Proceedings of the 2024 Winter Simulation Conference
H. Lam, E. Azar, D. Batur, S. Gao, W. Xie, S. R. Hunter, and M. D. Rossetti, eds.

A TUTORIAL ON NESTED SIMULATION

Guangwu Liu[1], and Kun Zhang[2]

[1]Department of Management Sciences, College of Business, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong SAR, CHINA
[2]Institute of Statistics and Big Data, Renmin University of China, Beijing, CHINA

ABSTRACT

Nested simulation refers to the problem of estimating a functional of a conditional expectation that cannot be evaluated analytically and requires simulation. It finds a wide range of applications in operations research and machine learning, including portfolio risk measurement, pricing of complex derivatives, and Bayesian experimental design. Nested simulation typically proceeds at two levels. At the outer level, a number of scenarios are simulated. Then, at the inner level, given each scenario, one or multiple samples are simulated for estimating the conditional expectation. Various approaches have been proposed for estimating the conditional expectation and conducting nested simulation itself, focusing on point estimation and/or interval estimation, and addressing general or specified forms of the functional. In this tutorial, we review several approaches to nested simulation and discuss their contributions to the field.

1 INTRODUCTION

Nested simulation refers to the problem of estimating a functional of a conditional expectation, where the conditional expectation cannot be evaluated analytically but instead requires simulation. Specifically, the quantity of interest can be expressed as

$$\alpha = \rho(\mathbb{E}[Y|X]), \tag{1}$$

where X is a random vector in $\mathbb{R}^d$ with $d \geq 1$, Y is a random variable in $\mathbb{R}$, and the functional ρ maps a random variable to a real number, such as the cumulative distribution function or quantiles.

The problem of estimating α in the form of (1) finds several applications in operations research, particularly in risk measurement for financial portfolios. For instance, Y might represent the (discounted) loss of a portfolio at maturity, e.g., one year, and $\mathbb{E}[Y|X]$ denotes its mark-to-market loss for a given scenario of risk factors X up to a risk horizon prior to maturity, e.g., one week. In financial risk management, quantifying various risk measures associated with the mark-to-market loss $\mathbb{E}[Y|X]$, such as the value-at-risk (VaR) and the conditional value-at-risk (CVaR), is of paramount importance.

The form of (1) is quite general and can accommodate different risk measures. For instance, VaR with a confidence level $\beta \in (0,1)$ is defined as:

$$\rho(\mathbb{E}[Y|X]) = \inf\{x : F(x) \geq \beta\}, \tag{2}$$

where F denotes the cumulative distribution function of $\mathbb{E}[Y|X]$, and another widely used risk measure, CVaR, with a confidence level β, is given by

$$\rho(\mathbb{E}[Y|X]) = \inf_{u\in\mathbb{R}} \left\{u + (1-\beta)^{-1}\mathbb{E}(\mathbb{E}[Y|X]-u)^+\right\}. \tag{3}$$

In addition to VaR and CVaR, other forms of ρ are also studied in the literature. For instance, when ρ is expressed as $\rho(\mathbb{E}[Y|X]) = \mathbb{E}[g(\mathbb{E}[Y|X])]$ for an appropriate function $g(\cdot)$, it leads to different risk measures

such as squared tracking error or expected excess loss (Broadie et al. 2015; Hong et al. 2017). In particular, when $g(\cdot)$ is a quadratic function, i.e., $g(z) = (z-z_0)^2$ for a given threshold value z_0, $\rho(\mathbb{E}[Y|X])$ is referred to as squared tracking error; when $g(\cdot)$ is a hockey-stick function, i.e., $g(z) = (z-z_0)^+$, $\rho(\mathbb{E}[Y|X])$ is referred to as expected excess loss; when $g(\cdot)$ is an indicator function, i.e., $g(z) = \mathbb{1}\{z \geq z_0\}$), $\rho(\mathbb{E}[Y|X])$ is referred to as probability of large losses beyond z_0.

Another application in financial engineering is the pricing of complex derivatives, such as collateralized debt obligations (CDOs). Let X represent the underlying variables, such as default event indicators, and Y represent the portfolio loss of defaultable assets with reference to the CDOs (Gordy and Juneja 2006). Then, α is the price of the tranches of the CDOs with an attachment point a and a detachment point b ($a < b$), and is expressed as

$$\alpha = \mathbb{E}[\min(b, \max(\mathbb{E}[Y|X], a)) - a] = \mathbb{E}\left[(\mathbb{E}[Y|X] - a)^+ - (\mathbb{E}[Y|X] - b)^+\right].$$

It is seen that this is a nested estimation with $g(x) = (x-a)^+ - (x-b)^+$, where $x^+ = \max\{x, 0\}$.

Beyond financial engineering, nested simulation techniques also find applications in Bayesian experimental design within the field of machine learning. Here, an optimal experimental design is sought to maximize the expected information gain (EIG); see, e.g., Goda et al. (2020). Specifically, the EIG, denoted by G, is expressed in the form $G = -\mathbb{E}_\theta[\log p(\theta)] - \mathbb{E}_X[-\mathbb{E}_\theta[\log p(\theta|X)]]$, where $\mathbb{E}_X$ and $\mathbb{E}_\theta$ are the expectations taken with respect to X and θ respectively, and $p(\theta)$ and $p(\theta|X)$ are respectively the probability densities of θ and θ given X. In Bayesian experimental designs, $-\mathbb{E}_\theta[\log p(\theta)]$ is the information entropy of the input random variable θ, and $\mathbb{E}_X[-\mathbb{E}_\theta[\log p(\theta|X)]]$ is the expected conditional information entropy of θ by collecting data X. Therefore, the EIG measures the expected amount of the reduction of the information entropy about θ by collecting data X.

By Bayes' theorem, $p(\theta|X) = p(X|\theta)p(\theta)/p(X)$. Note that $p(X) = \mathbb{E}_\theta[p(X|\theta)]$, then

$$G = \mathbb{E}[\log(p(X|\theta))] - \mathbb{E}_X[\log(\mathbb{E}_\theta[p(X|\theta)])]. \tag{4}$$

Note that only the second term on the right hand side of (4) is a nested simulation with $\rho(\mathbb{E}[Y|X]) = \mathbb{E}[g(\mathbb{E}[Y|X])]$ and $g(z) = \log z$ and $Y = p(X|\theta)$. Therefore, the EIG is a combination of a non-nested simulation and a nested simulation.

In this paper, we mainly discuss the following forms of the functional ρ, which accommodate most, if not all of the commonly encountered applications in practice. Existing literature often establishes methods for one or several of these forms.

($\mathscr{A}$.1) $\rho(\cdot) = \mathbb{E}g(\cdot)$, where $g(\cdot)$ is a smooth function,
($\mathscr{A}$.2) $\rho(\cdot) = \mathbb{E}g(\cdot)$, where $g(\cdot)$ is a hockey-stick function,
($\mathscr{A}$.3) $\rho(\cdot) = \mathbb{E}g(\cdot)$, where $g(\cdot)$ is an indicator function,
($\mathscr{A}$.4) $\rho(\cdot)$ is the VaR at a confidence level of β as in (2),
($\mathscr{A}$.5) $\rho(\cdot)$ is the CVaR at a confidence level of β as in (3).

To estimate a quantity of interest, e.g., (1), both point and interval estimations could be established. Point estimation aims to find a single estimate for the quantity of interest. Interval estimation, on the other hand, provides a range of values within which the true value of the quantity is expected to lie with a certain level of confidence. These two types of estimations serve different purposes and are often used together to provide a comprehensive understanding of the uncertainty associated with the estimated quantity.

In this paper, we review several methods proposed in the existing literature for estimating the functional of a conditional expectation given in (1). The review covers two aspects: point estimation and interval estimation, which are discussed in Sections 2 and 3, respectively. It is important to note that our review is not intended to be exhaustive, given the vast literature in this field, but rather to help readers understand this topic in a more systematic manner. More specifically, the reviewed methods are catagorized into general-purpose approaches and functional-specific approaches, respectively. The former can be used to

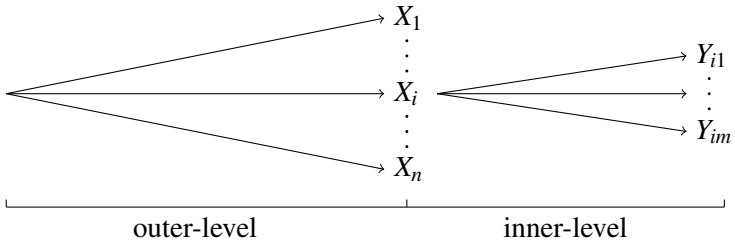

Figure 1: Illustration of standard nested simulation.

estimate the general form of (1), including ($\mathscr{A}$.1) $\sim$ ($\mathscr{A}$.5), which are discussed in Sections 2.1 and 3.1, respectively. The latter are proposed for estimating one or two of the above forms, and are discussed in Sections 2.2 and 3.2, respectively.

2 POINT ESTIMATION

In this section, we review methods for the point estimation of the functional given in (1). Due to the requirement of simulating inner-level samples in the estimation process, the convergence rate of the mean squared error (MSE) of the point estimator for a given sampling budget is often slower than that of the non-nested Monte Carlo simulation. Consequently, much of the existing literature focuses on proposing point estimators for (1) with the goal of accelerating the convergence rate of the MSE for the proposed estimators.

2.1 General-Purpose Approaches

In this part, we introduce methods capable of estimating the general form of the functional given in (1). These methods share a common objective: estimating the conditional expectation $\mathbb{E}[Y|X]$ in (1), irrespective of the specified form of the function $g(\cdot)$. In particular, standard nested simulation (SNS) involves proposing estimators of $\mathbb{E}[Y|X]$ at each realization of X, while metamodeling approaches, such as the least-squares method (LSM), aim to fit $\mathbb{E}[Y|X]$ using parametric or non-parametric methods.

2.1.1 Standard Nested Simulation

We introduce the SNS method for various functionals in (1). In the estimation of the quantity in (1), SNS proceeds in the following two levels shown in Figure 1.

1. *Outer-level simulation.* Generate n independent and identically distributed (i.i.d.) samples based on the distribution of X, denoted by $\{X_i, i = 1,2,...,n\}$.
2. *Inner-level simulation.* For each X_i, generate m i.i.d. samples of Y, denoted by $\{Y_{ij}, j = 1,2,...,m\}$.

Denote $L(X) = \mathbb{E}[Y|X]$. Then, by the laws of large numbers, $L(X_i)$ can be approximated by

$$L_m(X_i) \triangleq \bar{Y}_i = \frac{1}{m}\sum_{j=1}^{m} Y_{ij}. \tag{5}$$

As a result, the quantity of interest, i.e., $\alpha = \rho(L(X))$, can be estimated using $\{L_m(X_1),\ldots,L_m(X_n)\}$. For instance, when $\rho(\cdot) = \mathbb{E}g(\cdot)$ for a given function $g(\cdot)$, an SNS estimator of α is given by

$$\alpha_{m,n} = \frac{1}{n}\sum_{i=1}^{n} g(L_m(X_i)). \tag{6}$$

When $\rho(L(X))$ is the quantile (or VaR) of $L(X)$ at a confidence level of β, the estimator $\alpha_{m,n}$ can be set as the sample quantile of $\{L_m(X_1),...,L_m(X_n)\}$, denoted by $L_{m,\lceil \beta n \rceil}$, i.e.,

$$\alpha_{m,n} = L_{m,\lceil \beta n \rceil}, \tag{7}$$

where the operator $\lceil x \rceil$ denotes the smallest integer that is not less than x. When $\rho(L(X))$ is the CVaR of $L(X)$ with a confidence level β, the estimator is given by

$$\alpha_{m,n} = \inf_{u \in \mathbb{R}} \left\{ u + \frac{1}{1-\beta} \frac{1}{n} \sum_{i=1}^{n} [L_m(X_i) - u]^+ \right\}. \tag{8}$$

When ρ takes the forms as in ($\mathscr{A}$.3)-($\mathscr{A}$.5), Gordy and Juneja (2010) establish asymptotic bias and variance of the SNS estimator. Following the same spirit, Zhang et al. (2022) provide asymptotic bias and variance when ρ takes forms as in ($\mathscr{A}$.1) and ($\mathscr{A}$.2). These results corresponding to ($\mathscr{A}$.1)-($\mathscr{A}$.5) are summarized in the following,

$$\mathbb{E}\alpha_{m,n} - \alpha = A/m + o(1/m) \quad \text{and} \quad \mathrm{Var}(\alpha_{m,n}) = C/n + o(1/n), \tag{9}$$

where A and C are constants that depend on the forms of ρ and the distribution of (X,Y).

Based on (9), the optimal budget allocation between the inner and outer levels can be obtained, which has been presented in the literature; see, e.g., Gordy and Juneja (2010) and Zhang et al. (2022). Specifically, denoting the effort to generate an outer sample and an inner sample as γ_0 and γ_1, respectively, the total simulation budget is $\Gamma = n(m\gamma_1 + \gamma_0)$. Note that in practical applications, it is usually less time consuming to generate an outer sample than an inner sample in practice, so without loss of generality, we assume that $\gamma_0 = 0$ and $\gamma_1 = 1$ (Zhang et al. 2022). To guide the optimal allocation of the sampling budget, a useful approach is through the minimization of the (asymptotic) MSE. Specifically, we consider the following optimization problem:

$$\begin{aligned} \underset{m,n}{\text{minimize}} \quad & \frac{A^2}{m^2} + \frac{C}{n} \\ \text{subject to} \quad & nm = \Gamma, \\ & m, n \geq 0. \end{aligned} \tag{10}$$

Solving this problem leads to an asymptotically optimal budget allocation rule, with

$$m^* = \left(\frac{2A^2}{C} \right)^{1/3} \Gamma^{1/3} + o(\Gamma^{1/3}), \quad n^* = \left(\frac{C}{2A^2} \right)^{1/3} \Gamma^{2/3} + o(\Gamma^{2/3}). \tag{11}$$

That is, the order of n^* is asymptotically the square of that of m^*. As a result, Gordy and Juneja (2010) suggest that using relatively small sample size in the inner level could yield accurate estimators for α.

By substituting this asymptotically optimal budget allocation (11) into the results in (9), the optimal bias, variance and MSE of the SNS estimator are obtained as follows:

$$\begin{cases} \mathrm{Bias}^*_{\mathrm{SNS}} = & \left(\frac{AC}{2} \right)^{1/3} \Gamma^{-1/3} + o\left(\Gamma^{-1/3}\right), \\ \mathrm{Var}^*_{\mathrm{SNS}} = & \left(2A^2C^2\right)^{1/3} \Gamma^{-2/3} + o\left(\Gamma^{-2/3}\right), \\ \mathrm{MSE}^*_{\mathrm{SNS}} = & \frac{3}{2}\left(2A^2C^2\right)^{1/3} \Gamma^{-2/3} + o\left(\Gamma^{-2/3}\right). \end{cases} \tag{12}$$

To obtain the SNS estimator with the asymptotically optimal MSE, determining the optimal inner and outer samples sizes m^* and n^* is crucial. However, the obstacle lies in the unknown constants A and

C, as indicated in (11). To address this challenge, Zhang et al. (2022) propose a bootstrap-based and sample-driven budget allocation method. Pilot samples are utilized to estimate the unknown constants A and C through the bootstrap technique, enabling the estimation of the optimal budget allocation m^* and n^*. Furthermore, these pilot samples can also aid in constructing the estimator of α in (1).

More recently, Liang et al. (2024) develop an approach by combining the SNS method with the jackknife technique, proposing a so-called FAST estimator for (1). Their work has improved upon the results of Gordy and Juneja (2010). The jackknife technique proceeds as follows. For each outer-level sample X_i, the m samples in the inner level are divided into J nonoverlapping groups (where we assume that m is divisible by J without loss of generality), and each group contains m/J samples. $L_{m,-l}(X_i)$ is the inner-level estimate for $L(X_i)$ that is obtained when group l is omitted. Denote $L_{m,i,-l} = L_{m,-l}(X_i)$ and $L_{m,-l} = L_{m,-l}(X)$ for simplicity. The sampling method is shown in Figure 2 for $J = 3$. When ρ takes forms as in ($\mathscr{A}$.1) $\sim$ ($\mathscr{A}$.3), Liang et al. (2024) establish the FAST estimator $\alpha_{m,n,J}$ expressed as

$$\alpha_{m,n,J} = \frac{1}{n}\sum_{i=1}^{n} \hat{g}(L_{m,i}), \tag{13}$$

where

$$\hat{g}(L_{m,i}) = \frac{1}{J}\sum_{l=1}^{J} \left[J g(L_{m,i}) - (J-1) g(L_{m,i,-l}) \right].$$

Liang et al. (2024) first refine the bias of $\alpha_{m,n}$ and point out that the $o(1/m)$ in (9) is actually $B/m^2 + O(1/m^3)$, where B is also a constant depending on the forms of ρ and the distribution of (X,Y). Based on this, the bias and variance of $\alpha_{m,n,J}$ are provided:

$$\mathbb{E}[\alpha_{m,n,J}] - \alpha = -\frac{JB}{(J-1)m^2} + o\left(\frac{1}{m^2}\right), \ \mathrm{Var}[\alpha_{m,n,J}] = \frac{C}{n} + o\left(\frac{1}{n}\right). \tag{14}$$

Comparing (9) and (14), the convergence rate of the bias of the FAST estimator is much faster than that of the conventional SNS estimator. By solving the similar optimization problem to (10), the asymptotically optimal budget allocation for the FAST estimator is

$$m^* = \left(\frac{4J^2B^2}{(J-1)^2C}\right)^{1/5} \Gamma^{1/5}, \quad n^* = \left(\frac{(J-1)^2C}{4J^2B^2}\right)^{1/5} \Gamma^{4/5}. \tag{15}$$

Substituting this budget allocation into the results in (14), the asymptotically optimal bias, variance and MSE of the FAST estimator are as follows:

$$\begin{cases} \mathrm{Bias}^*_{\mathrm{FAST}} = & -\left(\dfrac{JBC^2}{16(J-1)}\right)^{1/5} \Gamma^{-2/5} + o\left(\Gamma^{-2/5}\right), \\ \mathrm{Var}^*_{\mathrm{FAST}} = & \left(\dfrac{4J^2B^2C^4}{(J-1)^2}\right)^{1/5} \Gamma^{-4/5} + o\left(\Gamma^{-4/5}\right), \\ \mathrm{MSE}^*_{\mathrm{FAST}} = & \dfrac{5}{4}\left(\dfrac{4J^2B^2C^4}{(J-1)^2}\right)^{1/5} \Gamma^{-4/5} + o\left(\Gamma^{-4/5}\right). \end{cases} \tag{16}$$

It can be seen that the MSE of the FAST estimator converges to 0 at the rate of $\Gamma^{-4/5}$, which is faster than that of the conventional SNS estimator.

Based on the bootstrap-based budget allocation method proposed by Zhang et al. (2022), Liang et al. (2024) establish efficient algorithm so that the MSE of the estimator could converge at the asymptotically optimal rate. In addition, Liang et al. (2024) discuss the application of nested simulation in Bayesian experimental design within the field of machine learning.

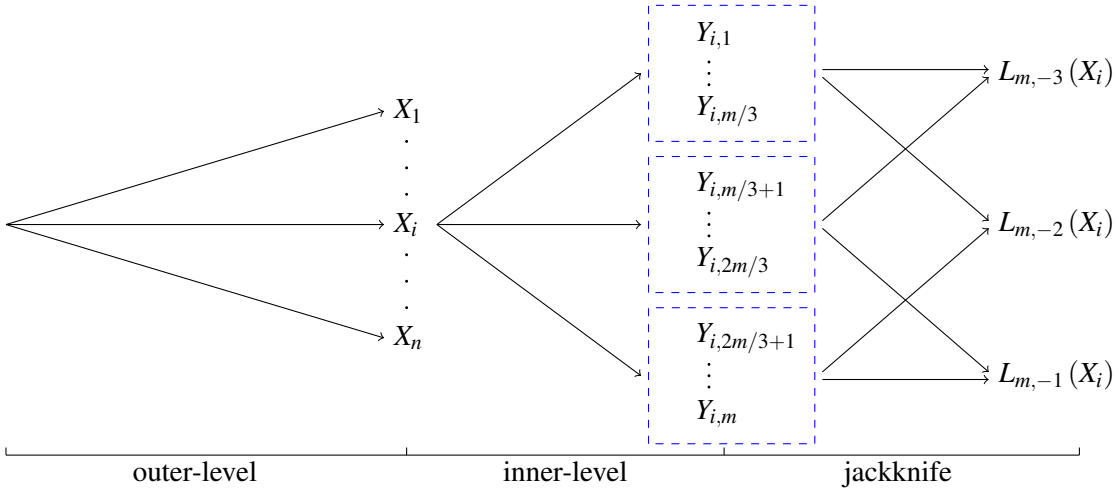

Figure 2: Illustration of nested simulation with the jackknife method ($J = 3$).

2.1.2 Stochastic Kriging

Stochastic kriging was initially proposed by Ankenman et al. (2010), and its roots trace back to Ankenman et al. (2008). This method approximates the conditional expectation as follows:

$$L(X_i) = \mathbb{E}[Y|X_i] \approx \bar{\mathscr{Y}}(X_i) = \frac{1}{m_i}\sum_{j=1}^{m_i} \mathscr{Y}_j(X_i), \tag{17}$$

where X_i represents the ith outer level sample, and $\mathscr{Y}_j(X_i)$ denotes the output of stochastic simulations at design point X_i on replication j:

$$\mathscr{Y}_j(x) = f(x)^\top \beta + \mathrm{M}(x) + \varepsilon_j(x). \tag{18}$$

Here, $f(x)$ constitutes a set of basis functions, β is a parameter vector, M stands for a realization of a mean 0 random field, and $\varepsilon_j(x)$'s, at a given design x, are naturally independent and identically distributed across replications.

Ankenman et al. (2010) proposed an algorithm for estimating the parameter vector β. This estimated vector is then substituted into (18) and subsequently into (17), resulting in an estimator of $\bar{\mathscr{Y}}(X_i)$, denoted by $\hat{\bar{\mathscr{Y}}}(X_i)$. By replacing $L_m(X_i)$ in (6), (7), and (8) with $\hat{\bar{\mathscr{Y}}}(X_i)$, one obtains a stochastic kriging estimator of α.

Liu and Staum (2010) first introduce the stochastic kriging method for estimating CVaR. However, the asymptotic bias, variance and MSE of the estimator are not provided. In fact, there is limited research in the literature on theoretical analyses of nested estimators resulting from stochastic kriging, partially due to the technical challenges arising in the integration of metamodel fitting and the nested simulation framework. Additionally, implementation difficulties may arise due to matrix inversions, especially if the matrix dimension is high. For the latter issue, methods proposed by Ding and Zhang (2024) offer potential powerful solutions. To maintain the focus of this paper, we skip a detailed review of the work by Ding and Zhang (2024), while referring interested readers to their paper for further details.

2.1.3 Least-Squares Method

The LSM is a classical parametric method. Broadie et al. (2011b) and Broadie et al. (2015) propose this method within the context of nested simulation. The LSM fits the conditional expectation in (1) by a group of basis functions in $\mathscr{L}^2$ space. In particular, let $\psi(\cdot) = (\psi_1(\cdot), \ldots, \psi_\eta(\cdot))^\top$ represent a set of

η basis functions, where each $\psi_k : \mathbb{R} \to \mathbb{R}$ is a one-dimensional function. For a d-dimensional vector $x \triangleq (x_1, \ldots, x_d)^\top \in \mathbb{R}^d$, for notational simplicity we let

$$\psi(x) = (\psi_1(x_1), \ldots, \psi_\eta(x_1), \psi_1(x_2), \ldots, \psi_\eta(x_2), \ldots, \psi_1(x_d), \ldots, \psi_\eta(x_d))^\top. \tag{19}$$

Then, $L(x)$ is approximated by the least-squares approximating function $\beta^{*\top}\psi(x)$, i.e.,

$$L(x) = \mathbb{E}[Y|X = x] \approx \beta^{*\top}\psi(x),$$

with β^* given by

$$\beta^* = \underset{\beta \in \mathbb{R}^{\eta d}}{\arg\min}\, \mathbb{E}\left[Y - \beta^\top \psi(X)\right]^2.$$

Given Monte Carlo samples $\{(X_i, \bar{Y}_i), 1 \le i \le n\}$, where $\bar{Y}_i$ is defined as in (5), β^* can be estimated by

$$\beta_n = \underset{\beta \in \mathbb{R}^{\eta d}}{\arg\min}\, \frac{1}{n}\sum_{i=1}^{n}\left[\bar{Y}_i - \beta^\top \psi(X_i)\right]^2,$$

and so $L(x)$ can be estimated by $\beta_n^\top \psi(x)$. Replacing $L_m(X_i)$ in (6), (7) and (8) by $\beta_n^\top \psi(X_i)$ leads to the LSM estimator of α.

The LSM exhibits a notable distinction from SNS: even when the sample size in the inner level m is finite or equal to 1, LSM remains applicable. This characteristic contrasts sharply with SNS, where m must tend to infinity to ensure the convergence of the estimator. Specifically, when ρ takes forms as in ($\mathscr{A}$.1) and ($\mathscr{A}$.2), Broadie et al. (2015) prove that the MSE of the LSM estimator converges at the rate of Γ^{-1} and $\Gamma^{-1+\delta}$ for any $\delta > 0$, respectively, until it reaches a regression model error. This error arises because, in practice, basis functions may not fully span the true conditional expectation in $\mathscr{L}^2$ space, meaning $\beta_n^\top \psi(x)$ is not an asymptotically unbiased estimator of $L(x)$. To mitigate this model error, Wang et al. (2024) propose fitting the conditional expectation $\mathbb{E}[Y|X = x]$ by using the kernel ridge regression in the reproducing kernel Hilbert space, a subspace of $\mathscr{L}^2$ space. The convergence rate of the MSE of this estimator is shown to depend on the smoothness of the density function of the conditional expectation and the dimension d.

Broadie et al. (2015) highlight an open problem regarding the estimation of bias in individual problem instances, which essentially refers to the model error. Addressing this challenge remains a compelling avenue for future research. Notably, Lai et al. (2024) propose a method to construct a confidence interval for CVaR using LSM, which effectively addresses the model error. We will delve into this approach in Section 3. However, (accurately) estimating the model error of LSM across diverse problem domains remains an open challenge.

2.1.4 Kernel Smoothing Method

The kernel smoothing method, a classic nonparametric approach, is introduced into nested simulation by Hong and Juneja (2009) and further elaborated upon by Hong et al. (2017). This method aims to fit the conditional expectation in (1) as follows.

$$L(x) = \mathbb{E}[Y|X = x] \approx \tilde{L}(x) \triangleq \sum_{l=1}^{n} \bar{Y}_l \frac{K_h(x - X_l)}{\sum_{k=1}^{n} K_h(x - X_k)},$$

where $\bar{Y}_l$ is defined as in (5), $K_h(x) = (1/h^d)K(x/h)$, $K(\cdot)$ is a kernel function that is a density function symmetric about 0 on $\mathbb{R}^d$, and h is a bandwidth parameter satisfying $h \to 0$ as $n \to \infty$. Evidently, the conditional expectation is approximated by a weighted average of $\bar{Y}_l (l = 1, ..., n)$, and the weights are

constructed by kernel functions. Replacing $L_m(X_i)$ in (6), (7) and (8) by $\tilde{L}(X_i)$ leads to the kernel smoothing estimator of α.

Many kernels can be used in the calculation of $\tilde{L}(x)$. A simple example is the uniform kernel $K(x) = \prod_{i=1}^{d} \mathbb{1}\{-\frac{1}{2} \leq x_i \leq \frac{1}{2}\}$ (Hong et al. 2017). For more kernels, such as Gaussian, tri-cube, and Epanechnikov kernels, see Chapter 6 in Hastie et al. (2009). Note that the choice of the kernel function k is often not critical (see, e.g., Section 4.5 of Härdle (1990)) for the performances of the estimators.

Similar to the LSM, the kernel smoothing method remains applicable even when the sample size in the inner level m is finite or equal to 1. Unlike the LSM, however, $\tilde{L}(x)$ serves as an asymptotically unbiased estimator of $L(x)$ under mild conditions, implying that for any x, $\tilde{L}(x)$ converges to $L(x)$ as $n \to \infty$. Consequently, it is expected that the kernel smoothing estimator could converge to the true value of α. Specifically, Hong et al. (2017) demonstrate that when ρ takes forms as in ($\mathscr{A}$.1) through ($\mathscr{A}$.3), the MSE of the kernel smoothing estimator converges to 0 at the rate of $\Gamma^{-\min\{1,\frac{4}{d+2}\}}$. For low-dimensional problems ($d \leq 2$), this convergence rate matches that of the non-nested Monte Carlo simulation method. For high-dimensional problems in the application of (1) to portfolio risk measurement, a decomposition technique is further proposed to enable efficient implementation of the kernel smoothing method in practice.

2.1.5 Green Nested Simulation

By ultilizing the likelihood ratios in the simulation process, Zhang et al. (2017) and Zhang et al. (2022) propose a method that recycles the same set of inner simulation outputs via the likelihood ratio method to estimate the conditional expectation in (1) across various scenarios. This simulation procedure is called the green nested simulation (GNS) procedure.

In particular, notice that for a typical nested simulation problem in the context of portfolio risk measurement, Y can usually be expressed as $Y = H(X, \xi)$ for certain functional $H(\cdot,\cdot)$ and a random vector ξ. Based on this observation, the GNS procedure proceeds as follows.

1. (Outer simulation) Simulate n i.i.d. outer scenarios, denoted by $X_1, \ldots, X_n$.
2. (Inner simulation) Simulate m i.i.d. inner replications, e.g., $\xi_1, \ldots, \xi_m \overset{i.i.d.}{\sim} \tilde{f}(\cdot)$ then estimate $L(X_i)$ by
$$\hat{L}(X_i) = \frac{1}{m} \sum_{j=1}^{m} H(X_i, \xi_j) \frac{f(\xi_j|X_i)}{\tilde{f}(\xi_j)}, \quad i = 1, \ldots, n.$$

Replacing $L_m(X_i)$ in (6), (7) and (8) by $\hat{L}(X_i)$ leads to the GNS estimator of α.

Zhang et al. (2022) discuss the appropriate choice of the sampling density $\tilde{f}(\cdot)$. When ρ assumes forms such as those in ($\mathscr{A}$.1) through ($\mathscr{A}$.3), the asymptotic bias, variance, and MSE of the GNS estimator are established. Notably, the MSE converges to 0 at the rate of Γ^{-1}, aligning with the convergence rate observed in non-nested Monte Carlo simulations. This accelerated rate of convergence is obtained mainly due to the recycling of the inner samples, which collectively contributes to the estimation of $L(X_i)$ for each i through the adjustment based on the likelihood ratios.

2.2 Functional-Specific Approaches

In addition to the methods discussed in Section 2.1, several approaches have been developed specifically for handling specific functionals in (1).

2.2.1 Probability of Large Losses

Consider $\rho(\cdot) = \mathbb{E}g(\cdot)$ with $g(z) = \mathbb{1}\{z \geq z_0\}$, i.e.,

$$\alpha = \mathbb{P}(\mathbb{E}[Y|X] \geq z_0),$$

where the known constant z_0 is a pre-specified threshold. In this case, α is the probability that the future portfolio loss exceeds z_0, and therefore, this risk measure is usually referred to as the probability of large losses. To efficiently estimate this risk measure, Broadie et al. (2011a) propose a method known as nested sequential simulation.

Unlike the SNS method illustrated in Figure 1, the nested sequential simulation approach sequentially allocates different inner-level sample sizes to different outer-level scenarios. The primary emphasis of this method lies in determining the inner-level sample size for each outer-level sample X_i, tailored to the indicator form of $g(\cdot)$. To achieve this, Broadie et al. (2011a) establish a theoretical criterion for this sequential allocation. Essentially, the criterion dictates that if $L(X_i)$ is closer to the threshold z_0 and/or the nested simulation estimator of $L(X_i)$ exhibits greater variance, then the inner-level sample size corresponding to X_i tends to be larger. By efficiently allocating inner-level samples according to this criterion, the MSE of the nested sequential simulation estimator converges to 0 at the rate of $\Gamma^{-4/5+\delta}$ for any $\delta > 0$.

In addition, to estimate the probability of large losses, Wang et al. (2023) combine the well-known optimal computing budget allocation method, which allows them to establish a similar criterion for the sequential allocation akin to that proposed by Broadie et al. (2011a). Furthermore, the almost sure convergence of their proposed estimator is provided, rather than the convergence rate.

Going beyond the estimation of the probability of large losses, Steckley et al. (2016) delve into estimating the density of a conditional expectation. Their approach combines kernel density estimation within the nested simulation framework, allowing for the estimation of the density of a conditional expectation.

2.2.2 Value-at-Risk

Consider the case of estimating (1) when ρ denotes a quantile or VaR. Note that the inverse of the probability of large losses is in fact the VaR/quantile of $\mathbb{E}[Y|X]$, and general-purpose approaches can be employed to estimate VaR by taking the inverse of the probability of large losses.

Recently, Liu et al. (2024) introduce the kernel quantile estimation (KQE) within the framework of SNS and propose nested KQE for estimating portfolio VaR. Notably, KQE exhibits a smaller variance compared to sample quantiles. Liu et al. (2024) prove that the nested KQE displays reduced variance compared to the conventional nested sample quantile methods discussed in previous works such as Gordy and Juneja (2010) and Zhang et al. (2022). Surprisingly, instances are also found where the bias of nested KQE is smaller than that of the conventional nested sample quantile.

It is also tempting to apply functional-specific approaches such as the nested sequential simulation method in Broadie et al. (2011a) for estimating the probability of large losses and then take the inverse of the estimated probability to estimate VaR. However, it should be pointed out that such an effort is highly non-trivial. To the best of our knowledge, there has not been any work in this direction, partly due to a main challenge that arises from the fact that these functional-specific approaches for the probability of large losses usually depend on a pre-specified threshold z_0, while taking the inverse involves multiple values of z_0. Despite the challenges, studies on simulation procedures tailored to the specific form of VaR to enhance sampling efficiency are highly desirable and present an interesting avenue for future research, especially given that VaR concerns the tail distribution, and sampling enhancement is crucial for the efficient use of simulation procedures in practical applications.

A related topic that deserves future research is the use of variance reduction techniques, especially importance sampling (IS), to improve the performance of the VaR estimator in the context of nested simulation. While the use of IS in the non-nested setting has been extensively studied in the literature (see, e.g., Glasserman et al. 2000), the integration of IS into nested simulation procedures appears non-trivial and deserves future research. It is also expected that the integration of functional-specific methods and/or IS into nested simulation for VaR may shed light on the estimation of CVaR, which is also a research area that has been underdeveloped.

2.2.3 Central Moments

Consider $\rho(\cdot) = \mathbb{E}g(\cdot)$ with $g(z) = (z - \mathbb{E}Y)^q$, $q \geq 2$, i.e.,

$$\alpha = \mathbb{E}(\mathbb{E}[Y|X] - \mathbb{E}Y)^q.$$

It is the qth central moment of $\mathbb{E}[Y|X]$.

Central moments are often preferred over moments about zero, particularly for the second order and higher order moments, because they offer a clearer understanding of the distribution's shape independent of its translation. The second order central moment, also known as the variance, quantifies the spread of the distribution, while the third order central moment indicates its asymmetry, with a zero value for symmetric distributions. Normalized, the third order central moment is termed skewness. The fourth order central moment gauges the heaviness of the distribution's tail relative to a normal distribution of equal variance, and its normalized version is known as kurtosis. Beyond the fourth order, higher order moments can describe additional shape parameters, but their estimation becomes more challenging, requiring larger sample sizes due to increased degrees of freedom. Both the characteristic function and the moment generating function can be approximated using power series in terms of the moments, enabling central moment estimates to be leveraged for approximating these functions (Sun et al. 2011; Cheng and Zhang 2021).

Evidently, the qth central moment can be estimated using the methods outlined in Section 2.1. However, because of its special configuration, its unbiased or bias-corrected estimators can be established in the nested simulation framework. In such cases, the estimator's bias is either 0, or its squared bias converges to 0 faster than its variance, eliminating the tradeoff between squared bias and variance. Therefore, its MSE converges to 0 at a rate of Γ^{-1}.

The first unbiased estimator of the central moment with second order, i.e., the variance of a conditional expectation, is established by Sun et al. (2011). Specifically, an ANOVA-like estimator is established for the variance of a conditional expectation, and as the simulation budget increases, the optimal inner level sample size remains bounded, and the variance (i.e., MSE) of estimator converges to 0.

In Goda (2017), a simpler method is proposed for establishing unbiased estimators of the second order central moment of a conditional expectation. The key idea is as follows:

$$\begin{aligned}
\mathbb{E}\left[(\mathbb{E}[Y|X] - \mathbb{E}Y)^2\right] &= \mathbb{E}\left(\mathbb{E}\left[Y|X^{(0)}\right]^2\right) - (\mathbb{E}Y)^2 \\
&= \mathbb{E}\left(\mathbb{E}\left[Y^{(1)}|X^{(0)}\right]\mathbb{E}\left[Y^{(2)}|X^{(0)}\right]\right) - \mathbb{E}\left(\mathbb{E}\left[Y^{(1)}|X^{(0)}\right]\right)\mathbb{E}\left(Y^{(3)}\right) \\
&= \mathbb{E}\left(\mathbb{E}\left[Y^{(1)}Y^{(2)}|X^{(0)}\right]\right) - \mathbb{E}\left(\mathbb{E}\left[Y^{(1)}|X^{(0)}\right]Y^{(3)}\right) \\
&= \mathbb{E}\left(\mathbb{E}\left[Y^{(1)}Y^{(2)}|X^{(0)}\right]\right) - \mathbb{E}\left(\mathbb{E}\left[Y^{(1)}Y^{(3)}|X^{(0)}\right]\right) \\
&= \mathbb{E}\left[Y^{(1)}\left(Y^{(2)} - Y^{(3)}\right)\right].
\end{aligned}$$

Thus the second order central moment of $\mathbb{E}[Y|X]$ can be estimated by

$$\frac{1}{n}\sum_{i=1}^{n} Y_i^{(1)}\left(Y_i^{(2)} - Y_i^{(3)}\right), \tag{20}$$

where, for each i, $X_i^{(0)}$ and $X_i^{(1)}$ are first independently and randomly sampled from F_X; $Y_i^{(1)}$ and $Y_i^{(2)}$ are then independently and randomly sampled from $F_{Y|X_i^{(0)}}$; and lastly, $Y_i^{(3)}$ is sampled from $F_{Y|X_i^{(1)}}$. It is easy to verify that (20) is an unbiased estimator.

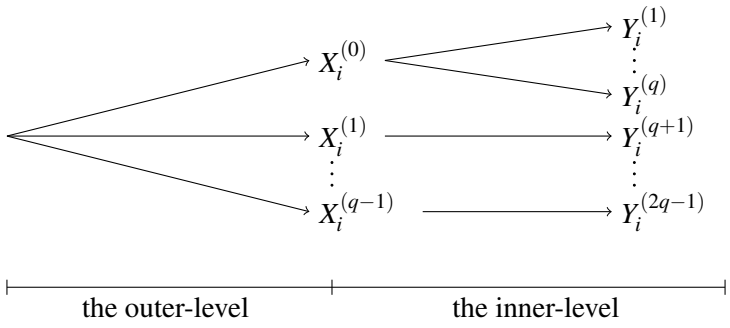

Figure 3: Illustration of the unbiased non-nested simulations for the ith group.

Similarly, Cheng and Zhang (2021) propose three unbiased estimators for the higher-order central moments of $\mathbb{E}[Y|X]$. In particular, one estimator is proposed based on

$$\mathbb{E}\left[(\mathbb{E}[Y|X]-\mathbb{E}Y)^q\right]=\mathbb{E}\left[\prod_{l=1}^{q}Y^{(l)}+\sum_{k=1}^{q-2}\binom{q}{k}(-1)^k\prod_{l=1}^{q-k}Y^{(l)}\prod_{j=q+1}^{q+k}Y^{(j)}+(-1)^q(1-q)Y^{(1)}\prod_{j=q+1}^{2q-1}Y^{(j)}\right].$$

As shown in Figure 3, for the ith group of samples, $\left\{X_i^{(k)}\right\}_{k=0}^{q-1}$ is first sampled independently and randomly from F_X. $\left\{Y_n^{(l)}\right\}_{l=0}^{q}$ is then sampled independently and randomly from $F_{Y|X_i^{(0)}}$. Lastly, $Y_i^{(q+j)}$ is sampled from $F_{Y|X_i^{(j)}}$ for $j\in\{1,2,...,q-1\}$. Therefore, a corresponding unbiased estimator is

$$\frac{1}{n}\sum_{i=1}^{n}\left[\prod_{l=1}^{q}Y_i^{(l)}+\sum_{k=1}^{q-2}\binom{q}{k}(-1)^k\prod_{l=1}^{q-k}Y_i^{(l)}\prod_{j=q+1}^{q+k}Y_i^{(j)}+(-1)^q(1-q)Y_i^{(1)}\prod_{j=q+1}^{2q-1}Y_i^{(j)}\right].$$

For the other two unbiased estimators, interest readers may refer to Cheng and Zhang (2021). Furthermore, Cheng and Zhang (2021) provide proofs for the asymptotic MSE of the estimators and establish the central limit theorem.

In the nested simulation framework, achieving unbiased estimation for general functionals in (1) remains an open problem. The approaches proposed by Rhee and Glynn (2015) and Blanchet et al. (2015) may provide potential avenues to address this challenge.

3 INTERVAL ESTIMATION

In the previous section, we have introduce various point estimators for (1). In addition to the point estimator, one may also be interested in quantifying the errors associated with the point estimate. To this end, a typical approach is through the construction of confidence intervals (CIs) with commonly used significance levels such as 90%, 95%, and 99%. By providing a range of plausible values for the quantity of interest, CIs quantify uncertainty and incorporate variability in the estimation. This allows for a more comprehensive assessment of precision and facilitating statistical inference.

While most literature primarily concentrates on developing point estimators for nested simulation (1) to expedite the convergence rate of the MSE or reduce variances, a few works in the literature focus on interval estimation for the nested simulation problem. In this section, we review a few methods for constructing CIs for (1), primarily based on the SNS technique and metamodels.

3.1 General-Purpose Approaches

While several approaches for estimating (1) have been developed, as outlined in Section 2.1, the construction of CIs for (1) in general cases has been primarily explored using the SNS and GNS methods.

3.1.1 Standard Nested Simulation

In this part, we review the SNS technique for constructing CIs. The construction of CI for (1) via SNS can be traced back to Lee (1998), where the distribution of $\mathbb{E}[Y|X]$ is assumed to be normal. Specifically, central limit theorems (CLTs) for probability of large losses and VaR are proved, and then their CIs are provided, whose forms are the same as (22). Furthermore, Lee and Glynn (2003) prove a CLT for the SNS estimator of the probability of large losses in the case that X is a discrete random variable. Based on the CLT, a CI is established for probability of large losses, and the convergence rate of the CI width is $(\log\Gamma/\Gamma)^{1/2}$.

Under similar mild conditions as those in Gordy and Juneja (2010), Cheng et al. (2022) prove CLTs for (1) when ρ takes forms as in ($\mathscr{A}$.1) $\sim$ ($\mathscr{A}$.5). The CLT has the form

$$\sqrt{n}\,(\alpha_{m,n}-\alpha) \Rightarrow \lambda A + \sqrt{C}\mathscr{N}(0,1), \tag{21}$$

when $\sqrt{n}/m \to \lambda$ as $m,n \to \infty$, where λ is a constant, and the constants A and C are the same as those in (9). In particular, (9) indicates that A and C are the bias and variance constants, respectively. Therefore, the CLT (21) stems from the tradeoff between the squared bias and variance by balancing the sample sizes n and m.

The CLT (21) can be used to construct CIs. The primary challenge lies in selecting λ. Cheng et al. (2022) propose a clever solution with the idea of minimizing the MSE of the point estimation. Specifically, considering that the budget allocation policy (n^*,m^*) in (11) minimizes the MSE, we set $\lambda = \sqrt{n^*}/m^* = \sqrt{C/2}/|A|$. Surprisingly, the CLT (21) reduces to

$$\sqrt{n}\,(\alpha_{m,n}-\alpha) \Rightarrow \sqrt{C/2}\mathrm{sign}(A) + \mathscr{N}(0,C)\,. \tag{22}$$

This selection of λ ensures that the point estimator has the minimum asymptotic MSE, as well as the CLT (22) holds. It is worth noting that calculating A, the bias constant, is "notoriously difficult" (Broadie et al. 2011a). Therefore, one advantage of (22) is that we only need to determine the sign of A rather than its precise value, which is significantly simpler.

Based on the CLT (22), Cheng et al. (2022) establish CIs for (1), with the property that the MSE of the point estimator and the width of the CI converge at their respective optimal rates. Therefore, the established CI is called *unified CI*. In particular, the UCI with $100(1-\gamma)\%$ level is

$$\left(\alpha_{m,n}+z_{\gamma/2}\zeta_{m,n}/\sqrt{n},\ \alpha_{m,n}-z_{\gamma/2}\zeta_{m,n}/\sqrt{n}\right) - (\sqrt{2}/2)\zeta^{(k)}_{m,n}/\sqrt{n}\mathrm{sign}(A), \tag{23}$$

where $\zeta^2_{m,n}$ is the estimator of C, and $z_{\gamma/2}$ is the $\gamma/2$ quantile of the standard normal distribution. Cheng et al. (2022) propose algorithms to calculate $\alpha_{m,n}$ and $\zeta_{m,n}$, so that the MSE of $\alpha_{m,n}$ and the width of the CI converge at their respective optimal rates.

In addition, Zhu et al. (2020) also derive the same CLT as (22) when ρ is VaR and CVaR. However, the CI is constructed based on a degenerated CLT, i.e., (21) with $\lambda = 0$.

3.1.2 Green Nested Simulation

The GNS point estimator for (1) has been introduced in Section 2.1.5. In this part, we further introduce the GNS CI for (1). In particular, Zhang et al. (2022) decompose $\alpha_{m,n}-\alpha$ into two terms: the first one being a or nearly a U-statistic, and the second one a reminder term. Then, a CLT is established for the first term based on the asymptotic normality of U-statistics, and further demonstrate that the reminder term converges to 0 much faster than the first term. Finally, when ρ assumes forms such as those in ($\mathscr{A}$.1) through ($\mathscr{A}$.3), the CLT is obtained as follows.

$$\frac{\alpha_{m,n}-\alpha}{\sigma_{mn}} \to \mathscr{N}(0,1),\ \text{as}\ \min\{m,n\}\to\infty,$$

where $\sigma_{mn}^2 = \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}$, $\sigma_1^2 = \mathrm{Var}\left[g\left(L\right)\right]$, and $\sigma_2^2 = \mathrm{Var}\left[\mathbb{E}\left[g'\left(L\right)H(X,\xi)\frac{f(\xi|X)}{\tilde{f}(\xi)}|\xi\right]\right]$. When $g(\cdot)$ is an indicator function, $g'(\cdot)$ in σ_2^2 does not exist. Instead a surrogate for $g'(\cdot)$ is established.

Then consistent estimators of σ_1^2 and σ_2^2 for the cases of ($\mathscr{A}$.1) and ($\mathscr{A}$.2) are constructed:

$$\widehat{\sigma}_{1,mn}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(g\left(\hat{L}(X_i)\right)\right)^2 - \left(\frac{1}{n}\sum_{i=1}^{n} g\left(\hat{L}(X_i)\right)\right)^2,$$

$$\widehat{\sigma}_{2,mn}^2 = \frac{1}{m}\sum_{j=1}^{m}\left(\frac{1}{n}\sum_{i=1}^{n} g'\left(\hat{L}(X_i)\right)H(X_i,\xi_j)\frac{f(\xi_j|X_i)}{\tilde{f}(\xi_j)}\right)^2 - \left(\frac{1}{n}\sum_{i=1}^{n} g'\left(\hat{L}(X_i)\right)\hat{L}(X_i)\right)^2.$$

Then, the following is an asymptotically valid confidence interval for the nested estimator α with a confidence level of $1-\gamma$:

$$(\alpha_{mn} - z_{1-\gamma/2}\cdot\widehat{\sigma}_{mn},\ \alpha_{mn} + z_{1-\gamma/2}\cdot\widehat{\sigma}_{mn}),$$

where $\widehat{\sigma}_{mn}^2 = \frac{\widehat{\sigma}_{1,mn}^2}{n} + \frac{\widehat{\sigma}_{2,mn}^2}{m}$ and $z_{1-\gamma/2}$ is the $1-\gamma/2$ quantile of the standard normal distribution.

It is worth noting that in the GNS method, setting $n = m = \Gamma$ is feasible. In this scenario, the convergence rate of the CI width is $\Gamma^{-1/2}$.

3.2 Functionals-Specific Approaches

A notable work on the construction of CIs for specific forms of the functional in (1) is Lan et al. (2010), which focuses on CIs for CVaR. Their method integrates the ranking and selection techniques into the nested simulation procedure, and ulitizes the fact that only outer-level scenarios located on the tail distribution of $L(X)$ matter in the estimation. To leverage the latter observation, a ranking-and-selection procedure is proposed to screening the scenarios using a portion of the samples, so that a large number of scenarios that are unlikely to locate on the tail can be discarded. Then, the remaining samples will be allocated to the surviving scenarios, thus enhancing sampling efficiently. Construction of the resulting CIs is based on the empirical likelihood tool, which assigns significance levels to the surviving scenarios through Bonferroni inequality. However, due to the use of the Bonferroni inequality, the resulting CI is usually conservative, which has also been observed in numerical studies.

Recently, Lai et al. (2024) construct a CI for CVaR via the least-squares metamodel. In particular, they first develop lower and upper bounds of CVaR as follows.

$$\inf_{u\in\mathbb{R}}\left\{u + \frac{1}{1-\beta}\mathbb{E}\left[(Y-u)\mathbb{1}\{h(X)\geq u\}\right]\right\} = \alpha^{\mathrm{low}} \leq \alpha \leq \alpha^{\mathrm{up}} = \inf_{u\in\mathbb{R}}\left\{u + \frac{1}{1-\beta}\mathbb{E}\left[(Y-\Delta-u)^+\right]\right\},$$

which holds for any approximating function $h(\cdot)$ and martingale difference Δ. Furthermore, they provide the following guarantees.

$$0 \leq \alpha - \alpha^{\mathrm{low}} \leq \frac{1}{1-\beta}\mathbb{E}\left|L(X) - h(X)\right|, \quad \text{and} \quad 0 \leq \alpha^{\mathrm{up}} - \alpha \leq \frac{1}{1-\beta}\mathbb{E}\left|Y - L(X) - \Delta\right|.$$

This suggests that a reasonable way of construction is to make $h(\cdot)$ and Δ as close to $L(\cdot)$ and $Y - L(X)$ as possible.

Then they approximate $h(\cdot)$ and Δ via the least-squares metamodel, and propose Monte Carlo estimators for the lower bound α^{low} and upper bound α^{up}. Moreover, they establish CLTs for these two estimators, and based on CLTs, CIs for the lower and upper bounds can be constructed. Then, the lower end of the CI for the lower bound and the upper end of the CI for the upper bound together form a CI of CVaR with justifiable statistical guarantee, which accounts for both the metamodel error and the noises of Monte Carlo samples. The proposed CI procedure reuses the samples simulated for LSM point estimation, thus requiring no additional simulation budget.

4 CONCLUSIONS

In this tutorial, we delve into the nested simulation problem, investigating various estimation approaches from the perspectives of point estimation and interval estimation, and examining their contributions within the field. While existing literature predominantly concentrates on establishing point estimators for nested simulation, the exploration of interval estimation corresponding to each point estimator emerges as a potential future research avenue.

Moreover, we have identified several promising future research directions. The first involves quantifying model error when employing LSM-like metamodels to estimate the conditional expectation. This open problem, highlighted by Broadie et al. (2015), is also critical for metamodel methods in the simulation domain. Wang et al. (2017) propose a method to test for metamodel misspecification, but this approach does not quantify the model error.

The second direction involves developing bias reduction and variance reduction techniques for nested simulation. Broadie et al. (2011a) propose the nested sequential simulation method to estimate the probability of large losses, efficiently reducing bias compared to the conventional SNS estimator. This improvement enhances the MSE of nested sequential simulation to $O(\Gamma^{-4/5+\delta})$ for any $\delta > 0$. Consequently, a natural question arises: how can the idea of nested sequential simulation be applied to estimate other risk measures, such as VaR and CVaR? Another pertinent question is whether the rate of $\Gamma^{-4/5+\delta}$ is the optimal one achievable under the nested sequential simulation approach. In addition, bias reduction techniques, such as FAST and two-fold FAST (2-FAST), have been proposed by Liang et al. (2024). In-depth investigation of bias and variance reduction techniques may be a future research direction.

ACKNOWLEDGMENTS

The research of the first author was supported partially by the National Natural Science Foundation of China (NNSFC) and the Research Grants Council of Hong Kong (RGC-HK), under the RGC-HK General Research Fund Project 11508620, and NSFC/RGC-HK Joint Research Scheme under project N_CityU 105/21. This research of the second author was supported by NNSFC grants 72101260.

REFERENCES

Ankenman, B., B. L. Nelson, and J. Staum. 2008. "Stochastic Kriging for Simulation Metamodeling". In *Proceedings of the 2015 Winter Simulation Conference*, 362–370 https://doi.org/10.1109/WSC.2008.4736089.

Ankenman, B., B. L. Nelson, and J. Staum. 2010. "Stochastic Kriging for Simulation Metamodeling". *Operations Research* 58(2):371–382.

Blanchet, J. H., N. Chen, and P. W. Glynn. 2015. "Unbiased Monte Carlo Computation of Smooth Functions of Expectations via Taylor Expansions". In *Proceedings of the 2015 Winter Simulation Conference*, 360–367 https://doi.org/10.1109/WSC.2015.7408178.

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

Broadie, M., Y. Du, and C. C. Moallemi. 2011b. "Risk Estimation via Weighted Regression". In *Proceedings of the 2011 Winter Simulation Conference*, 3854–3865 https://doi.org/10.1109/WSC.2011.6148077.

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

Cheng, H.-F., X. Liu, and K. Zhang. 2022. "Constructing Confidence Intervals for Nested Simulation". *Naval Research Logistics* 69:1138–1149.

Cheng, H.-F. and K. Zhang. 2021. "Non-nested Estimators for the Central Moments of a Conditional Expectation and their Convergence Properties". *Operations Research Letters* 49(5):625–632.

Ding, L. and X. Zhang. 2024. "Sample and Computationally Efficient Stochastic Kriging in High Dimensions". *Operations Research* 72(2):660–683.

Glasserman, P., P. Heidelberger, and P. Shahabuddin. 2000. "Variance Reduction Techniques for Estimating Value-at-Risk". *Management Science* 16(10):1349–1364.

Goda, T. 2017. "Computing the Variance of a Conditional Expectation via Non-nested Monte Carlo". *Operations Research Letters* 45(1):63–67.

Goda, T., T. Hironaka, and T. Iwamoto. 2020. "Multilevel Monte Carlo Estimation of Expected Information Gains". *Stochastic Analysis and Applications* 38(4):581–600.

Gordy, M. and S. Juneja. 2006. "Efficient Simulation for Risk Measurement in Portfolio of CDOs". In *Proceedings of the 2006 Winter Simulation Conference*, 749–756 https://doi.org/10.1109/WSC.2006.323155.

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

Härdle, W. K. 1990. *Applied Nonparametric Regression*. Cambridge University Press.

Hastie, T., R. Tibshirani, and J. Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer.

Hong, L. J. and S. Juneja. 2009. "Estimating the Mean of a Non-linear Function of Conditional Expectation". In *Proceedings of the 2009 Winter Simulation Conference*, 1223–1236 https://doi.org/10.1109/WSC.2009.5429428.

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.

Lai, Q., G. L. Liu, B. Zhang, and K. Zhang. 2024. "Simulating Confidence Intervals for Conditional Value-at-Risk via Least-squares Metamodel". *Working paper*.

Lan, H., B. L. Nelson, and J. Staum. 2010. "A Confidence Interval Procedure for Expected Shortfall Risk Measurement via Two-level Simulation". *Operations Research* 58(5):1481–1490.

Lee, S.-H. 1998. *Monte Carlo Computation of Conditional Expectation Quantiles*. PhD Thesis, Stanford University, https://www.proquest.com/openview/a2a16d4d08f02267d83d23a481915159/1?pq-origsite=gscholar&cbl=18750&diss=y.

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* 13(3):238–258.

Liang, G., K. Zhang, and J. Luo. 2024. "A FAST Method for Nested Estimation.". *INFORMS Journal on Computing* Publish Online.

Liu, M. and J. Staum. 2010. "Stochastic Kriging for Efficient Nested Simulation of Expected Shortfall". *Journal of Risk* 12(3):3–27.

Liu, X., X. Yan, and K. Zhang. 2024. "Kernel Quantile Estimators for Nested Simulation with Application to Portfolio Value-at-Risk Measurement". *European Journal of Operational Research* 312(3):1168–1177.

Rhee, C.-H. and P. W. Glynn. 2015. "Unbiased Estimation with Square Root Convergence for SDE Models". *Operations Research* 63(5):1026–1043.

Steckley, S. G., S. G. Henderson, D. Ruppert, R. Yang, D. W. Apley and J. Staum. 2016. "Estimating the Density of a Conditional Expectation". *Electronic Journal of Statistics* 10(1):736 – 760.

Sun, Y., D. W. Apley, and J. Staum. 2011. "Efficient Nested Simulation for Estimating the Variance of a Conditional Expectation". *Operations Research* 59(4):998–1007.

Wang, S., G. Liu, and K. Zhang. 2017. "A Misspecification Test for Simulation Metamodels". In *Proceedings of the 2017 Winter Simulation Conference*, 1938–1949 https://doi.org/10.1109/WSC.2017.8247929.

Wang, T., J. Xu, J.-Q. Hu, and C.-H. Chen. 2023. "Efficient Estimation of a Risk Measure Requiring Two-stage Simulation Optimization". *European Journal of Operational Research* 305(3):1355–1365.

Wang, W., Y. Wang, and X. Zhang. 2024. "Smooth Nested Simulation: Bridging Cubic and Square Root Convergence Rates in High Dimensions". *Management Science* Publish Online.

Zhang, K., B. M. Feng, G. Liu, and S. Wang. 2022. "Sample Recycling for Nested Simulation with Application in Portfolio Risk Measurement". *arXiv preprint arXiv:2203.15929*.

Zhang, K., G. Liu, and S. Wang. 2017. "Portfolio Risk Measurement via Stochastic Mesh". In *Proceedings of the 2017 Winter Simulation Conference*, 1796–1811 https://doi.org/10.1109/WSC.2017.8247917.

Zhang, K., G. Liu, and S. Wang. 2022. "Bootstrap-based Budget Allocation for Nested Simulation". *Operations Research* 70(2):1128–1142.

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

AUTHOR BIOGRAPHIES

GUANGWU LIU is a Professor in College of Business at City University of Hong Kong. His research interests include stochastic simulation, financial engineering, and risk management. His e-mail address is msgw.liu@cityu.edu.hk and his website is https://www.cb.cityu.edu.hk/staff/guanliu/.

KUN ZHANG is an Assistant Professor in the Institute of Statistics and Big Data at Renmin University of China. He holds a PhD in management science from City University of Hong Kong. His research interests include simulation optimization, machine learning, business analytics, financial engineering and risk management. His email address is kunzhang@ruc.edu.cn.