form

$$q_{\boldsymbol{\beta}}(x) = \sum_{i=1}^{m}\beta_i p(x|\theta_i), \quad \text{where } \beta_i \ge 0 \text{ and } \sum_{i=1}^{m}\beta_i = 1. \tag{15}$$

We propose to use $q_{\boldsymbol{\beta}}(x)$, for some mixture weights $\boldsymbol{\beta}$, to approximate *the shape* of $q^*(x)$, i.e., to approximate a function that is proportional to $q^*(x)$. For ease of reference, we define $q^*_\gamma(x) = \gamma|g(x)|\sqrt{\frac{1}{m}\sum_{i=1}^{m} p^2(x|\theta_i)}$ where proportionality constant $\gamma > 0$ is a free variable that we will optimize over. Approximating only the shape of $q^*(x)$ allows some degrees of variance reduction while circumventing the unknown normalizing constant λ^*. We choose the mixture weights $\boldsymbol{\beta}$ that minimizes the average expected squared distance between $q_{\boldsymbol{\beta}}(x)$ and $q^*_\gamma(x)$, which is given by

$$\frac{1}{m}\sum_{i=1}^{m}\mathbb{E}[\|q_{\boldsymbol{\beta}}(X) - q^*_\gamma(X)\|_2^2|\theta_i] = \frac{1}{m}\sum_{i=1}^{m}\int_{\mathscr{X}}(q_{\boldsymbol{\beta}}(x) - q^*_\gamma(x))^2 p(x|\theta_i)dx = \mathbb{E}_{\bar{p}}[\|q_{\boldsymbol{\beta}}(X) - q^*_\gamma(X)\|_2^2], \tag{16}$$

where $\bar{p}(x)$ is as defined in (13). We can then choose the mixture weights $\boldsymbol{\beta}$ by solving a sample estimate of (16), i.e., $(\boldsymbol{\beta}^*, \gamma^*) = \arg\min_{\boldsymbol{\beta}^* \geq 0, \gamma^*} \left\{ \sum_{j=1}^{\Gamma_1} \left(q_{\boldsymbol{\beta}}(X_j) - q^*_{\gamma}(X_j) \right)^2 \middle| \sum_{i=1}^{m} \beta_i = 1 \right\}$ where $X_j \sim \bar{p}(x)$ for $j = 1, \ldots, \Gamma_1$. Recognizing that $\gamma \in \mathbb{R}$, we can reformulate the problem as a NNLS regression:

$$\min_{\widetilde{\boldsymbol{\beta}} \geq 0} \left\{ \sum_{j=1}^{\Gamma_1} \left(\sum_{i=1}^{m} \widetilde{\beta}_i p(X_j|\theta_i) - |g(X_j)| \sqrt{\frac{1}{m} \sum_{i=1}^{m} p^2(X_j|\theta_i)} \right)^2 \right\}. \tag{17}$$

For any optimal weight $\widetilde{\boldsymbol{\beta}}^*$ for (17) that satisfies $\sum_{i=1}^{m} \widetilde{\beta}^*_i > 0$, one can show that $\left(\boldsymbol{\beta}^* = \frac{\widetilde{\beta}^*}{\sum_{i=1}^{m} \widetilde{\beta}^*_i}, \gamma^* = \frac{1}{\sum_{i=1}^{m} \widetilde{\beta}^*_i} \right)$ is optimal for the original problem. Summarizing the above discussions, Algorithm 1 presents a two-stage GIS nested simulation procedure. Note that the split of the simulation budget, i.e., Γ_1 and Γ_2, is the user's choice. In our numerical experiments, we find that increasing Γ_1 in Stage 1 can significantly increase the fitting time but has little improvement in Stage 2's estimation quality. So, we recommend allocating only a small fraction of the budget, e.g., 1% to Stage 1 and majority of the budget to Stage 2.

Algorithm 1: GIS nested simulation procedure using NNLS heuristic

Input: Given scenarios $\theta_1, \ldots, \theta_m$. Total simulation budget Γ. Stage 1 simulation budget Γ_1. Stage 2 simulation budget $\Gamma_2 = \Gamma - \Gamma_1$.

Output: GIS estimators $\widetilde{\mu}_1, \ldots, \widetilde{\mu}_m$ for $\mu_1, \ldots, \mu_m$.

Stage I: Fit a mixture distribution via NNLS.

Simulate $X_j \sim \bar{p}(x)$, $j = 1, \ldots, \Gamma_1$;

Solve (17) via NNLS for $\widetilde{\boldsymbol{\beta}}$ then normalize the weight $\boldsymbol{\beta}^* = \frac{\widetilde{\boldsymbol{\beta}}^*}{\sum_{i=1}^{m} \widetilde{\beta}^*_i}$.

Stage II: Construct GIS estimators with Γ_2 samples from $q_{\boldsymbol{\beta}^*}(x)$.

Simulate $\widetilde{X}_j \overset{i.i.d.}{\sim} q_{\boldsymbol{\beta}^*}(x)$ then calculate and store simulation outputs $g(\widetilde{X}_j)$ for $j = 1, \ldots, \Gamma_2$;

Calculate and store conditional pdfs $p\left(\widetilde{X}_j|\theta_i\right)$ for all $j = 1, \ldots, \Gamma_2$ and $i = 1, \ldots, m$;

Retrieve the stored values to estimate the conditional expectation μ_i by

$$\widetilde{\mu}_i = \frac{1}{\Gamma_2} \sum_{j=1}^{\Gamma_2} g(\widetilde{X}_j) \frac{p\left(\widetilde{X}_j|\theta_i\right)}{q_{\boldsymbol{\beta}^*}(\widetilde{X}_j)}, \text{ for } i = 1, \ldots m.$$

Lastly, when sampling from $\bar{p}(x)$ and $q_{\boldsymbol{\beta}^*}(x)$, we propose using stratified sampling to further reduce variance. That is, we generate Γ_1/m replications from each component distribution in Stage I and $\beta_i \Gamma_2$ replications from the ith component distribution, $i = 1, \ldots, m$, in Stage II, provided that these values are integers. As demonstrated by Hesterberg (1988), Feng and Staum (2017), using stratified sampling when sampling from mixture distributions improves estimation accuracy and numerical stability.

5 NUMERICAL EXPERIMENTS

To illustrate the effectiveness of the proposed GIS procedure, we consider a reverse iron butterfly option portfolio and estimate its expected profits and losses (P&Ls) in 6-month using nested simulation. We compare and contrast three nested simulation procedures:the standard nested simulation procedure (STD), GIS procedure with mixture sampling distribution (MLR), and GIS procedure with NNLS implementation of the optimally weighted mixture distribution (OIS).

We consider a reverse iron butterfly option portfolio that consists of 4 vanilla European options written on the same underlying stock and have the same expiry: 1 short call with strike \$125, 1 long call with

strike \$145, 1 long put with strike \$145, and 1 short put with strike \$165. We assume that these options mature in $T = 1$ year and we are interested in estimating risk measures of this option portfolio's P&L in 6 months, i.e., $\tau = 0.5$. Denote the time-T stock price by S_T, the option portfolio's time-T payoff is

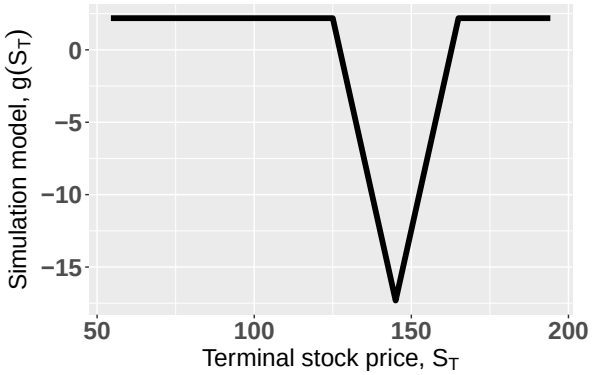
$$P(S_T) = -(S_T - 125)^+ + (S_T - 145)^+ + (145 - S_T)^+ - (165 - S_T)^+.$$

We model the underlying stock's price by the Geometric Brownian Motion. We assume that the asset has an initial price $S_0 = \$100$. Also, for any two times $t_2 > t_1 \geq 0$, $S_{t_2}|S_{t_1}$ is log-normally distributed, i.e.,

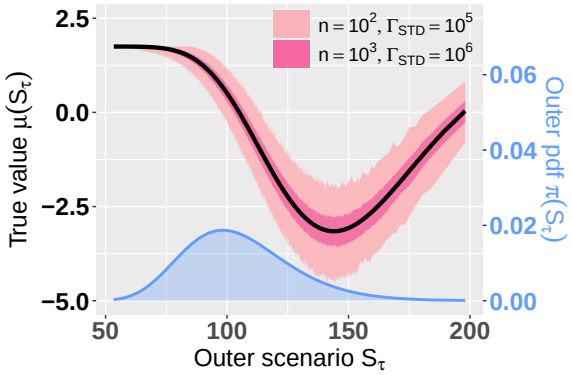
$$S_{t_2} = S_{t_1} \exp\left(\left(\mu - \frac{\sigma^2}{2}\right)(t_2 - t_1) + \sigma\sqrt{t_2 - t_1}Z\right), \quad Z \sim \mathcal{N}(0,1),$$

where the annualized volatility $\sigma = 30\%$ and the annualized expected return $\mu = 10\%$ under the real-world measure $\mathbb{P}$ and equals to the risk-free rate $r = 5\%$ under the risk-neutral measure $\mathbb{Q}$. For these modeling parameters, the initial price of the reverse iron butterfly portfolio is $P_0 = \mathbb{E}^{\mathbb{Q}}[e^{-rT}P(S_T)|S_0] = \$17.32$. For a given time-$\tau$ stock price S_τ, the portfolio P&L at time-τ is $\mu(S_\tau) = e^{-r(T-\tau)}\mathbb{E}^{\mathbb{Q}}[P(S_T)|S_\tau] - P_0 = \mathbb{E}^{\mathbb{Q}}[g(S_T)|S_\tau]$, where $g(S_T) = e^{-r(T-\tau)}P(S_T) - P_0$. At time 0, $\mu(S_\tau)$ is a random variable because S_τ is random. Under our settings, for any scenario S_τ we can calculate $\mu(S_\tau)$ analytically using Black-Scholes option pricing formulas. Nonetheless, we run inner simulations to estimate $\mu(S_\tau)$ in order to examine the performances of the different nested simulation procedures.

Figure 1 depicts the problem setting for this example: Figure 1a plots the simulation model $g(S_T)$ as a function of the terminal stock price S_T. By design, the small value of $g(S_T)$ is at $S_T = 145$. The top solid line in Figure 1b shows the conditional expectation $\mu(S_\tau)$ as a function of the scenario S_τ, which has a similar shape as $g(S_T)$ but smoother. The bottom bell-shaped shaded area shows the outer scenario's distribution $\pi(S_\tau)$. We see that the dip of $\mu(S_\tau)$ and the hump of $\pi(S_\tau)$ are off phase. This design helps us compare the MLR estimators and OIS estimators, as discussed later.



(a) Simulation logic $g(S_T)$ for option portfolio example.



(b) Outer pdf $\pi(s_\tau)$ and conditional expectation $\mu(s_\tau)$

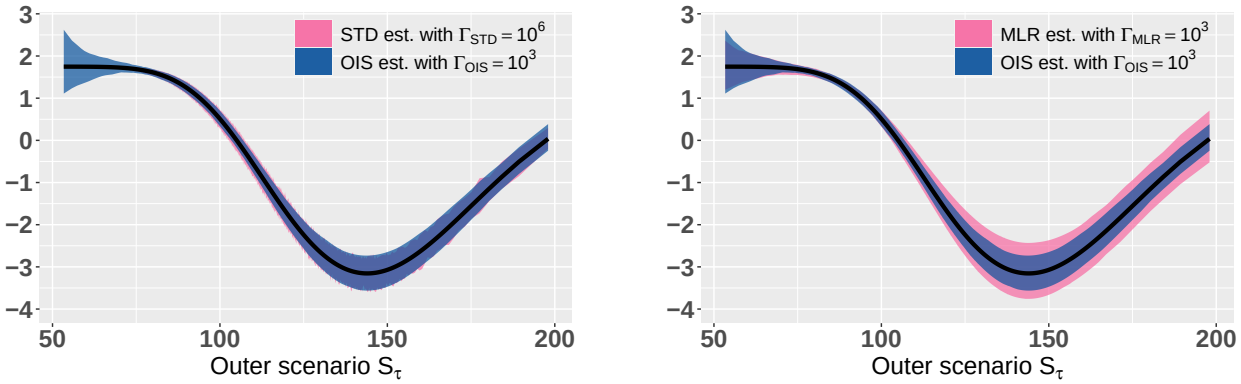
Figure 1: Problem settings for reverse iron butterfly option portfolio example.

To estimate a risk measure $\rho(\mu(S_\tau))$ via nested simulation, one first simulates m outer scenarios, $S_{\tau,1}, \ldots, S_{\tau,m} \overset{i.i.d.}{\sim} \pi(s_\tau)$, where $\pi(s_\tau)$ is the pdf of $S_\tau|S_0$ under the $\mathbb{P}$-measure. The next step is to estimate the conditional expectations $\mu(S_{\tau,i})$ for $i = 1, \ldots, m$. Specifically, for each scenario $S_{\tau,i}$:

1. The standard nested simulation (STD) simulates n inner samples, $S_{T,1}, \ldots, S_{T,n} \overset{i.i.d.}{\sim} p(s_T|S_{\tau,i})$, where $p(s_T|S_{\tau,i})$ is the conditional density of $S_T|S_{\tau,i}$ under the $\mathbb{Q}$-measure then estimates $\mu(S_{\tau,i})$ by the standard MC estimators $\widehat{\mu}_i = \widehat{\mu}(S_{\tau,i}) = \frac{1}{n}\sum_{j=1}^{n} g(S_{T,j})$. The total simulation budget is $\Gamma = mn$.

2. The GIS procedure simulates Γ inner samples, $\widetilde{S}_{T,1},\ldots,\widetilde{S}_{T,\Gamma} \overset{i.i.d.}{\sim} q(s_T)$ then estimates $\mu(S_{\tau,i})$ by the GIS estimators $\widetilde{\mu}_i = \widetilde{\mu}(S_{\tau,i}) = \frac{1}{\Gamma}\sum_{j=1}^{\Gamma} g(\widetilde{S}_{T,j})\frac{p(\widetilde{S}_{T,j}|S_{\tau,i})}{q(\widetilde{S}_{T,j})}$. For MLR estimators, $q(s_T)$ is the mixture pdf $\bar{p}(s_T)$ as defined in (13). For OIS estimators, $q(s_T)$ is the weighted mixture distribution $q_{\boldsymbol{\beta}^*}(s_T)$ as defined in (15). The simulation budget Γ is chosen by the user.

To compare the different nested simulation procedures and for ease of visualization, we consider $m = 1{,}000$ outer scenarios, which are the $\left(\frac{k}{m+1}\right)$-th quantile of the outer distribution, $k = 1,\ldots,m$. For each scenario $S_{\tau,i}$, we estimate $\mu(S_{\tau,i})$ with different estimators 200 times, i.e., 200 macro replications. We use the 2.5%- and 97.5%-tiles of the 200 estimates of $\mu(S_{\tau,i})$ as its 95% confidence interval (CI). Connecting the 95% CIs of 1,000 outer scenarios, we plot the 95% confidence bands for the $\mu(S_\tau)$. Two 95% confidence bands for the STD estimators are shown in Figure 1b around the solid line. Two different colors are used to denote experiments with different simulation budgets. As expected, the confidence band is narrower as the simulation budget increases. We also see that the inner simulation variance is large for scenarios near the dip (e.g., around $S_\tau = 145$) of the true value function $\mu(S_\tau)$.



(a) 95% confidence bands for OIS and STD estimators.

(b) 95% confidence bands for OIS and MLR estimators.

Figure 2: 95% confidence bands for different estimators. Please note the different simulation budgets.

Figure 2 compares the 95% confidence bands of the OIS estimator to those of the STD and the MLR estimators. We first note that the simulation budget for the OIS and the MLR estimators is 10^3 while STD estimator's simulation budget is 10^6. To ensure fair comparison, i.e., the OIS estimator indeed uses the simulation budget as stated, when implementing Algorithm 1, 100 inner samples were used in Stage I and 900 inner samples were used in Stage II. We see in Figure 2a that, even when the simulation budget is 1,000 times smaller, the OIS estimator's confidence band is comparable to the STD estimator's. This finding is consistent with Proposition 2. Figure 2b shows that, with the same simulation budget, the OIS estimator's confidence band is narrower than the MLR estimator's, particularly in the region where the inner simulation has high variance (e.g., around $S_T = 145$). This finding is consistent with Proposition 1. Comparing Figure 2a and Figure 2b, we see that both OIS and MLR estimators' confidence bands have similar widths with that of the STD estimator's, even though the latter's simulation budget is 1,000 times larger. The key message in Figure 2 is two-fold: (1) Reusing simulation outputs, such the OIS and MLR estimators, is a promising nested simulation design principle. (2) A well-chosen sampling distribution, like $q_{\boldsymbol{\beta}^*}(\cdot)$, can further improve the performance of a GIS procedure.

Table 1 summarizes the same experiments in Figures 1 and 2 and provides a quantitative comparison, i.e., comparing the average MSEs (AMSEs) among the OIS, MLR, and STD estimators. Recall that we ran 200 macro-replications to estimate $\mu(S_{\tau,i})$ for the same scenarios $S_{\tau,i}$, $i = 1,\ldots,m = 1{,}000$. Thus we could estimate the MSEs and AMSEs of the corresponding estimators of $\mu(S_{\tau,i})$. We see from the boxed

Table 1: Comparison of Average MSEs (AMSEs) of different nested estimators for $\mu(S_{\tau,i})$, $i = 1,\ldots,m$.

Estimators for $\mu(S_{\tau,i})$	OIS	MLR	STD different simulation budgets			
Simulation budget	$\boxed{\Gamma = 10^3}$	$\boxed{\Gamma = 10^3}$	$\boxed{\Gamma = 10^3}$	$\Gamma = 10^4$	$\Gamma = 10^5$	$\Gamma = 10^6$
AMSE reduction	$\boxed{1}$	$\boxed{2.03}$	$\boxed{1{,}114}$	110	11.06	1.10
AMSE	0.0167	0.0339	18.59	1.84	0.18	0.018

numbers in Table 1 that, when using the same simulation budget, the OIS estimator's AMSE is about half of the MLR estimator's and over 1,000 times smaller than the STD estimator's. This orders of magnitudes improvement is the result of (1) recycling and reusing simulation outputs, (2) using a sampling distribution that approximates the optimal one, and (3) stratified sampling.

Lastly, we examine different sampling densities in Figure 3 to better understand the OIS estimators. We first see that the true optimal sampling distribution $q^*(S_T)$ has a peak around $S_T = 145$, which corresponds to the dip in the simulation logic $g(S_T)$ and is a region where the inner simulation variance is high. In the language of importance sampling, this is an important region. The NNLS weighted mixture $q_{\beta^*}(S_T)$ approximates the shape of $q^*(S_T)$ and has a mode in that region. In contrast, the equal mixture $\bar{p}(S_T)$ as defined in (13) does not depend on the simulation logic $g(S_T)$ and its mode is located near the model of $\pi(S_\tau)$. Therefore, its inner simulation variance is high in the "important region". Combining this observation with the AMSE reduction in Table 1, we see that it is an advantageous feature for the optimal sampling distribution $q^*(S_T)$ to be dependent on $g(S_T)$.

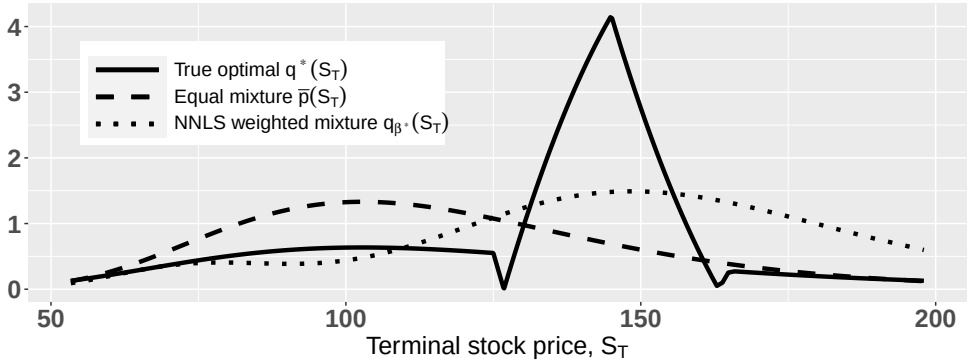


Figure 3: Plots of different sampling density functions for a GIS procedure.

6 CONCLUDING REMARKS

In this study, we generalize the classical importance sampling technique to identify an optimal sampling distribution for many different target distributions in a nested simulation procedure. Also, the random samples generated from this optimal sampling distribution are reused in different scenarios to improve efficiency. The optimal sampling distribution is derived in closed-form, but it has limited practical use. We further propose approximating the optimal sampling distribution via a weighted mixture distribution based on non-negative least squares regression (NNLS). Our numerical experiments show that the proposed method achieves a smaller average MSE than that of the standard nested simulation by using only 0.1% of the latter's simulation budget. The proposed method also achieves smaller average MSE compared to one of the best known likelihood-ratio-based sample reusing techniques in the literature.

As shown by our numerical experiments, the proposed GIS procedure is highly efficient compared to the standard nested simulation procedure. But there are drawbacks. For example, though we use the simulation budgets to represent computations of different simulation procedures, implementing NNLS and

calculating the likelihood ratios involved add computations to the GIS procedure too. Therefore the GIS procedure has limited applicability for problems with simple simulation logic. Also, further developments are needed to identify an optimal sampling distribution for stochastic scenarios rather than given scenarios.

ACKNOWLEDGMENTS

This work is partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grants RGPIN-2018-03755.

REFERENCES

Barton, R. R. 2012. "Tutorial: Input Uncertainty in Output Analysis". In *Proceedings of the 2012 Winter Simulation Conference*, edited by C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A. M. Uhrmacher, 1–12. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Beckman, R. J., and M. D. McKay. 1987. "Monte Carlo Estimation under Different Distributions Using the Same Simulation". *Technometrics* 29(2):153–160.

Broadie, M., Y. Du, and C. C. Moallemi. 2011. "Efficient Risk Estimation via Nested Sequential Simulation". *Management Science* 57(6):1172–1194.

Broadie, M., Y. Du, and C. C. Moallemi. 2015. "Risk Estimation via Regression". *Operations Research* 63(5):1077–1097.

Cheng, R. C. H., and W. Holland. 1997. "Sensitivity of Computer Simulation Experiments to Errors in Input Data". *Journal of Statistical Computation and Simulation* 57(1-4):219–241.

Feng, M., and J. Staum. 2015. "Green Simulation Designs for Repeated Experiments". In *Proceedings of the 2015 Winter Simulation Conference*, edited by L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, 403–413. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Feng, M., and J. Staum. 2017. "Green Simulation: Reusing the Output of Repeated Experiments". *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 27(4):23.

Gelfand, I. M., and S. V. Fomin. 2000. *Calculus of Variations*. Revised English ed. Mineola, N.Y.: Dover Publications. Translated and edited by Richard A. Silverman.

Glasserman, P. 2013. *Monte Carlo Methods in Financial Engineering*, Volume 53. Springer Science & Business Media.

Gordy, M. B., and S. Juneja. 2010. "Nested Simulation in Portfolio Risk Measurement". *Management Science* 56(10):1833–1848.

Hesterberg, T. C. 1988. *Advances in Importance Sampling*. Ph. D. thesis, Stanford University.

Hong, L. J., S. Juneja, and G. Liu. 2017. "Kernel Smoothing for Nested Estimation with Application to Portfolio Risk Measurement". *Operations Research* 65(3):657–673.

L'Ecuyer, P. 1990. "A Unified View of the IPA, SF, and LR Gradient Estimation Techniques". *Management Science* 36(11):1364–1383.

Lee, S.-H. 1998. *Monte Carlo Computation of Conditional Expectation quantiles*. Ph. D. thesis, Stanford University.

Lee, S.-H., and P. W. Glynn. 2003. "Computing the Distribution Function of a Conditional Expectation via Monte Carlo: Discrete Conditioning Spaces". *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 13(3):238–258.

Owen, A. B. 2013. *Monte Carlo Theory, Methods and Examples*.

Rubinstein, R. Y., and A. Shapiro. 1993. *Discrete Event Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Method*, Volume 13. Wiley.

Veach, E. 1997. *Robust Monte Carlo Methods for Light Transport Simulation*. Ph. D. thesis, Stanford University.

Zhang, K., B. M. Feng, G. Liu, and S. Wang. 2022. "Sample Recycling for Nested Simulation with Application in Portfolio Risk Measurement". *ArXiv*. Accessed April 13, 2023. https://arxiv.org/abs/2203.15929.

Zhu, H., T. Liu, and E. Zhou. 2020. "Risk Quantification in Stochastic Simulation under Input Uncertainty". *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 30(1):1–24.

AUTHOR BIOGRAPHIES

QINGYUAN CHEN is a first-year PhD student in Department of Operations Research and Information Engineering at Cornell University. Her research interests are analysis and design simulation experiments, applied to real-life probelms. Her email is qc72@cornell.edu.

MINGBIN BEN FENG is an Assistant Professor in actuarial science at the University of Waterloo. He is an Associate of the Society of Actuaries (ASA) and a Certified Analytics Professional (CAP). His research interests include stochastic simulation design and analysis, optimization via simulation, nonlinear optimization, and financial and actuarial applications of simulation and optimization methodologies. His e-mail address is ben.feng@uwaterloo.ca. His website is http://www.math.uwaterloo.ca/∼mbfeng/.