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

INPUT UNCERTAINTY QUANTIFICATION VIA SIMULATION BOOTSTRAPPING

Manjing Zhang
Yulin He

Guangdong Laboratory of
Artificial Intelligence and Digital Economy (SZ)
Kerun Building, Guangming District
Shenzhen, 518107, CHINA

Guangwu Liu

Department of Management Sciences
City University of Hong Kong
Tat Chee Avenue, Kowloon
Hong Kong, CHINA

Shan Dai

Shenzhen Research Institute of Big Data
2001 Longxiang Boulevard, Longgang District
Shenzhen, 518172, CHINA

ABSTRACT

Input uncertainty, which refers to the output variability arising from statistical noise in specifying the input models, has been intensively studied recently. Ignoring input uncertainty often leads to poor estimates of system performance. In the non-parametric setting, input uncertainty is commonly estimated via bootstrap, but the performance by traditional bootstrap resampling is compromised when input uncertainty is also associated with simulation uncertainty. Nested simulation is studied to improve the performance by taking variance estimation into account, but suffers from a substantial burden on required simulation effort. To tackle the above problems, this paper introduces a non-nested method to build asymptotically valid confidence intervals for input uncertainty quantification. The convergence properties are studied, which establish statistical guarantees for the proposed estimators related to real-data size and bootstrap budget. An easy-implemented algorithm is also provided. Numerical examples show that the estimated confidence intervals perform satisfactorily under given confidence levels.

1 INTRODUCTION

The problem of input uncertainty refers to the effect of not knowing the true, correct distributions of the basic stochastic processes that drive the simulation. Input models are often based on a finite list of the observed real-world data, which are therefore subject to error. A stochastic output usually suffers two types of uncertainties. With the availability of cheap and flexible computing power, the existing simulation softwares, which measure simulation error based on Monte-Carlo methods, have gained increasing importance in probability and statistics. On the contrary, there are rare software considering input-uncertainty errors. Few practitioners are even aware of this problem. Although the existence of input uncertainty has been put forward in the 1990s, the modelers may still ignore this source of uncertainty.

When the input distributions are obtained from the observed real-world data, it is possible to quantify the impact of input uncertainty on the output results. Therefore, they have excessive confidence in the results of the usual simulation output analysis, which is conditional on the data set used to fit the input probability distributions. Many results, such as Barton and Schruben (2001), Henderson (2003), have indicated that the error in such conditional confidence intervals can be quite large, which results in information loss to a large extent. And in most cases, the simulation outputs not only depend on the unknown input distributions but also are affected by some extra random variables whose distributions are already known, which can be massively simulated. For example, the queueing system with unknown distribution of interarrival-time and known distribution of service time, where both distributions should be considered to quantify the average waiting time. Thus, the simulation outputs suffer two kinds of uncertainties simultaneously, i.e., input uncertainty and simulation uncertainty.

A common approach to quantify input uncertainty is the bootstrap resampling method, which constructs the bootstrap empirical distributions as the input model estimators, see Barton and Schruben (1993), Barton and Schruben (2001). To capture the uncertainty in the input distributions without additional real-world samples, bootstrap technique is usually applied to create a new empirical distribution function (EDF) for use in each simulation run. In the bootstrap resampling method, the empirical distribution $\mathbf{F}_n$ based on the real-world data is firstly constructed as the input model estimator. Then the bootstrap samples of $\mathbf{F}_n$ are generated accordingly and the corresponding bootstrap empirical distributions are calculated, which will be used to quantify the input uncertainty. The bootstrap resampling method is easy to be implemented as not requiring any prior information on the input models and the simulation outputs, so it is applied prominently in the non-parametric regime as in Cheng (2001), Ankenman and Nelson (2012) and parametric regime as in Cheng and Holloand (1997), Song and Nelson (2015).

There are many different methods using bootstrap to construct confidence intervals of the simulation output for input uncertainty quantification, among which the first is called the Bootstrap Standard Interval introduced by Efron (1981). Suppose that $\hat{\theta}$ is a bootstrap point estimator for the target θ and its variance is estimated as $\hat{\sigma}^2$, then the approximate confidence interval by bootstrap can be set as $[\hat{\theta}-\hat{\sigma}z_{\alpha/2}, \hat{\theta}+\hat{\sigma}z_{\alpha/2}]$, where $z_{\alpha/2}$ is the $100-\alpha/2$ percentile of a standard normal distribution. The interval is claimed to have an approximate coverage probability of $1-\alpha$. Except for the standard confidence intervals, there are some other bootstrap confidence intervals such as percentile confidence intervals, bias-corrected confidence intervals, and bias-corrected and accelerated confidence intervals, and all of them depend upon the percentile of the bootstrap distributions. In Efron and Tibshirani (1986), the authors conducted a comparison among these different bootstrap confidence intervals mentioned above.

However, when applying the bootstrap standard confidence intervals to quantify input uncertainty, people face a few potential difficulties. One main difficulty lies in the estimation of variance, the solution of which is mainly the method of nested simulation. More precisely, to handle both the input and simulation noises, the bootstrap resampling method usually adopts nested simulation, which typically comprises a two-layer sampling procedure. The outer-level simulation resamples the input data to build the bootstrap empirical distribution as the input model estimator, while the inner-level runs simulation replications under each bootstrap resampling to estimate the corresponding conditional expectation, which could be computationally prohibitive. Lam and Qian (2022) developed a subsampling framework to improve the computational burden by the conventional bootstrap approach. The other challenge brought by nested simulation in the implementation of the bootstrap resampling is how to allocate the limited simulation budget between the outer-level size B and inner-level size k, which has a direct implication on the simulation output. Yi and Xie (2017) introduced a sequential experiment design to efficiently allocate the simulation resources to bootstrapped samples of input models. Except for the nested simulation, metamodel-assisted bootstrap is first proposed by Barton, Nelson, and Xie (2010) to quantify both input uncertainty and simulation uncertainty. However, the algorithm only works for parametric input models, and the metamodel may cause additional uncertainty.

There are many other methods to quantify input uncertainty besides the bootstrap procedure. The Bayesian method, first introduced by Chick (1997), typically emphasizes the choice of parameters for parametric distributions. In terms of the Bayesian model, averaging strategy has been described by Chick (1997), Zouaoui and Wilson (2003). Zouaoui and Wilson (2004) considered the case with unknown input models and quantified the simulation output variance due to simulation uncertainty, parameter uncertainty, and model uncertainty. Biller and Gunes (2010) provided a Bayesian approach that can be applied to multivariate input models, and Biller and Corlu (2011) extended the Bayesian framework to handle a large number of correlated inputs. The delta method, first discussed in Cheng (1994), was further studied in Cheng and Holloand (1997), Cheng and Holland (2004). It applies the sampling distribution by the maximum likelihood estimates of the input parameters to quantify the parameter uncertainty. Lam and Qian (2019) studied the use of the delta method in estimating non-parametric input variance. The robust optimization approach to input uncertainty was further studied and is still a very active research area, see Lam and Qian (2016), Ghosh and Lam (2019). However, the responses rely on the selection of distance measures and problem constraints, and cannot be directly applied to non-convex cases.

Motivated by the non-nested variance estimators proposed by Goda (2017), where the authors provided several non-nested Monte Carlo estimators for the variance of a conditional expectation, we adopt the non-nested variance estimator for the input uncertainty problem. In this paper we implement the bootstrap resampling method for interval construction and propose the new non-nested bootstrap estimators when input uncertainty and simulation uncertainty co-exist, which avoids the substantial burden on the required simulation effort caused by nested simulation. We also provide an algorithm for estimating the mean and variance estimators and constructing asymptotic confidence intervals accordingly to quantify the input uncertainty. Moreover, the theoretical results about the relationship between the size of real-world data n and the bootstrap times B are also provided to verify the convergence properties. The numerical results demonstrate that the proposed method works effectively.

The remainder of this paper is organized as follows. Section 2 describes the formulation of the input uncertainty quantification problem. Section 3 introduces how to build confidence intervals for the input uncertainty quantification problem. Section 4 reports related experimental results. Finally, a conclusion is drawn in Section 5.

2 NON-PARAMETRIC INPUT UNCERTAINTY QUANTIFICATION

Throughout this paper, the following notations are used. The notations '$\xrightarrow{p}$', '$\xrightarrow{d}$' and '$\xrightarrow{a.s.}$' signify convergence in probability, convergence in distribution, and convergence with probability one, respectively.

2.1 Problem Formulation

The case that there is no information available about the input distribution families except for a limited amount of real data is primarily considered, and the input uncertainty is measured in a non-parametric way. The aim is to estimate confidence intervals for the mean response of a system that is represented by a stochastic simulation, of which the input models are unknown but can be estimated by real-world data.

All the simulation output can be summarized as depending on two types of random variables, i.e., one with unknown distributions but a limited amount of data, and the other with known distributions. In this paper, we mainly consider a 1-dimensional simulation output $h(\mathbf{X},\mathbf{U})$ with an S_1-dimensional input variable $\mathbf{X}^T = (X_1, X_2, \cdots, X_{S_1})$, where the joint probability distribution $\mathbf{F}$ for $\mathbf{X}$ is unknown, and an S_2-dimensional input variable $\mathbf{U}^T = (U_1, U_2, \cdots, U_{S_2})$, where the joint probability distribution $\mathbf{G}$ for $\mathbf{U}$ is known. This paper estimates the expectation value of $h(\mathbf{X},\mathbf{U})$, which is essentially a mapping of the unknown distribution $\mathbf{F}$. Denote $T(\mathbf{F})$ as

$$T(\mathbf{F}) = \mathbb{E}[h(\mathbf{X},\mathbf{U})].$$

Notably, the output $h(\mathbf{X},\mathbf{U})$ is a general representation of the simulation output of the random variables $\mathbf{X}$ and $\mathbf{U}$, for instance, it can be an indicator, a quantile, a variance and so on.

Moreover, as the distribution for $\mathbf{U}$ is known, $T(\mathbf{F})$ can be written as the expectation of a particular mapping of the random variable $\mathbf{X}$, i.e.,

$$T(\mathbf{F}) = \mathbb{E}[q(\mathbf{X})],$$

where

$$q(\mathbf{X}) = \mathbb{E}_{\mathbf{U}}[h(\mathbf{X},\mathbf{U})|\mathbf{X}].$$

Extra simulation is always necessary to calculate $q(\mathbf{x})$ by generating different values of $\mathbf{U}$ under a fixed realized value of $\mathbf{X}$. The distribution of $\mathbf{X}$ is unknown, which brings into the input uncertainty, while the distribution of $\mathbf{U}$ is known, which still leads to the simulation uncertainty. Therefore, the total variation for $T(\mathbf{F})$ estimation comes from the joint uncertainty composed by the simulation uncertainty when estimating $q(\mathbf{X})$ for each fixed value of $\mathbf{X}$, and the input uncertainty when estimating the expectation by using real-world data. In this paper, we mainly focus on building confidence intervals for estimating $T(\mathbf{F})$ to quantify the uncertainty caused by the limited sample of real-world data with size n.

2.2 Traditional Bootstrap Quantification Framework for the Pure Input Uncertainty Cases

This part propagates input uncertainty from the input model to the simulation output by using bootstrap resampling of the real-world data.

Firstly, the unknown distribution $\mathbf{F}$ can be estimated by the observed real-world data $\{\mathbf{x}_1,\mathbf{x}_2,\cdots\mathbf{x}_n\}$ in a non-parametric manner, where $\mathbf{x}_i^T = (x_{i1}, x_{i2},\cdots,x_{iS_1})$ and $\mathbf{x}_i$'s are independently and identically distributed (i.i.d.) observations following the probability distribution $\mathbf{F}$.

We estimate the joint input distribution by its empirical cumulative distribution function (ecdf) $\mathbf{F}_n$, where

$$\mathbf{F}_n(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} I_{\{\mathbf{X}_i\leq\mathbf{x}\}}.$$

Then $T(\mathbf{F}_n)$ is an estimator of $T(\mathbf{F})$ and can be expressed as

$$T(\mathbf{F}_n) = \mathbb{E}[q(\mathbf{X})|\mathbf{x}_1,\mathbf{x}_2,\cdots\mathbf{x}_n] = \frac{1}{n}\sum_{i=1}^{n} q(\mathbf{x}_i).$$

$T(\mathbf{F}_n)$ is a mapping of the ecdf $\mathbf{F}_n$, which represents the expectation of $q(\mathbf{X})$ when $\mathbf{X}$ follows the ecdf $\mathbf{F}_n$. In the following, we denote $\mathbb{E}^*$ as taking the expectation under the probability measure conditional on the real-world data $\mathbf{x}_1,\mathbf{x}_2,\cdots\mathbf{x}_n$, i.e.,

$$\mathbb{E}^*[Y] = \mathbb{E}[Y|\mathbf{x}_1,\mathbf{x}_2,\cdots\mathbf{x}_n],$$

such that

$$T(\mathbf{F}_n) = \mathbb{E}^*[q(\mathbf{X})].$$

Secondly, to analyze how $T(\mathbf{F}_n)$ changes with respect to $\mathbf{F}_n$, the most commonly used method involves generating many ecdfs with different sets of data and then calculating the corresponding different $T(\mathbf{F}_n)$'s. However, with highly limited data, the method mentioned above will suffer from substantial error if one divides the limited data into several parts for calculation. An alternative to estimating the empirical distributions is to apply resampling methods. The bootstrap, introduced by Efron (1979), is one of the most popular techniques adopted in resampling to measure the variability of a given statistic, which makes full use of the whole real-world data.

Given the data $\{\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_n\}$, a bootstrap sample is defined to be a random sample of the same size n drawn from $\mathbf{F}_n$, say $\{\mathbf{x}_1^*,\mathbf{x}_2^*,\cdots,\mathbf{x}_n^*\}$. The star notation indicates that the bootstrap sample $\{\mathbf{x}_1^*,\mathbf{x}_2^*,\cdots,\mathbf{x}_n^*\}$ is not the actual data set $\{\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_n\}$, but rather a randomized, or resampled version of the real data. Then we can calculate the corresponding empirical cumulative distribution function from $\{\mathbf{x}_1^*,\mathbf{x}_2^*,\cdots,\mathbf{x}_n^*\}$ as

$$\mathbf{F}_n^*(\mathbf{x}) = \frac{1}{n}\sum_{j=1}^{n} I_{\{\mathbf{x}_j^*\leq\mathbf{x}\}}.$$

$\mathbf{F}_n^*$ is called a bootstrap empirical cumulative distribution function (bootstrap ecdf). Then the estimator based on this bootstrap ecdf can be denoted as

$$T(\mathbf{F}_n^*) = \mathbb{E}[q(\mathbf{X})|\mathbf{x}_1^*, \mathbf{x}_2^*, \cdots, \mathbf{x}_n^*] = \frac{1}{n}\sum_{j=1}^{n} q(\mathbf{x}_j^*).$$

$T(\mathbf{F}_n^*)$ refers to the expectation of $q(\mathbf{X})$, in which $\mathbf{X}$ follows the bootstrap ecdf $\mathbf{F}_n^*$.

After bootstrapping B times, there will be B different bootstrap samples, where the i-th bootstrap sample is denoted as $\{\mathbf{x}_1^{*i}, \mathbf{x}_2^{*i}, \cdots, \mathbf{x}_n^{*i}\}$. For the B different bootstrap samples, the bootstrap ecdfs can be constructed as $\{\mathbf{F}_n^{*1}, \mathbf{F}_n^{*2}, \cdots, \mathbf{F}_n^{*B}\}$. For a bootstrap ecdf $\mathbf{F}_n^{*i}$, the corresponding estimator for $T(\mathbf{F})$ should be

$$T(\mathbf{F}_n^{*i}) = \mathbb{E}[q(\mathbf{X})|\mathbf{x}_1^{*i}, \mathbf{x}_2^{*i}, \cdots, \mathbf{x}_n^{*i}] = \frac{1}{n}\sum_{j=1}^{n} q(\mathbf{x}_j^{*i}).$$

Substituting the B bootstrap ecdfs, one can analyze how $T(\mathbf{F}_n^{*i})$'s changes with respect to $\mathbf{F}_n^{*i}$'s to quantify the input uncertainty. In the following, we denote $\mathbb{E}^{*i}$ as taking the expectation under the probability measure conditional on the i-th bootstrap sample, i.e.,

$$\mathbb{E}^{*i}[Y] = \mathbb{E}[Y|\mathbf{x}_1^{*i}, \mathbf{x}_2^{*i}, \cdots, \mathbf{x}_n^{*i}],$$

and the equation

$$T(\mathbf{F}_n^{*i}) = \mathbb{E}^{*i}[q(\mathbf{X})]$$

can be similarly derived.

Therefore, the different $T(\mathbf{F}_n^{*i})$'s are the estimators of $T(\mathbf{F})$. To establish the confidence intervals for $T(\mathbf{F})$ by $T(\mathbf{F}_n^{*i})$'s, it is necessary to figure out the difference between $T(\mathbf{F}_n^{*i})$'s and $T(\mathbf{F})$, which is the major concern studied in this paper.

2.3 Bootstrap Quantification Frameworks for the Joint Uncertainty Cases

2.3.1 Nested Quantification Framework

However, as we discussed previously, the exact form of $q(\mathbf{X})$ can not be obtained directly in many cases, and in our setting, recall that

$$q(\mathbf{x}) = \mathbb{E}_{\mathbf{U}}[h(\mathbf{X}, \mathbf{U})|\mathbf{X} = \mathbf{x}].$$

Nested simulation is typically applied in this situation. In a two-level nested simulation, an outer level of simulation samples different scenarios, while the inner level uses simulation to estimate a conditional expectation given the scenario. Specifically, extra simulation is needed to estimate $q(\mathbf{x})$ by generating a lot of samples of $\mathbf{U}$ in this setting, thus an additional error by simulation is also introduced to analyze $T(\mathbf{F})$, i.e., the extra error when we estimate $q(\mathbf{x})$ by simulating $\mathbf{U}$. Then the estimator of $q(\mathbf{x})$ can be constructed as $\frac{1}{k}\sum_{l=1}^{k} h(\mathbf{x}, \mathbf{u}_l)$.

Unfortunately, Sun, Apley, and Staum (2011) claimed that the nested variance estimator can be badly biased unless the inner level sample size k is large enough. In fact, the bias of the variance estimator for a two-level nested procedure is $O(k^{-1})$. This further exemplifies a typical situation in the two-level nested simulation: for its mean-squared error to converge to zero, it is necessary that both the outer sample size and inner level sample size go to infinity. Therefore, the two-level nested simulation can be extremely time-consuming. Moreover, the relationship between the outer-level sample and the inner-level sample size is hard to derive. Sun, Apley, and Staum (2011) also provided an unbiased nested Monte Carlo estimator by a pre-computation step to choose proper inner-level sample sizes (each scenario may have a different number of inner-level samples), which is referred to as $1\frac{1}{2}$-level simulation. However, the pre-computation is complex, depending on the specific problem and a given computational cost.

To resolve the problem, this paper introduces a non-nested method to build the asymptotically valid confidence intervals for the joint uncertainty quantification and discusses the convergence conditions for the relationship between bootstrap times B and real-world data size n.

2.3.2 Non-nested Quantification Framework

Goda (2017) constructed several non-nested Monte Carlo estimators for the variance of a conditional expectation. Let X be a random variable with probability density function p_X defined on Ω_X, and $f:\Omega_X \to \mathbb{R}$ be a function. Let Y be another random variable which is correlated with X, then they built the following estimators V_1, V_2, and V_3 for computing the variance of a conditional expectation. The estimators for $\mathrm{Var}_Y(\mathbb{E}_{X|Y}[f(X)])$ are:

$$V_1 = \frac{1}{n}\sum_{i=1}^{n} f(x_i)f(x_i') - \hat{\mu}^2,$$

$$V_2 = \frac{1}{n}\sum_{i=1}^{n} f(x_i)f(x_i') - (\frac{\hat{\mu}+\hat{\mu}'}{2})^2,$$

$$V_3 = \frac{1}{n}\sum_{i=1}^{n} f(x_i)f(x_i') - \hat{\mu}\hat{\mu}',$$

where $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} f(x_i)$, $\hat{\mu}' = \frac{1}{n}\sum_{i=1}^{n} f(x_i')$, and x_i and x_i' independently and randomly sample from $X|Y = y_n$. Goda (2017) proved that all of the estimators V_1, V_2 and V_3 are biased, and the bias of V_3 is smaller or equal to that of V_2, which itself is smaller or equal to that of V_1. The bias for every realized estimator decays at the rate n^{-1}, so that the bias is negligible for large n. A bias correction method of the estimators is also provided.

Inspired by Goda (2017), we proposed our non-nested quantification framework based on the observation that estimating the variance of $T(\mathbf{F}_n^{*i})$ can be also regarded as a variance estimation problem of a conditional expectation essentially. The non-nested estimators can be modified to solve the input uncertainty quantification problem then. Since the estimators are no longer of the nested form, there is no need to take care of a proper choice of inner-level sample sizes. Here we construct the estimators using V_2 in our setting. And in fact, the estimators based on V_1 or V_3 can be implemented similarly.

The non-nested mean and variance estimators for $q(\mathbf{X})$ should be

$$\bar{m} = \sum_{i=1}^{B} m_i/B,$$

$$\hat{\sigma}^2 = n\left[\frac{1}{B}\sum_{i=1}^{B} m_{i1}\cdot m_{i2} - \left(\frac{1}{2B}\sum_{i=1}^{B}(m_{i1}+m_{i2})\right)^2\right],$$

where

$$m_{i1} = \frac{1}{n}\sum_{j=1}^{n} h(\mathbf{x}_j^{*i}, \mathbf{u}_j^{i1}),$$

$$m_{i2} = \frac{1}{n}\sum_{j=1}^{n} h(\mathbf{x}_j^{*i}, \mathbf{u}_j^{i2}),$$

$$m_i = (m_{i1}+m_{i2})/2.$$

and $\{\mathbf{u}_1^{i1},\cdots,\mathbf{u}_n^{i1},\mathbf{u}_1^{i2},\cdots,\mathbf{u}_n^{i2}\}$ are i.i.d. following the known distribution $\mathbf{G}$. Here m_i's are regarded as the mean estimators of $T(\mathbf{F}_n^{*i})$'s. Goda (2017) demonstrated that constructing Monte Carlo estimators require two function evaluations for each bootstrap sample. By using two conditionally independent estimators for $T(\mathbf{F}_n^{*i})$, the bias of the non-nested variance estimator should be $O(B^{-1})$ and the total computational budget equals $2B$. Therefore, we can evaluate input uncertainty by constructing confidence intervals with the proposed mean and variance estimators.

Next, we will provide the theoretical and numerical instructions for the estimated confidence intervals. Moreover, we will also discuss the relationship between the size of real-world data n and the bootstrap times B to verify the convergence properties for the estimators.

3 CONFIDENCE INTERVALS CONSTRUCTION

3.1 The Convergence Property for $\bar{m}$

Assumption 1 $\mathbb{E}|q(\mathbf{X})|^2 < \infty$.

Assumption 1 ensures that $\mathrm{Var}(T(\mathbf{F}))$ is finite, which guarantees that $\sqrt{n}(T(\mathbf{F}_n) - T(\mathbf{F}))$ admits a Central Limit Theorem (CLT). This assumption yields the following propositions.

Proposition 1 Given $\{\mathbf{X}_1,\cdots,\mathbf{X}_n\}$, suppose that there are B conditionally independent bootstrap estimators, denoted as $\{T(\mathbf{F}_n^{*1}), T(\mathbf{F}_n^{*2}),\cdots,T(\mathbf{F}_n^{*B})\}$, then under Assumption 1,

$$\sqrt{n}(\bar{T}(\mathbf{F}_n^*) - T(\mathbf{F})) \xrightarrow{d} N(0,\sigma^2(\mathbf{F})),$$

as $n\to\infty$, where $\sigma^2(\mathbf{F})$ means the variance of $q(\mathbf{X})$ when $\mathbf{X}$ follows $\mathbf{F}$, and

$$\bar{T}(\mathbf{F}_n^*) = \frac{1}{B}\sum_{i=1}^{B} T(\mathbf{F}_n^{*i}).$$

Interested readers may refer to Bickel and Freedman (1981) and Singh (1981) for the main proof of proposition 1. Next, we discuss the convergence property of $\bar{m}$, and it's necessary to make a restriction on $h(\mathbf{X},\mathbf{U})$.

Assumption 2 $\mathbb{E}|h(\mathbf{X},\mathbf{U})|^2 < \infty$.

Then we can conclude the following theorem.

Theorem 1 Under Assumption 2, as $B,n\to\infty$,

$$\sqrt{n}(\bar{m} - \bar{T}(\mathbf{F}_n^*)) \xrightarrow{p} 0.$$

Consequently, we can obtain that

$$\sqrt{n}(\bar{m} - T(\mathbf{F})) \xrightarrow{d} N(0,\sigma^2(\mathbf{F})),$$

as $B,n\to\infty$.

3.2 The Convergence Property for $\hat{\sigma}^2$

Under the following assumption, we can obtain the convergence result of the variance estimators.

Assumption 3 $\mathbb{E}|h(\mathbf{X},\mathbf{U})|^4 < \infty$.

Theorem 2 The variance estimator for $q(\mathbf{X})$ is

$$\hat{\sigma}^2 = n\left[\frac{1}{B}\sum_{i=1}^{B} m_{i1}\cdot m_{i2} - \bar{m}^2\right],$$

and under Assumption 3, when $B/n \to \infty$,

$$\hat{\sigma}^2 \xrightarrow{p} \sigma^2(\mathbf{F}),$$

as $n \to \infty$.

For the input uncertainty problem associated with simulation uncertainty, the proposed non-nested method requires $B/n \to \infty$ to achieve consistency, such that the total computational budget should be $2B$. Compared to the nested estimators, the bias of the nested variance estimator should be $O(k^{-1})$ and the total computational budget equals to $2Bk$, where $B, k \to \infty$.

So far, we propose the mean and the variance estimators for input uncertainty quantification that contains input and simulation uncertainty together, and establish their convergence property under regular conditions. To ensure their consistency, the bootstrap times B must satisfy that $B/n \to \infty$, where n represents the real-world data size. The following algorithm is further proposed to build bootstrap confidence intervals in practice based on the previous results.

Algorithm 1 Non-nested Bootstrap (NNB) Confidence Intervals.

1: **Input:** The real-world data $\{\mathbf{x}_1, \mathbf{x}_2, \cdots \mathbf{x}_n\}$, the bootstrap times B, and the fixed confidence level α.
2: **Output:** The estimated confidence interval for $T(\mathbf{F})$.
3: **Estimate F.** Calculate the empirical distribution $\mathbf{F}_n$ as an estimator of $\mathbf{F}$.

$$\mathbf{F}_n(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} I_{\{\mathbf{x}_i \le \mathbf{x}\}}$$

4: **for** $i = 1$ to B **do**
5: Generate a bootstrap sample $\{\mathbf{x}_1^{*i}, \mathbf{x}_2^{*i}, \cdots, \mathbf{x}_n^{*i}\}$ from the empirical distribution $\mathbf{F}_n$, and then compute the corresponding bootstrap empirical distribution $\mathbf{F}_n^{*i}$ as an estimator of $\mathbf{F}_n$.

$$\mathbf{F}_n^{*i}(\mathbf{x}) = \frac{1}{n}\sum_{j=1}^{n} I_{\{\mathbf{x}_j^{*i} \le \mathbf{x}\}}$$

6: Simulate $\mathbf{u}_1^{i1}, \cdots, \mathbf{u}_n^{i1}, \mathbf{u}_1^{i2}, \cdots, \mathbf{u}_n^{i2}$ from the known distribution $\mathbf{G}$.

$$m_{i1} = \frac{1}{n}\sum_{j=1}^{n} h(\mathbf{x}_j^{*i}, \mathbf{u}_j^{i1}),\ m_{i2} = \frac{1}{n}\sum_{j=1}^{n} h(\mathbf{x}_j^{*i}, \mathbf{u}_j^{i2})$$

7: **end for**
8: Calculate the mean estimator $\bar{m}$,

$$\bar{m} = \sum_{i=1}^{B} (m_{i1} + m_{i2})/2B.$$

9: Calculate the variance estimator $\hat{\sigma}^2$,

$$\hat{\sigma}^2 = n\Big(\sum_{i=1}^{B} m_{i1} \cdot m_{i2}/B - \bar{m}^2\Big).$$

10: Calculate the confidence interval,

$$[\bar{m} - \frac{\hat{\sigma}}{\sqrt{n}} z_{\alpha/2},\ \bar{m} + \frac{\hat{\sigma}}{\sqrt{n}} z_{\alpha/2}].$$

In Algorithm 1, the final confidence interval is directly built by using $\bar{m}$ and $\frac{\hat{\sigma}}{\sqrt{n}}$ (not by $\hat{\sigma}$), and therefore it is possible that the convergence conditions for the coverage probility of constucted confidence interval can be relaxed to $B,n \to \infty$. In the simulations, we also study the cases in which the conditions are relaxed to $B,n \to \infty$ for both nested and non-nested methods, and compare their finite-sample performances extensively, while $k \to \infty$ is still necessary for nested simulation.

4 NUMERICAL RESULTS

This section illustrates the performance of the proposed algorithm using numerical experiments. The following numerical results show the coverage probabilities of the estimated confidence intervals based on the proposed estimators. All of the reported coverage probabilities are validated based on 5000 independent replications.

The Black-Scholes model is a famous option pricing model for the dynamics of a risk-neutral financial market containing derivative investment instruments. In the simulation problem, the objective is to calculate the price of a European call option. For a strike of K and maturity T, a standard result is given by

$$C(S(\tau),K,T) = e^{-rT}\mathbb{E}[\max(S(T)-K,0)],$$

and the statistical process for the stock price grows like

$$\begin{aligned} S(T) &= S(\tau)\cdot \exp((r-\sigma^2/2)(T-\tau)+\sigma\sqrt{T-\tau}W) \\ &= S(\tau)\cdot X(T-\tau), \end{aligned}$$

where $S(\tau)$ represents the stock price at time τ, following a geometric Brownian motion, and W is a standard Wiener process. In this setting, we let $S(0) = 100, r = 0.15, T = 2, \tau = 1, K = 95$, and $\sigma = 0.15$. Assume that the stock price at time T can be represented as the product of two independent random variables $S(\tau)$ and $X(T-\tau)$, where the distribution of $S(\tau)$ is known, and the distribution of $X(T-\tau)$ is unknown with a list of real data.

Substituting in the proposed algorithm, we can obtain the estimated confidence intervals for $S(T)$, and calculate the coverage probabilities by computing the average rate of whether the estimated confidence intervals contain the true value. We also provide the coverage probabilities of the confidence intervals built by the nested simulation when the inner size $k = 2$ as a comparison, as the total computational budget for the non-nested and nested simulation are the same. To test the applicability of our non-parametric approach, we investigate the coverage probabilities of the non-nested bootstrap (NNB) confidence intervals, constructed by the proposed mean and variance estimators. The numerical results show the coverage rate of the estimated 90% and 95% confidence intervals. Moreover, we also provide comparative experiments with nested estimators. Under different n and B, the performances of the algorithms are presented in Tables 1–3.

Table 1: 90% Confidence Intervals for Option Pricing (k=2).

B \ n	50		100		500		1000		5000	
	NNB	Nested k=2	NNB	Nested k=2	NNB	Nested k=2	NNB	Nested k=2	NNB	Nested k=2
100	0.8642	0.9128	0.8828	0.9272	0.8942	0.9370	0.8934	0.9336	0.8920	0.9380
$n^{0.8}$	0.8462	0.9066	0.8664	0.9230	0.8854	0.9326	0.8928	0.9354	0.8968	0.9338
$n/2$	0.8370	0.9018	0.8762	0.9206	0.8912	0.9308	0.8930	0.9284	0.8962	0.9358
n	0.8630	0.9158	0.8784	0.9258	0.8962	0.9338	0.8934	0.9396	0.9014	0.9438
$n^{1.2}$	0.8804	0.9216	0.8960	0.9358	0.8958	0.9304	0.8958	0.9356	0.9000	0.9378

Table 2: 95% Confidence Intervals for Option Pricing (k=2).

B \ n	50		100		500		1000		5000	
	NNB	Nested k=2	NNB	Nested k=2	NNB	Nested k=2	NNB	Nested k=2	NNB	Nested k=2
100	0.9172	0.9582	0.9348	0.9670	0.9444	0.9692	0.9432	0.9716	0.9468	0.9734
$n^{0.8}$	0.9010	0.9526	0.9250	0.9610	0.9462	0.9734	0.9458	0.9702	0.9440	0.9706
$n/2$	0.8962	0.9487	0.9264	0.9612	0.9436	0.9700	0.9406	0.9686	0.9470	0.9744
n	0.9172	0.9582	0.9334	0.9646	0.9464	0.9730	0.9488	0.9738	0.9528	0.9764
$n^{1.2}$	0.9274	0.9610	0.9426	0.9676	0.9440	0.9714	0.9474	0.9740	0.9472	0.9716

It is evident that the coverage probabilities of the proposed NNB method show convergence to the confidence levels, which demonstrates the algorithm can be applied effectively in quantifying input uncertainty in a non-parametric manner. Notably, the coverage probabilities do not converge well to the confidence levels for the nested confidence intervals, which is consistent with the result that the nested variance estimator introduces extra simulation error when quantifying input uncertainty, especially for a small k. Although it seems that the nested method performs better for the cases of $n = 50$ in Table 1 and Table 2, the results should come from the fluctuations caused by small samples. In fact, when $k = 2$, the limitations of the coverage probabilities will not be equal to the confidence levels for the nested method.

Moreover, we also provide the numerical comparison to the nested method with $k \to \infty$ under the same budget. In the following table, let $k = B$ for nested simulation, and B, k will also go to infinity as n goes to infinity. The numerical results show the coverage rate of the estimated 90% confidence intervals.

Table 3: 90% Confidence Intervals for Option Pricing (k=B).

Bk \ n	50		100		500		1000		5000	
	NNB	Nested k=B	NNB	Nested k=B	NNB	Nested k=B	NNB	Nested k=B	NNB	Nested k=B
$2*100$	0.8642	0.8454	0.8828	0.8524	0.8942	0.8542	0.8934	0.8624	0.8920	0.8542
$2*n^{0.8}$	0.8462	0.8004	0.8664	0.8270	0.8854	0.8724	0.8928	0.8718	0.8968	0.8788
$2*n/2$	0.8370	0.8032	0.8762	0.8366	0.8912	0.8788	0.8930	0.8806	0.8962	0.8862
$2*n$	0.8630	0.8368	0.8784	0.8512	0.8962	0.8808	0.8934	0.8846	0.9014	0.8826
$2*n^{1.2}$	0.8804	0.8458	0.8960	0.8644	0.8958	0.8908	0.8958	0.8894	0.9000	0.8918

In this case, the coverage probabilities show convergence to the given confidence level for both nested and non-nested simulation methods, as k increases with n and B. However, the coverage probability of the proposed NNB method converges much faster than that of the nested simulation, which effectively implies that, under the same computational budget, the proposed non-nested variance estimator reduces the variability of the variance estimate and thus provides more accurate confidence intervals compared to the existing nested method.

5 CONCLUSIONS

This paper mainly considers the input uncertainty quantification problem in a non-parametric way, which enables the input uncertainty and simulation uncertainty to occur simultaneously. Since the existing nested bootstrap methods suffer from high computational budget under the joint uncertainty case, we propose a new non-nested bootstrap method to build asymptotically valid confidence intervals for input uncertainty quantification, inspired by the variance estimators in Goda (2017). The theoretical properties are also studied, which establish statistical guarantees for the proposed estimators. An easy-implemented algorithm for building confidence intervals is also provided. Numerical results demonstrate that the estimated confidence intervals perform satisfactorily under given confidence levels, and show faster convergence rate under the same computational budget compared to nested methods, which further verify the effectiveness of the proposed estimators and the corresponding non-nested method.

6 ACKNOWLEDGMENTS

This paper was supported by NSFC/RGC Joint Research Scheme under project N_CityU105/21, National Natural Science Foundation of China (61972261), Natural Science Foundation of Guangdong Province (2023A1515011667), Key Basic Research Foundation of Shenzhen (JCYJ20220818100205012), Shenzhen Science and Technology Program (RCBS20221008093331068), and Basic Research Foundation of Shenzhen (JCYJ20210324093609026).

REFERENCES

Ankenman, B. E., and B. L. Nelson. 2012. "A Quick Assessment of Input Uncertainty". In *Proceedings of the 2012 Winter Simulation Conference*, edited by C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A. Uhrmacher, 1–10. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Barton, R. R., B. L. Nelson, and W. Xie. 2010. "A Framework for Input Uncertainty Analysis". In *Proceedings of the 2010 Winter Simulation Conference*, edited by B. Johansson, S. Jain, J. Montoya–Torres, J. Hugan, and E. Yucesan, 1189–1198. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Barton, R. R., and L. W. Schruben. 1993. "Uniform and Bootstrap Resampling of Empirical Distributions". In *Proceedings of the 1993 Winter Simulation Conference*, edited by G. W. Evans, M. Mollaghasemi, W. E. Biles, and E. C. Russell, 503–508. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Barton, R. R., and L. W. Schruben. 2001. "Resampling Methods for Input Modeling". In *Proceeding of the 2001 Winter Simulation Conference*, edited by B. A. Peters, J. S. Smith, D. J. Medeiros and M. W. Rohrer, 372–378. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Bickel, P. J., and D. A. Freedman. 1981. "Some Asymptotic Theory for the Bootstrap". *The Annals of Statistics* 9(6):1196–1217.

Biller, B., and C. G. Corlu. 2011. "Accounting for Parameter Uncertainty in Large-scale Stochastic Simulations with Correlated Inputs". *Operations Research* 59(3):661–673.

Biller, B., and C. Gunes. 2010. "Capturing Parameter Uncertainty in Simulations with Correlated Inputs". In *Proceedings of the 2010 Winter Simulation Conference*, edited by B. Johansson, S. Jain, J. Montoya–Torres, J. Hugan, and E. Yucesan, 1167–1177. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Cheng, R. 1994. "Selecting Input Models". In *Proceedings of the 1994 Winter Simulation Conference*, edited by J. D. Tew, M. S. Manivannan, D. A. Sadowski and A. F. Seila, 184–191. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Cheng, R. 2001. "Analysis of Simulation Experiments by Bootstrap Resampling". In *Proceedings of the 2001 Winter Simulation Conference*, edited by B. A. Peters, J. S. Smith, D. J. Medeiros, and M. W. Rohrer, 179–186. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Cheng, R. C., and W. Holland. 2004. "Calculation of Confidence Intervals for Simulation Output". *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 14(4):344–362.

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

Chick, S. E. 1997. "Bayesian Analysis for Simulation Input and Output". In *Proceedings of the 1997 Winter Simulation Conference*, edited by S. Andradottir, K. J. Healy, D. H. Withers, and B. L. Nelson, 253–260. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Efron, B. 1979. "Bootstrap Methods: Another Look at the Jackknife". *The Annals of Statistics* 7(1):1–26.

Efron, B. 1981. "Nonparametric Standard Errors and Confidence Intervals". *Canadian Journal of Statistics* 9(2):139–158.

Efron, B., and R. Tibshirani. 1986. "Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy". *Statistical Science* 1(1):54–75.

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

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

Henderson, S. G. 2003. "Input Model Uncertainty: Why Do We Care and What Should We Do About It?". In *Proceedings of the 2003 Winter Simulation Conference*, edited by S. Chick, P. J. Sdnchez, D. Ferrin, and D. J. Morrice, 90–100. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Lam, H., and H. Qian. 2016. "The Empirical Likelihood Approach to Simulation Input Uncertainty". In *Proceedings of the 2016 Winter Simulation Conference*, edited by T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S. E. Chick, 791–802. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Lam, H., and H. Qian. 2019. "Random Perturbation and Bagging to Quantify Input Uncertainty". In *Proceedings of the 2019 Winter Simulation Conference*, edited by N. Mustafee, K.–H. G. Bae, S. Lazarova–Molnar, M. Rabe, and C. Szabo, 320–331. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Lam, H., and H. Qian. 2022. "Subsampling to Enhance Efficiency in Input Uncertainty Quantification". *Operations Research* 70(3):1891–1913.

Singh, K. 1981. "On the Asymptotic Accuracy of Efron's Bootstrap". *The Annals of Statistics* 9(6):1187–1195.

Song, E., and B. L. Nelson. 2015. "Quickly Assessing Contributions to Input Uncertainty". *IIE Transactions* 47(9):893–909.

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.

Yi, Y., and W. Xie. 2017. "An Efficient Budget Allocation Approach for Quantifying the Impact of Input Uncertainty in Stochastic Simulation". *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 27(4):1–23.

Zouaoui, F., and J. R. Wilson. 2003. "Accounting for Parameter Uncertainty in Simulation Input Modeling". *IIE Transactions* 35(9):781–792.

Zouaoui, F., and J. R. Wilson. 2004. "Accounting for Input-model and Input-parameter Uncertainties in Simulation". *IIE Transactions* 36(11):1135–1151.

AUTHOR BIOGRAPHIES

MANJING ZHANG is an associate research fellow in Guangdong Laboratory of Artificial Intelligence and Digital Economy (SZ), Shenzhen, China. She holds a PhD in management science from City University of Hong Kong. Her research interests mainly focus on stochastic simulation, data mining and machine learning. Her email address is zhangmanjing@gml.ac.cn.

GUANGWU LIU is a Professor in the Department of Management Sciences, College of Business at City University of Hong Kong, Hong Kong, China. His research interests include stochastic simulation, financial engineering, risk management and machine learning. He currently serves as an associate editor for Naval Research Logistics and Operations Research. His email address is msgw.liu@cityu.edu.hk.

SHAN DAI is a research scientist in Shenzhen Research Institute of Big Data, Shenzhen, China. He holds a PhD in statistics from The Chinese University of Hong Kong. His research interests mainly focus on time series, nonparametric statistics and machine learning. His email address is shandai@sribd.cn.

YULIN HE is an associate research fellow in Guangdong Laboratory of Artificial Intelligence and Digital Economy (SZ), Shenzhen, China. He holds a PhD from Hebei University. From 2013 to 2014, he has served as a Research Assistant with the Department of Computing, Hong Kong Polytechnic University. From 2014 to 2017, he worked as a Post-doctoral Fellow in the College of Computer Science and Software Engineering in Shenzhen University. His main research interests include big data approximate computing, multi-sample statistic and analysis, and algorithms and applications of data mining and machine learning. Dr. He is a CCF member, CAAI member, CIE member, ACM member, IEEE member, and Editorial Review Board member of several international journals. His email address is heyulin@gml.ac.cn.