

EFFICIENT INPUT UNCERTAINTY QUANTIFICATION VIA GREEN SIMULATION USING SAMPLE PATH LIKELIHOOD RATIOS

Ben M. Feng

Department of Statistics and Actuarial Science
University of Waterloo
Waterloo, Ontario, CANADA

Eunhye Song

Department of IME
The Pennsylvania State University
310 Leonhard Building
University Park, PA 16802, USA

ABSTRACT

Bootstrapping is a popular tool for quantifying input uncertainty, inflated uncertainty in the simulation output caused by finite-sample estimation error in the input models. Typical bootstrap-based procedures have a nested simulation structure that requires $B \times R$ simulation runs: the outer loop bootstraps B input distributions, each of which requires R inner simulation runs. In this article, we present a measure-theoretic framework for constructing a sample path likelihood ratio and propose an efficient input uncertainty quantification procedure using two green simulation estimators. The proposed procedures reuse the same R inner simulation outputs in all outer loops by reweighting them using appropriately defined likelihood ratios. Both procedures produce asymptotically valid confidence intervals for the expected simulation output under the true input distribution. Our numerical results show that the proposed procedures have efficiency gains compared to other popular bootstrap-based alternatives.

1 INTRODUCTION

Probability distributions that drive the randomness in a stochastic simulation are often referred to as input models. These input models are fitted from finite observations of true but unknown real-world processes, and therefore are subject to estimation errors. When the fitted input models are used to run simulations, their estimation errors are propagated to the simulation output and cause input uncertainty, which inflates the simulation output variance beyond the usual Monte Carlo simulation noise. In general, input uncertainty can be only reduced by collecting additional real-world data and improving the input models. Standard simulation output analysis is often based solely on Monte Carlo noise, which is insufficient in such settings. For example, in a queueing simulation illustrated by (Barton and Schruben 2001), when input uncertainty is overlooked, the 90% confidence interval for the expected waiting time only has 20% coverage probability. So quantifying the input uncertainty is an important task as it provides simulation users with a better understanding of the real-world system being modeled from simulation outputs.

Input uncertainty quantification (IUQ) has been an active research area, as reviewed by Barton (2012), Song et al. (2014), and Lam (2016). With few recent exceptions (Zhou and Liu 2018), the standard setting in the IUQ literature concerns cases where additional data collection is expensive or even infeasible. Adopting this problem setting, we focus on deriving asymptotically valid confidence intervals (CIs) for the expected simulation output under the unknown true input distributions with a target coverage probability guarantee as the real-world sample size increases. This problem has been studied using bootstrap-based methods (Barton and Schruben 1993; Barton and Schruben 2001; Ankenman and Nelson 2012), the delta method (Cheng and Holland 1997; Ng and Chick 2006; Song and Nelson 2015; Lin et al. 2015), and metamodeling-based methods (Xie et al. 2014).

One of the main challenges of the IUQ problem lie in its nested simulation structure, which results in excessive computational burden. Ideally, if one has abundant (outer) samples of fitted input models and the conditional performance measure given the fitted input models can be exactly evaluated without (inner) simulation, then the CI of the true mean performance measure can be estimated using empirical quantiles of the conditional performance measures. However, in actuality, the sampling distribution of fitted input models is unknown and the conditional performance measures must be estimated via simulations. The former is often approximated by bootstrapping the data when nonparametric input models are used, or by an asymptotic sampling distribution when parametric distribution families are assumed for the input models; both approximations are theoretically justified, asymptotically, as the real-world sample size increases to infinity. Estimating the conditional mean is harder to get around. For instance, the direct bootstrap approach (Barton and Schruben 2001) bootstraps B times from the real-world sample and runs R simulation replications at each of the B fitted input model, resulting in a total of BR simulation runs. Here, both B and R need to be reasonably large to reduce the quantile estimation error and the Monte Carlo error in estimating the conditional performance measure given fitted input models. The delta effects model or metamodeling approach alleviate the computational burden by reducing the inner simulations required but at the expense of imposing structural models of the conditional mean function. However, the parameters of the imposed structural models need to be estimated too, which incurs estimation errors and requires large simulation efforts if there are several input models (i.e., high-dimensional parameter vector).

There are recent studies that aim to tackle the aforementioned computational challenge. Barton et al. (2018) apply a shrinkage scheme to the direct bootstrap approach so that the conditional mean estimates are first adjusted for its Monte Carlo error to mitigate the inflation in the CI even for smaller R . Glynn and Lam (2018) apply a sectioning method to the real-world data sample and construct the CI using a pivotal statistic. Lam and Qian (2018) investigate a subsampling scheme in conjunction with bootstrap. These studies achieve efficiency gain by cleverly extracting information from the conditional performance measures of the replications, but pay little attention to the sample paths of inputs generated within each outer replication. In contrast, Lin et al. (2015) use within-replication sample paths to compute control variates and estimate the gradients for the delta method without any additional simulation runs. Our proposed research is inspired by this last approach in that we extract more information from the within-replication sample paths, where the efficiency gain is achieved by applying green simulation at the sample-path level.

Green simulation was introduced by Feng and Staum (2015) and further expanded in Feng and Staum (2017) as a new experiment design paradigm that reuses simulation outputs from past experiments to improve the efficiency of future experiments. One implementation of green simulation employs the so-called likelihood ratio (LR) method, also known as the score function method. The LR method has been applied in metamodeling, sensitivity analysis, and optimization; see, for example, Beckman and McKay (1987), L'Ecuyer (1990), L'Ecuyer (1993), Rubinstein and Shapiro (1993), Kleijnen and Rubinstein (1996), Glasserman and Xu (2014), Dong et al. (2018) and Eckman and Feng (2018). The LR method is mathematically identical to the well-known importance sampling method but they have different purposes: the LR method aims to reuse simulation outputs while the importance sampling method aims to reduce estimation variance. Extensions of importance sampling, i.e., multiple importance sampling (Hesterberg 1988; Veach 1997), defensive importance sampling (Owen and Zhou 2000), and adaptive importance sampling (Cornuet et al. 2012; Martino et al. 2015), are particularly relevant to the LR method and its application in green simulation.

Zhou and Liu (2018) also apply green simulation for IUQ, however, their approach differs from ours in two ways; 1) we consider the case with fixed data set, whereas they concern streaming data; 2) we apply green simulation at the sample-path level, whereas, they apply at the replication level by defining the likelihood ratios from the posterior distribution of input parameters. Specifically, via the likelihood ratio method at the sample path level, we reuse the same set of R inner simulation sample paths to estimate all B outer conditional performance measures and to construct an asymptotically valid CI for the true performance measure. The main merit of this approach is computational efficiency: only R simulation runs

are required, rather than BR runs in the direct bootstrap method, to estimate the conditional means at all B bootstrapped input models. This approach also has its limitations: It is only applicable to simulations whose input models affect only the distributions of random variates in the simulation model. It requires the user to be not only able to sample from the random variates but also able to calculate their likelihoods. Lastly, in reusing the inner sample paths it requires additional likelihood ratio calculations, which could be itself a substantial computational burden. In this study, we consider two green LR estimators to illustrate the main ideas, implementations, and potentials of green simulation for solving IUQ problems.

The remainder of the paper is organized as follows. Section 2 illustrates the nested simulation structure of input uncertainty quantification of a stochastic simulation model and highlight computational challenges. Section 3 defines the probability space of the sample paths generated within replications, which enables us to mathematically analyze two green simulation estimators of the conditional output mean given bootstrapped input distributions proposed in Section 4. Based on these estimators, we present an IUQ procedure that requires only R replications and show it provides an asymptotically correct CI for the output mean under the true real-world distributions. Section 5 compares empirical performance of the proposed method with a bootstrap percentile CI method and demonstrates its efficiency gain. Due to the space limit, we state our theoretical results without proofs, which will be presented in follow-up work.

2 INPUT UNCERTAINTY QUANTIFICATION VIA NESTED BOOTSTRAP SIMULATIONS

Consider a simulation experiment that is driven by L independent parametric input distributions h_1, h_2, \dots, h_L whose families are known, but the correct parameters $\boldsymbol{\theta}^c := \{\boldsymbol{\theta}_1^c, \boldsymbol{\theta}_2^c, \dots, \boldsymbol{\theta}_L^c\}$ are unknown. We consider situations where the unknown input parameters $\boldsymbol{\theta}^c$ can be estimated by $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\mathbf{X})$ based on the observations, denoted by $\mathbf{X} = \{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_L\}$, of some real-world processes. Let $m_\ell := |\mathbf{X}_\ell|$ be the sample size of the ℓ -th real-world process. For notational convenience, the parameter space, i.e., the set of all possible values of $\hat{\boldsymbol{\theta}}$, is denoted by $\Theta \subseteq \mathbb{R}^p$ for some $p \geq L$. As a generic example, $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2) > 0$ may be the maximum likelihood estimates of the interarrival rate and the service rate based on the observed interarrival times \mathbf{X}_1 and serviced times \mathbf{X}_2 in a real-world system. Also, $h_1(\cdot; \hat{\theta}_1)$ and $h_2(\cdot; \hat{\theta}_2)$ are the density functions for the interarrival time and service time distribution in a simulation model of the system.

Given an input parameter $\boldsymbol{\theta} \in \Theta$, the i -th replication of the stochastic simulation model generates a sample path $\mathbf{Z}_i = (Z_{i,\ell_1}, \dots, Z_{i,\ell_{N_i}})$ where $Z_{i,\ell_j} \sim h_{\ell_j}(\cdot, \boldsymbol{\theta}_{\ell_j})$ and h_{ℓ_j} is the distribution of the j th input generated within the sample path. Then the simulation model calculates a performance measure, or the simulation output, $Y_i = g(\mathbf{Z}_i)$ for some deterministic function g . For instance, \mathbf{Z}_i may be the simulated interarrival and service times of a queueing system and $g(\mathbf{Z}_i)$ is the average time-in-system. Note that the total number of realized inputs, N_i , is also random.

The simulation outputs obtained from R independent and identically distributed (i.i.d.) replications can be represented as

$$Y_i(\mathbf{Z}_i; \boldsymbol{\theta}) = \mu(\boldsymbol{\theta}) + \varepsilon_i(\mathbf{Z}_i; \boldsymbol{\theta}), \quad \forall i = 1, \dots, R,$$

where $\varepsilon(\mathbf{Z}_i; \boldsymbol{\theta})$ represents the simulation error such that $E[\varepsilon(\mathbf{Z}_i; \boldsymbol{\theta})|\boldsymbol{\theta}] = 0$ and $\text{Var}[\varepsilon(\mathbf{Z}_i; \boldsymbol{\theta})|\boldsymbol{\theta}] < \infty$ for all $\boldsymbol{\theta} \in \Theta$, and $\mu(\boldsymbol{\theta}) = E[g(\mathbf{Z})|\boldsymbol{\theta}] = \int_{\mathcal{Z}} g(\mathbf{z}) d\nu_{\boldsymbol{\theta}}(\mathbf{z})$ is conditional expected performance given input parameter $\boldsymbol{\theta}$. Note that \mathcal{Z} and $\nu_{\boldsymbol{\theta}}$ are the sample space and the probability measure for the sample paths, respectively. The standard Monte Carlo (SMC) estimator for $\mu(\boldsymbol{\theta})$ is given by

$$\hat{\mu}_R^{\text{SMC}}(\boldsymbol{\theta}) = \frac{1}{R} \sum_{i=1}^R Y_i(\mathbf{Z}_i; \boldsymbol{\theta}) = \frac{1}{R} \sum_{i=1}^R [\mu(\boldsymbol{\theta}) + \varepsilon_i(\mathbf{Z}_i; \boldsymbol{\theta})]. \quad (1)$$

When $\hat{\boldsymbol{\theta}}$ is used to run simulations, the variance of (1) can be decomposed into two parts:

$$\text{Var} \left[\hat{\mu}_R^{\text{SMC}}(\hat{\boldsymbol{\theta}}) \right] = \text{Var}_{\hat{\boldsymbol{\theta}}} \left[\mu(\hat{\boldsymbol{\theta}}) \right] + \frac{1}{R} E_{\hat{\boldsymbol{\theta}}} \left[\text{Var}_{\mathbf{Z}} \left[\varepsilon(\mathbf{Z}_i; \hat{\boldsymbol{\theta}}) | \hat{\boldsymbol{\theta}} \right] \right].$$

Input uncertainty in Y is characterized by $\text{Var}_{\hat{\theta}}[\mu(\hat{\theta})]$, the uncertainty in simulation output due to estimation error in $\hat{\theta}$, whereas Monte Carlo error is characterized by $E_{\hat{\theta}}[\text{Var}_{\mathbf{Z}}[\varepsilon(\mathbf{Z}; \hat{\theta})|\hat{\theta}]]$. To construct a CI for $\mu(\theta^c)$, both uncertainties must be accounted for. Quantifying input uncertainty requires the sampling distribution of $\hat{\theta}$, which is generally unobtainable from a single set of observations \mathbf{X} . Instead, one common approach is to bootstrap \mathbf{X} (nonparametric bootstrap) or sample from the estimated distribution parameterized with $\hat{\theta}$ (parametric bootstrap) and re-estimate θ^c .

Algorithm 1: Parametric Bootstrap Percentile CI

Input: Real-world observations $\mathbf{X} = \{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_L\}$. Two-sided confidence level, α . Number of bootstrap samples, B . Number of simulated sample paths per bootstrapped distribution, R . Parametric input distribution families h_1, h_2, \dots, h_L .

Output: Two-sided CI for $\mu(\theta^c)$

- 1 **Initialization:** Estimate input parameters $\hat{\theta} = \{\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_L\}$ from real-world observations \mathbf{X} ;
 - 2 **for** $b = 1, 2, \dots, B$ **do**
 - 3 Generate synthetic size- m_ℓ observations \mathbf{X}_ℓ^* from $h_\ell(\cdot; \hat{\theta}_\ell)$ for all $\ell = 1, \dots, L$ then estimate bootstrapped input parameters θ_b^* from \mathbf{X}^* in the same way as in the Initialization;
 - 4 **for** $r = 1, 2, \dots, R$ **do**
 - 5 Based on bootstrapped input parameters $\theta_b^* = \{\theta_{b1}^*, \dots, \theta_{bL}^*\}$, obtain the r -th simulated performance measure $Y_{br} = g(\mathbf{Z}_{br})$;
 - 6 Compute the b -th performance-measure estimate by $\hat{\mu}_R^{\text{SMC}}(\theta_b^*)$;
 - 7 **Output** $(\hat{q}_{\alpha/2}, \hat{q}_{1-\alpha/2})$ as the $1 - \alpha$ CI for $\mu(\theta^c)$, where $\hat{q}_{\alpha/2}$ and $\hat{q}_{1-\alpha/2}$ are the $\alpha/2$ and $1 - \alpha/2$ sample quantiles of $\hat{\mu}_R^{\text{SMC}}(\theta_1^*), \hat{\mu}_R^{\text{SMC}}(\theta_2^*), \dots, \hat{\mu}_R^{\text{SMC}}(\theta_B^*)$;
-

Algorithm 1 is a benchmark procedure to be compared with our proposals; it is a parametric version of the bootstrap percentile CI approach in Barton et al. (2018). Algorithm 1 has a nested simulation structure: each of the B outer loop (Lines 2 to 6) simulates one bootstrapped input parameter θ_b^* , which requires R inner loops (Lines 4 to 5) to simulate performance measures Y_{br} 's. In each outer loop, we essentially approximate real-world distributions $h_\ell(\cdot; \theta_\ell^c)$'s with $h_\ell(\cdot; \hat{\theta}_\ell)$'s and the sampling distribution of $\hat{\theta}$ with that of θ_b^* . Because the Monte Carlo error of $\hat{\mu}_R^{\text{SMC}}(\theta_b^*)$ is convoluted with input uncertainty so the CI produced by Algorithm 1 tends to show overcoverage. Therefore, large R is required to obtain a tight CI in Algorithm 1 (Barton and Schruben 2001).

Due to the nested simulation structure, Algorithm 1 requires a total BR inner simulation runs. We propose two green simulation estimators in Section 4 which produces asymptotically CIs for $\mu(\theta^c)$ with much smaller computational cost, i.e., only R simulation runs are required. Both our proposals and that of Zhou and Liu (2018) use green simulation estimators to reuse simulation outputs via likelihood ratios, but at different granularity. Zhou and Liu (2018) reuse simulation outputs at the outer level, $\hat{\mu}_R^{\text{SMC}}(\theta_b^*)$'s, by weighting them with likelihood ratios constructed from the posterior distribution of $\hat{\theta}$. We reuse simulation outputs at the inner level, $Y = g(\mathbf{Z})$, by weighting them with a sample-path likelihood ratio.

3 LIKELIHOOD RATIOS FOR SAMPLE PATHS OF DISCRETE EVENT SIMULATIONS

For some simulation models, each replication requires the same number of inputs from each of L input distributions. In this case, the sample path likelihood function is simply a product of individual inputs' probability functions. In many stochastic simulation models, however, the length of sample paths and the order of inputs generated within the sample paths in different replications vary. To illustrate this point, Figure 1 depicts a tree representation of some possible sample paths for the discrete event simulation of

an $M/M/1/1$ queue that starts from an empty system and terminates after two departures. Between the root node (\square) and the leaf node (\diamond), each node represents either the simulated interarrival or service time (denoted by A_i and S_i , respectively). Because the simulation starts from an empty system, the first arrival triggers an immediate service and the scheduling of the next arrival. Consequently, A_1, S_1 , and A_2 , are generated in sequence for all sample paths. From then on, the sequence of inputs may differ depending on the realized values of the preceding inputs. For instance, if the first service ends before the second arrival, i.e., $A_2 \geq S_1$, then the second service time, S_2 , is generated after A_2 . Otherwise, the second arrival is rejected from joining the system as the first customer has not yet left the system and the next interarrival time, A_3 , is generated. In this example, the sample space of sample paths generated within a replication is not a typical subset of \mathbb{R}^n because the sample paths have different lengths.

It turns out that even if $\mathcal{Z} \not\subset \mathbb{R}^n$ as in the above example, the sample path likelihood function is still equivalent to the product of individual inputs' probability distribution functions that are realized within the sample path. However, defining the likelihood function rigorously requires more work. In the following, we present a measure theoretic framework that help defining the sample path likelihood ratios.

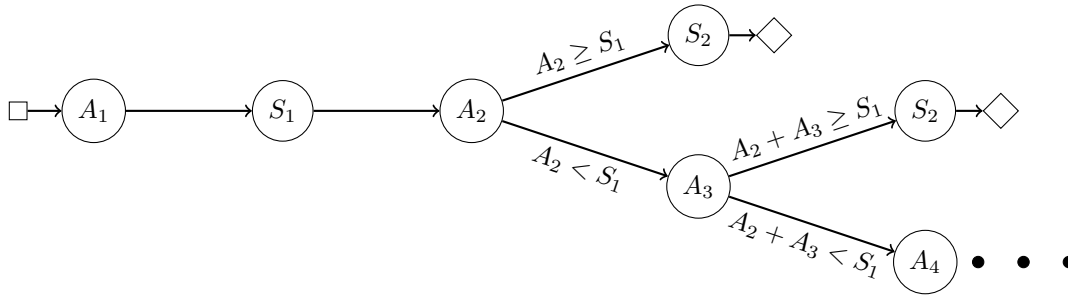


Figure 1: Tree representation of sample paths for a $M/M/1/1$ discrete event simulation model. Each sample path starts from the root node (\square) to a leaf node (\diamond).

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the underlying probability space of the inputs of a discrete event simulation model. Given any input parameter θ , simulation inputs are generated from $h_\ell(\cdot; \theta_\ell)$, $\ell = 1, \dots, L$ according to the simulation logic until a termination condition is met. We denote the sample space of the sample path \mathbf{Z} and its σ -algebra by \mathcal{Z} and \mathcal{G} , respectively, and define mapping $\zeta_\theta : (\Omega, \mathcal{F}) \mapsto (\mathcal{Z}, \mathcal{G})$ so that $\mathbf{Z} = \zeta_\theta(\omega)$ is completely determined by ω . We will show that ζ_θ^{-1} is well-defined and $\nu_\theta := \mathbb{P} \circ \zeta_\theta^{-1}$ is a valid probability measure on $(\mathcal{Z}, \mathcal{G})$.

Assumption 1 For all $\theta \in \Theta$, the L input distributions have the same support.

Assumption 1 ensures that the sample space \mathcal{Z} is common for any input distributions parameterized by $\theta \in \Theta$. It is a sufficient condition to ensure the absolute continuity between some measures so that likelihood ratio functions of interest are well-defined.

Let $\pi(n)$ be the number of depth- n paths in the tree representation depicted in Figure 1; by depth we the number of nodes in a path, excluding the root (\square) and leaf (\diamond) nodes. Clearly, $\pi(n) \leq L^n$, where the equality holds if every node in the tree has L child nodes, but typically the strict inequality holds. For instance, $\pi(4) = 1$ in Figure 1 because there is only one path, i.e., $A_1 \rightarrow S_1 \rightarrow A_2 \rightarrow S_2$, that ends after generating four inputs. If $\pi(n) > 1$, then each of $\pi(n)$ paths has a distinct sequence of inputs. Let $\mathcal{Z}_{n,j}$ be the event that exactly n inputs are generated in the j -th sequence for $j = 1, 2, \dots, \pi(n)$. Then, $\mathcal{Z}_{n,j}, \forall (n, j)$, are mutually exclusive and the sample space can be partitioned as $\mathcal{Z} = \cup_{n \in \mathbb{N}} \cup_{j=1}^{\pi(n)} \mathcal{Z}_{n,j}$. Lastly, let $\eta_{\theta,n,j}$ be the mapping from Ω to \mathbb{R}^n that generates i.i.d. random variables from the same sequence of input distributions as $\mathcal{Z}_{n,j}$.

The following lemma states that for each subset E of $\mathcal{Z}_{n,j}$, the preimage of E under mappings ζ_{θ} and $\eta_{\theta,n,j}$ are the same. Since $\eta_{\theta,n,j}$ is a well-defined measure, this implies that $\zeta_{\theta}^{-1}(E)$ is measurable by \mathbb{P} .

Lemma 1 For $E \in \mathcal{Z}_{n,j}$, $\zeta_{\theta}^{-1}(E) = \eta_{\theta,n,j}^{-1}(E)$.

From Lemma 1, Theorem 1 establishes that ν_{θ} is a probability measure on $(\mathcal{Z}, \mathcal{G})$ by showing that the preimage of any $A \in \mathcal{G}$ under ζ_{θ} is measurable by \mathbb{P} .

Theorem 1 For each $\theta \in \Theta$, ζ_{θ} is a measurable mapping and $\nu_{\theta} := \mathbb{P} \circ \zeta_{\theta}^{-1}$ is a probability measure on $(\mathcal{Z}, \mathcal{G})$.

We proceed to defining a likelihood ratio between ν_{θ_1} and ν_{θ_2} for $\theta_1, \theta_2 \in \Theta$ on \mathcal{Z} . Let $f_{\theta}(\cdot)$ be the product of probability density functions of random variables generated within a sample path, i.e., $f_{\theta}(\mathbf{Z}) = \prod_{j=1}^{|\mathbf{Z}|} h_{\ell_j}(\mathbf{Z}_j; \theta_{\ell_j})$. If $\mathcal{Z} \subset \mathbb{R}^n$, the likelihood ratio is simply defined as $f_{\theta_1}(\mathbf{Z})/f_{\theta_2}(\mathbf{Z})$. However, since \mathcal{Z} is not a subset of \mathbb{R}^n , we define the likelihood ratio in a more general way as the Radon-Nikodym derivative of ν_{θ_1} with respect to ν_{θ_2} on \mathcal{Z} . Theorem 2 shows that $f_{\theta_1}(\mathbf{Z})/f_{\theta_2}(\mathbf{Z})$ is indeed the Radon-Nikodym derivative of ν_{θ_1} with respect to ν_{θ_2} on \mathcal{Z} .

Theorem 2 Given $\theta_1, \theta_2 \in \Theta$, $f_{\theta_1}/f_{\theta_2}$ is the Radon-Nikodym derivative of ν_{θ_1} with respect to ν_{θ_2} .

Theorem 2 is proved by showing that $\int_A \frac{f_{\theta_1}(\mathbf{z})}{f_{\theta_2}(\mathbf{z})} d\nu_{\theta_2}(\mathbf{z}) = \nu_{\theta_1}(A)$ for any $A \in \mathcal{G}$.

4 GREEN SIMULATION FOR INPUT UNCERTAINTY QUANTIFICATION

Green simulation is introduced by Feng and Staum (2015) and further expanded in Feng and Staum (2017) as a new experiment design paradigm that reuses existing simulation outputs to improve efficiency of future experiments. Specifically, green simulation estimators based on the likelihood ratio method were shown to be an effective way to recycle and reuse simulation outputs. In the context of input uncertainty quantification, we use green simulation estimators to reuse the same R replications to estimate the conditional expected system performances for all bootstrapped input parameters.

Under Assumption 1, and Theorem 2, the conditional expected performance given input parameter θ can be written as

$$\mu(\theta) = E[g(\mathbf{Z})|\theta] = \int_{\mathcal{Z}} g(\mathbf{z}) d\nu_{\theta}(\mathbf{z}) = \int_{\mathcal{Z}} g(\mathbf{z}) \frac{f_{\theta}(\mathbf{z})}{f_{\theta_s}(\mathbf{z})} d\nu_{\theta_s}(\mathbf{z}) = E \left[g(\mathbf{Z}_s) \frac{f_{\theta}(\mathbf{Z}_s)}{f_{\theta_s}(\mathbf{Z}_s)} \middle| \theta, \theta_s \right],$$

for any $\theta_s \in \Theta$; θ_s is the parameter of the distribution from which the inputs are generated. In short, we refer to θ_s as the sampling input parameter. Then, for a target input parameter $\theta \in \Theta$ for which we want to estimate $\mu(\theta)$, the following individual likelihood ratio (ILR) green simulation estimator is an unbiased estimator of $\mu(\theta)$:

$$\hat{\mu}_R^{\text{ILR}}(\theta) = \frac{1}{R} \sum_{s=1}^R g(\mathbf{Z}_s) \frac{f_{\theta}(\mathbf{Z}_s)}{f_{\theta_s}(\mathbf{Z}_s)}, \text{ where } \mathbf{Z}_s \sim h(\cdot; \theta_s), \forall s = 1, \dots, R. \tag{2}$$

Note that the sampling parameters $\theta_1, \theta_2, \dots, \theta_R$ may or may not be the same. The input parameter estimated from the real-world process $\hat{\theta}$ is a natural choice as the sampling input parameter, e.g., $\theta_s = \hat{\theta}$; we consider this case in our numerical experiments. Since the likelihood function $f_{\theta}(\cdot)$ is a product of probability distribution functions, the resulting likelihood ratios and the ILR estimator can have a large variance. To improve numerical stability, the sum of exponents are calculated first in our implementation. In our numerical experiments, the ILR estimator produces wide confidence intervals. This effect may worsen when each individual replication is longer (more inputs per sample path). We assume that each replication generates finite number of inputs until termination, which is a practically reasonable assumption. Another remedy for this issue is to select a subset of realized inputs in the sample path for likelihood ratio calculation, which reduces the variance of the resulting green simulation estimator, but introduces bias. We defer this investigation to the future work.

When θ_s 's are distinct, a provably more efficient estimator studied in the literature is the following mixture likelihood ratio (MLR) estimator:

$$\hat{\mu}_R^{\text{MLR}}(\boldsymbol{\theta}) = \frac{1}{R} \sum_{s=1}^R g(\mathbf{Z}_s) \frac{f_{\boldsymbol{\theta}}(\mathbf{Z}_s)}{\bar{f}_R(\mathbf{Z}_s)}, \text{ where } \bar{f}_R(\mathbf{z}) = \frac{1}{R} \sum_{s=1}^R f_{\boldsymbol{\theta}_s}(\mathbf{z}). \quad (3)$$

The MLR estimator views the collection of sample paths $\{\mathbf{Z}_1, \dots, \mathbf{Z}_R\}$ as a stratified sample from the mixture distribution \bar{f}_R , even though they are indeed generated from $f_{\boldsymbol{\theta}_s}$ for $s = 1, \dots, R$. The MLR estimator is studied in Hesterberg (1988), Veach and Guibas (1995), Feng and Staum (2015), Feng and Staum (2017) and has been shown to have smaller variance compared to the ILR estimator; our numerical experiments have similar findings.

We use the same notation R as the number of independent sample paths in (1), (2), and (3) only for ease comparing their performances with the same number of simulation runs. However, we reiterate that the green simulation estimators (2) and (3) use the same R sample paths to estimate conditional performance measures for all B parameters while the SMC estimator requires BR sample paths to achieve the same task.

Algorithm 2 constructs a CI for $\mu(\boldsymbol{\theta}^c)$ from green simulation estimators. Note $\hat{\mu}_R^{\text{GS}}(\cdot)$ refers to either $\hat{\mu}_R^{\text{MLR}}(\cdot)$ or $\hat{\mu}_R^{\text{ILR}}(\cdot)$. Algorithm 2 has three main loops: bootstrapping input parameters $\boldsymbol{\theta}_b^*$ in Lines 2-3, running the simulator in Lines 4-6, and estimating the conditional expected performances $\mu(\boldsymbol{\theta}_b^*)$'s in Lines 7-8. These steps are performed in sequence rather than in a nested loop, which is the main reason for the efficiency gain achieved by Algorithm 2 over Algorithm 1. In general, the sampling parameter $\boldsymbol{\theta}_s$ may be $\hat{\boldsymbol{\theta}}$ or one of the bootstrapped parameters $\boldsymbol{\theta}_b^*$. However, it is desirable to select $\boldsymbol{\theta}_s$ so that the resulting green simulation estimators have good performance, e.g., have finite variance, for any $\boldsymbol{\theta}_b^*$. When the input distributions are members of the exponential family, we identify a sufficient condition in Proposition 1. In cases where the simulation runs take much longer than likelihood ratio calculations, Algorithm 2 can be B times faster than Algorithm 1.

Algorithm 2: Outline of Green Simulation Percentile CI

Input: Real-world observations $\mathbf{X} = \{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_L\}$. Two-sided confidence level, α . Number of bootstrap samples, B . Number of simulated sample paths, R . Parametric input distribution families h_1, h_2, \dots, h_L .

Output: Two-sided CI for $\mu(\boldsymbol{\theta}^c)$

- 1 **Initialization:** Estimate input parameters $\hat{\boldsymbol{\theta}} = \{\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2, \dots, \hat{\boldsymbol{\theta}}_L\}$ from real-world observations \mathbf{X} ;
 - 2 **for** $b = 1, 2, \dots, B$ **do**
 - 3 Generate synthetic size- m_ℓ observations \mathbf{X}_ℓ^* from $h_\ell(\cdot; \hat{\boldsymbol{\theta}}_\ell)$ for all $\ell = 1, \dots, L$ then estimate bootstrapped input parameters $\boldsymbol{\theta}_b^*$ from \mathbf{X}^* in the same way as in the Initialization;
 - 4 **for** $s = 1, 2, \dots, R$ **do**
 - 5 Select sampling input parameter $\boldsymbol{\theta}_s$ based on some criteria;
 - 6 Based on $\boldsymbol{\theta}_s$, simulate sample path \mathbf{Z}_s and calculate $g(\mathbf{Z}_s)$, then store both quantities;
 - 7 **for** $b = 1, 2, \dots, B$ **do**
 - 8 Compute the b -th performance-measure estimate $\hat{\mu}_R^{\text{GS}}(\boldsymbol{\theta}_b^*)$, where GS may be ILR or MLR;
 - 9 Output $(\hat{q}_{\alpha/2}, \hat{q}_{1-\alpha/2})$ as the $(1 - \alpha)$ -CI for $\mu(\boldsymbol{\theta}^c)$, where $\hat{q}_{\alpha/2}$ and $\hat{q}_{1-\alpha/2}$ are the $\alpha/2$ and $1 - \alpha/2$ sample quantiles of $\hat{\mu}_R^{\text{GS}}(\boldsymbol{\theta}_1^*), \hat{\mu}_R^{\text{GS}}(\boldsymbol{\theta}_2^*), \dots, \hat{\mu}_R^{\text{GS}}(\boldsymbol{\theta}_B^*)$;
-

In the following analysis, we show that the CI produced by Algorithm 2 are asymptotically valid for both ILR and MLR. Moreover, we will articulate a criterion in Line 5 of Algorithm 2 for choosing $\boldsymbol{\theta}_s$. Notice that Algorithm 2 is conditional on real-world observations \mathbf{X} , so is most of the analysis in this section. However, for notational simplicity, we drop “ \mathbf{X} ” from the mathematical expressions unless specified otherwise.

The key to establishing the asymptotic validity of the percentile CIs produced in Algorithm 2 is to analyze the conditional biases and variances for the ILR and MLR estimators. For notational convenience, we define the target- θ_b^* -sample- θ_s variance as

$$\sigma^2(\theta_b^*, \theta_s) = \text{Var} \left[g(\mathbf{Z}_s) \frac{f_{\theta_b^*}(\mathbf{Z}_s)}{f_{\theta_s}(\mathbf{Z}_s)} \middle| \theta_b^*, \theta_s \right].$$

When the target and sampling input parameters coincide, i.e., $\theta_b^* = \theta_s = \theta \in \Theta$, $\sigma_\theta^2 := \sigma^2(\theta, \theta)$ is the Monte Carlo variance $\text{Var}_{\mathbf{Z}}[\varepsilon(\mathbf{Z}; \theta) | \theta]$. In the simulation literature, it is standard to assume that the Monte Carlo variance is finite for any $\theta \in \Theta$. A stronger assumption, i.e., Assumption 2 is standard in analyzing likelihood-ratio-based green simulation estimators such as ILR and MLR.

Assumption 2 For any sampling and target input parameters $\theta_b^*, \theta_s \in \Theta$, the target- θ_b^* -sample- θ_s variance is finite, i.e., $\sigma^2(\theta_b^*, \theta_s) < \infty$ for any $\theta_b^*, \theta_s \in \Theta$.

Conditioning on the sampling input parameters $\theta_1, \dots, \theta_R$, the sample paths \mathbf{Z}_s 's in the ILR estimator are independent. Therefore $\text{Var} \left[\hat{\mu}_R^{\text{ILR}}(\theta_b^*) | \theta_b^*, \theta_1, \dots, \theta_R \right] = \frac{1}{R^2} \sum_{s=1}^R \sigma^2(\theta_b^*, \theta_s)$, so Assumption 2 ensures that the ILR estimator has a finite conditional variance. While Assumption 2 is difficult to verify in general, we present some mild sufficient conditions when f_θ belongs to the exponential family below.

The probability density or mass function of an exponential family distribution in its natural form can be written as $f_\theta(\mathbf{z}) = c(\mathbf{z}) \exp(\theta \cdot T(\mathbf{z}) - A(\theta))$, where $\theta = (\theta_1, \dots, \theta_p)$ are the natural parameters and $c(\mathbf{z})$, $T(\mathbf{z})$, and $A(\theta)$ are known functions. The natural parameter space, denoted by Ξ , is the set of natural parameters such that $f(\mathbf{z} | \theta)$ is a well-defined probability density or mass function on the support \mathcal{Z} , i.e., $\Xi := \{\theta \in \mathbb{R}^p | \int_{\mathcal{Z}} f_\theta(\mathbf{z}) d\mathbf{z} < \infty\}$.

Proposition 1 Consider a simulation model $g(\mathbf{z})$ whose input distribution is given by $f_\theta(\mathbf{z})$ and has finite Monte Carlo variance for any input parameter $\theta \in \Xi$. Suppose $f_\theta(\mathbf{z})$ is a member of the exponential family where θ is its natural parameter. If the input parameter space Θ satisfies $2\theta_b^* - \theta_s \in \Xi$ for any $\theta_b^*, \theta_s \in \Theta$, then $\sigma^2(\theta_b^*, \theta_s) < \infty$ for any $\theta_b^*, \theta_s \in \Theta$.

With Proposition 1, we can verify Assumption 2 by examining the input distribution and the geometry of the input parameter space. For example, consider the pdf of an exponential distribution, $f_\theta(z) = \theta \exp(-\theta z)$, and a simulation model g such that $\text{Var}[g(Z) | \theta] < \infty$ for any $\theta \in \Xi := \mathbb{R}^+$. Then Proposition 1 implies that Assumption 2 holds for any parameter space $\Theta := [\theta_{\min}, \theta_{\max}]$ such that $2\theta_{\min} > \theta_{\max}$.

For completeness we state two known results for ILR and MLR, on which our analysis is based. Lemma 2 dictates that both ILR and MLR are conditionally unbiased estimators. Lemma 3 states that the conditional variance of the MLR estimator, given the sampling and target input parameters, is not greater than that of the ILR estimator. Such inequality provides a shortcut for analyzing the convergence property of the variance of MLR estimator, which is bounded by that of the ILR estimator.

Lemma 2 [Paraphrase of Lemma 9.1 in Veach (1997) and Proposition 2.4 in Feng and Sturm (2017)] Given the set of sampling input parameters $\theta_s \in \Theta$, $s = 1, \dots, R$ and any bootstrapped parameter $\theta_b^* \in \Theta$. If Assumption 1 holds, then both the MLR estimator is unbiased, i.e.,

$$\mathbb{E} \left[\hat{\mu}_R^{\text{ILR}}(\theta_b^*) | \theta_b^*, \theta_1, \dots, \theta_R \right] = \mathbb{E} \left[\hat{\mu}_R^{\text{MLR}}(\theta_b^*) | \theta_b^*, \theta_1, \dots, \theta_R \right] = \mu(\theta_b^*)$$

Lemma 3 [Paraphrase of Theorem A.2 in Martino et al. (2015) and Proposition 2.5 in Feng and Sturm (2017)] Given the set of sampling input parameters $\theta_s \in \Theta$, $s = 1, \dots, R$ and any bootstrapped parameter $\theta_b^* \in \Theta$. If Assumptions 1 and 2 hold, then the following inequality holds,

$$\text{Var} \left[\hat{\mu}_R^{\text{MLR}}(\theta_b^*) | \theta_b^*, \theta_1, \dots, \theta_R \right] \leq \text{Var} \left[\hat{\mu}_R^{\text{ILR}}(\theta_b^*) | \theta_b^*, \theta_1, \dots, \theta_R \right].$$

In Algorithm 2, even for a given set of real-world observations \mathbf{X} the sampling input parameters θ_s 's may be random as they may depend on the randomly bootstrapped input parameters θ_b^* 's and may be

randomly selected from some input parameter space. We will analyze the expected value and variance of the ILR and MLR estimators subject to such randomness. For notational convenience, denote the random selection of R input parameters by $\tilde{\boldsymbol{\theta}}_R = \{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_R\}$. Proposition 2 shows that both the ILR and MLR estimators are conditionally unbiased given any bootstrapped input parameter $\boldsymbol{\theta}_b^*$ and their conditional variances given $\boldsymbol{\theta}_b^*$ converges to zero at $\mathcal{O}(R^{-1})$.

Proposition 2 Consider a given simulation model g , input distributions h_1, \dots, h_L , and an input parameter space Θ that satisfy Assumptions 1 and 2. Then, for any target input parameter $\boldsymbol{\theta}_b^* \in \Theta$,

$$\mathbb{E} \left[\hat{\mu}_R^{\text{ILR}}(\boldsymbol{\theta}_b^*) | \boldsymbol{\theta}_b^* \right] = \mathbb{E} \left[\hat{\mu}_R^{\text{MLR}}(\boldsymbol{\theta}_b^*) | \boldsymbol{\theta}_b^* \right] = \mu(\boldsymbol{\theta}_b^*).$$

Furthermore, if there exists $M \in \mathbb{R}$ such that $\sup\{\sigma^2(\boldsymbol{\theta}_b^*, \boldsymbol{\theta}_s) | \boldsymbol{\theta}_b^*, \boldsymbol{\theta}_s \in \Theta\} = M < \infty$, then

$$\text{Var} \left[\hat{\mu}_R^{\text{ILR}}(\boldsymbol{\theta}_b^*) | \boldsymbol{\theta}_b^* \right] = \mathcal{O}(R^{-1}) \text{ and } \text{Var} \left[\hat{\mu}_R^{\text{MLR}}(\boldsymbol{\theta}_b^*) | \boldsymbol{\theta}_b^* \right] = \mathcal{O}(R^{-1}).$$

The analysis presented so far is conditional on real-world observations $\mathbf{X} = \{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_L\}$. Proposition 3 shows that the marginal distribution of the green simulation estimators converge to a limiting distribution as the size of \mathbf{X} increases. Recall that $m_\ell := |\mathbf{X}_\ell|$ may be different for each ℓ . We define $m = \sum_\ell m_\ell$ and assume m_ℓ/m converges to a nonzero constant as $m \rightarrow \infty$ for each ℓ .

Proposition 3 Consider Algorithm 2 with real-world observation \mathbf{X} with size m . If R grows faster than m asymptotically, i.e., $R = \omega(m)$, then $\sqrt{m}(\hat{\mu}_R^{\text{GS}}(\boldsymbol{\theta}_b^*) - \mu(\hat{\boldsymbol{\theta}})) \Big| \mathbf{X} \xrightarrow{\mathcal{D}} \sqrt{m}(\mu(\hat{\boldsymbol{\theta}}) - \mu(\boldsymbol{\theta}^c))$, as $m \rightarrow \infty$, in probability over the sample space of \mathbf{X} , where GS may be ILR or MLR.

Proposition 3 states the convergence of the marginal distribution of $\hat{\mu}_R^{\text{GS}}(\boldsymbol{\theta}_b^*)$. If the conditional performance estimators $\hat{\mu}_R^{\text{GS}}(\boldsymbol{\theta}_b^*), b = 1, 2, \dots, B$ were independent, this would be sufficient to justify that the CIs constructed in Algorithm 2 are asymptotically valid for $\mu(\boldsymbol{\theta}^c)$. But $\hat{\mu}_R^{\text{GS}}(\boldsymbol{\theta}_b^*)$'s are dependent because they reuse the same set of R sample paths. With additional assumptions, such dependence does not affect the consistency of the resulting CIs based on empirical quantile estimators as stated in the following theorem.

Theorem 3 Given real-wold observations \mathbf{X} with sizes m_ℓ and $m = \sum_\ell m_\ell$ as well as a prescribed confidence level $0 < \alpha < 1$. If $\sqrt{m}(\mu(\hat{\boldsymbol{\theta}}) - \mu(\boldsymbol{\theta}^c)) \xrightarrow{\mathcal{D}} N(0, \tau^2)$ for some $\tau < \infty$, then the empirical α -quantile of $\{\hat{\mu}_R^{\text{GS}}(\boldsymbol{\theta}_b^*), b = 1, 2, \dots, B\}$ converges, in probability, to the α -quantile of $\{\mu(\boldsymbol{\theta}_b^*), b = 1, 2, \dots, B\}$ as $B \rightarrow \infty, m \rightarrow \infty$, and $R = \omega(m)$.

5 NUMERICAL EXPERIMENTS

To demonstrate the effectiveness of the proposed green simulation procedures, we compare the empirical coverage probabilities and the average widths of the CIs produced by different procedures as well as their runtime. In particular, we simulate M/M/1/10 systems whose interarrival and service time distributions have rates 0.9 and 1, respectively. We assume that these distributions are unknown can be estimated from $m = 50$ observations. The simulation is initialized with an empty system and the performance measure of interest is the steady-state expected time in system (TIS). For each simulation replication, the initial 50 observations are deleted for warm up. The SMC estimator for $\mu(\boldsymbol{\theta}^*)$ will average the 150 TISs after warm-up. For both ILR and MLR estimators, the simulation output in each replication is also the average of 150 TISs after warm-up, but all simulated interarrival and service times, including those in the warm-up, are used in the likelihood ratio calculations. Since the interarrival and service times in the M/M/1/10 system are exponentially distributed and are members of the exponential family, the sampling input parameters are selected such that conditions in Proposition 1 are satisfied. Specifically,

- ILR estimator is used when one sampling parameter is used; the sampling parameter is $\theta = \min\{\hat{\theta}, 2 \times \min\{\theta_1^*, \dots, \theta_B^*\}\}$, where θ may be the rate of the interarrival or service time distributions.

- MLR estimator is used when multiple sampling input parameters used; the R parameters are randomly sampled with replacement from a subset of the bootstrapped input parameters. The subset in question is $\{\theta \in \{\theta_1^*, \dots, \theta_B^*\} | \theta < 2 \times \min\{\theta_1^*, \dots, \theta_B^*\}\}$, where θ may be the rate of the interarrival or service time distributions.

We consider fixed simulation budget experiments where the total number of simulation replications is at most $C = 1,000$. For the number of bootstrapped input parameters, i.e., number of outer loops, we tested two cases $B = 100$ and $1,000$. The following procedures are considered in our experiments:

- Algorithm 1 without inner simulations, which is labeled “Oracle” in Table 1. For the M/M/1/k queueing simulation model, there exists an analytical oracle to calculate the TIS given the rates of the interarrival and service time distributions.
- Algorithm 1 with inner simulation, which is labeled “SMC” in Table 1. In this case the simulation budget is $C = BR$ so we tested two cases: $R = 10$ and $R = 1$.
- Algorithm 2 using ILR or MLR, which are labeled “ILR” and “MLR”, respectively, in Table 1. In these cases, the simulation budget $C = R$ and we tested two cases $R = 1,000$ and $R = 100$. Note that the experiments with $R = 100$ use only 10% of the given simulation budget.

All experiments are repeated 10,000 times and the target coverage was set to $1 - \alpha = 0.9$.

The first two columns of Table 1 show the empirical coverage probabilities and the average widths for CIs produced by different estimators. When the number of outer loop is insufficient, i.e., $R = 100$, the oracle method’s CI shows under-coverage, which is due to the errors associated with quantile estimations. Comparing the second rows to the first rows in both panels of Table 1 we see that the CIs produced by the SMC method are wide and has over-coverage. This is consistent with the common knowledge in the literature. Comparing the third rows to the second rows in both panels of Table 1 we see that, with the same simulation budget, the CIs produced by the MLR estimator has both narrower widths and better empirical coverage probabilities, i.e., closer to the empirical coverage probabilities of the CIs produced by the oracle method, than those of the SMC procedure. Moreover, adding the fourth rows in both panels of Table 1 to the above comparison we see that, with only 10% of the simulation budget, the MLR CIs still have better empirical coverage probabilities than the SMC CIs; the MLR CIs’ widths are still narrower than the SMC CIs’ widths when $B = 1,000$ and are only slightly wider than the latter when $B = 100$. The last two rows in both panels of Table 1 show that the ILR CIs show under-coverage and are wide; reducing the simulation runs aggravates under-coverage and widens the CIs. This last observation show some deficiencies of the ILR, which indicates a potential venue for future research.

The three columns on the right of Table 1 show a breakdown of runtime for different procedures: The “Simulation” column is the time for running the simulation replications. The 90% reduction of simulation budget for some ILR and MLR experiments is reflected in this column. The “Likelihood Ratios” column shows the time it takes to perform all the required likelihood ratio calculations. This column shows that the likelihood ratio calculations can be significant or even takes longer than the simulation runs. The computational burden for the MLR estimator is more prominent, especially when the number of simulation runs, R , is large. Reducing the computational burden of the MLR estimator is another potential venue for future research. Note that, when MLR uses only 10% of the simulation budget, i.e., $R = 100$, it takes only a fraction of the total runtime required for the SMC estimator. Specifically, in the top panel when $B = 1,000$, the MLR estimator produces CIs that are narrower and have better coverage probabilities compared to the SMC procedure, and reduce the total runtime by over 50% compared to the latter. In the bottom panel when $B = 100$, the CIs produced by the MLR estimator still have better coverage probabilities and are only slightly wider than those by the SMC procedure, and the total runtime is reduced by almost 75%.

In summary, taking into account for empirical coverage probabilities, CI widths, and the total runtime, we recommend using the MLR estimator in Algorithm 2 with reduced number of simulation runs for input uncertainty quantification.

Table 1: Empirical coverage probabilities, widths of confidence intervals, and breakdowns of runtime produced by different procedures. All results are calculated based on numerical experiments with 10,000 macro replications. Numbers in brackets are the respective standard errors.

	Coverage	Width	Runtime per macro replication (secs)		
			Simulation	LR Calculation	Total
No. of bootstraps = 1,000					
Oracle	89.97% (0.30%)	6.04 (0.015)	N/A	N/A	2.66E-03
SMC ($R = 1$)	97.53% (0.16%)	8.38 (0.020)	1.58	N/A	1.58
MLR ($R = 1,000$)	89.85% (0.30%)	6.05 (0.017)	1.56	36.34	37.90
MLR ($R = 100$)	90.26% (0.30%)	6.72 (0.025)	0.16	0.57	0.73
ILR ($R = 1,000$)	87.94% (0.33%)	7.69 (0.129)	1.60	1.89	3.49
ILR ($R = 100$)	83.89% (0.37%)	10.71 (0.184)	0.16	0.23	0.39
No. of bootstraps = 100					
Oracle	89.25% (0.31%)	5.97 (0.016)	N/A	N/A	4.22E-04
SMC ($R = 10$)	90.70% (0.29%)	6.24 (0.017)	1.58	N/A	1.58
MLR ($R = 1,000$)	89.14% (0.31%)	5.98 (0.017)	1.58	5.26	6.84
MLR ($R = 100$)	89.30% (0.31%)	6.48 (0.024)	0.16	0.08	0.24
ILR ($R = 1,000$)	86.01% (0.35%)	7.38 (0.078)	1.57	0.18	1.74
ILR ($R = 100$)	83.47% (0.37%)	10.50 (0.377)	0.16	0.02	0.18

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. Himmelspace, R. Pasupathy, O. Rose, and A. M. Uhrmacher, 1–10. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Barton, R. R. 2012. "Tutorial: Input Uncertainty in Output Analysis". In *Proceedings of the 2012 Winter Simulation Conference*, edited by C. Laroque, J. Himmelspace, R. Pasupathy, O. Rose, and A. M. Uhrmacher, 1–12. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Barton, R. R., H. Lam, and E. Song. 2018. "Revisiting Direct Bootstrap Resampling for Input Model Uncertainty". In *Proceedings of the 2018 Winter Simulation Conference*, edited by M. Rabe, A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, 1635–1645. 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. Mollabhasemi, E. Ruseel, and W. Biles, 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 *Proceedings 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.
- Beckman, R. J., and M. D. McKay. 1987. "Monte Carlo Estimation Under Different Distributions Using the Same Simulation". *Technometrics* 29(2):153–160.
- Cheng, R. C. H., and W. Holland. 1997. "Sensitivity of Computer Simulation Experiments to Errors in Input Data". *Journal of Statistical Computation and Simulation* 57(1-4):219–241.
- Cornuet, J.-M., J.-M. Martin, A. Mira, and C. P. Robert. 2012. "Adaptive Multiple Importance Sampling". *Scandinavian Journal of Statistics* 39(4):798–812.
- Dong, J., M. B. Feng, and B. L. Nelson. 2018. "Unbiased Metamodeling via Likelihood Ratios". In *Proceedings of the 2018 Winter Simulation Conference*, edited by M. Rabe, A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, 1778–1789. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Eckman, D. J., and B. M. Feng. 2018. "Green Simulation Optimization Using Likelihood Ratio Estimators". In *Proceedings of the 2018 Winter Simulation Conference*, edited by M. Rabe, A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, 4146–4147. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Feng, M., and J. Staum. 2015. "Green Simulation Designs for Repeated Experiments". In *Proceedings of the 2015 Winter Simulation Conference*, edited by L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Feng, M., and J. Staum. 2017. "Green Simulation: Reusing the Output of Repeated Experiments". *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 27(4):1–28.

- Glasserman, P., and X. Xu. 2014. "Robust Risk Measurement and Model Risk". *Quantitative Finance* 14(1):29–58.
- Glynn, P. W., and H. Lam. 2018. "Constructing Simulation Output Intervals Under Input Uncertainty via Data Sectioning". In *Proceedings of the 2018 Winter Simulation Conference*, edited by M. Rabe, A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, 1551–1562. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Hesterberg, T. C. 1988. *Advances in Importance Sampling*. Ph.D. thesis, Stanford University, Stanford, California.
- Kleijnen, J. P. C., and R. Y. Rubinstein. 1996. "Optimization and Sensitivity Analysis of Computer Simulation Models by the Score Function Method". *European Journal of Operational Research* 88(3):413–427.
- Lam, H. 2016. "Advanced Tutorial: Input Uncertainty and Robust Analysis in Stochastic Simulation". 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, 178–192. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Lam, H., and H. Qian. 2018. "Subsampling Variance for Input Uncertainty Quantification". In *Proceedings of the 2018 Winter Simulation Conference*, edited by M. Rabe, A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, 1611–1622. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- L'Ecuyer, P. 1990. "A Unified View of the IPA, SF, and LR Gradient Estimation Techniques". *Management Science* 36(11):1364–1383.
- L'Ecuyer, P. 1993. "Two Approaches for Estimating the Gradient in Functional Form". In *Proceedings of the 1993 Winter Simulation Conference*, edited by G. W. Evans, M. Mollabhasemi, E. Ruseel, and W. Biles, 338–346. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Lin, Y., E. Song, and B. L. Nelson. 2015. "Single-Experiment Input Uncertainty". *Journal of Simulation* 9(3):249–259.
- Martino, L., V. Elvira, D. Luengo, and J. Corander. 2015. "An Adaptive Population Importance Sampler: Learning from Uncertainty". *IEEE Transactions on Signal Processing* 63(16):4422–4437.
- Ng, S. H., and S. E. Chick. 2006. "Reducing Parameter Uncertainty for Stochastic Systems". *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 16(1):26–51.
- Owen, A., and Y. Zhou. 2000. "Safe and Effective Importance Sampling". *Journal of the American Statistical Association* 95(449):135–143.
- Rubinstein, R. Y., and A. Shapiro. 1993. *Discrete Event Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Method*. Chichester, England: John Wiley & Sons Inc.
- Song, E., and B. L. Nelson. 2015. "Quickly Assessing Contributions to Input Uncertainty". *IIE Transactions* 47(9):893–909.
- Song, E., B. L. Nelson, and C. D. Pegden. 2014. "Advanced Tutorial: Input Uncertainty Quantification". In *Proceedings of the 2014 Winter Simulation Conference*, edited by A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, 162–176. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Veach, E. 1997. *Robust Monte Carlo Methods for Light Transport Simulation*. Ph.D. thesis, Stanford University, Stanford, California.
- Veach, E., and L. J. Guibas. 1995. "Optimally Combining Sampling Techniques for Monte Carlo Rendering". In *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques*, 419–428. Los Angeles, CA, USA: Association for Computing Machinery.
- Xie, W., B. L. Nelson, and R. R. Barton. 2014. "A Bayesian Framework for Quantifying Uncertainty in Stochastic Simulation". *Operations Research* 62(6):1439–1452.
- Zhou, E., and T. Liu. 2018. "Online Quantification of Input Uncertainty for Parametric Models". In *Proceedings of the 2018 Winter Simulation Conference*, edited by M. Rabe, A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, 1587–1598. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

AUTHOR BIOGRAPHIES

BEN M. FENG is an assistant professor in actuarial science at the University of Waterloo. He earned his Ph.D. in the Department of Industrial Engineering and Management Sciences at Northwestern University. He is an Associate of the Society of Actuaries (ASA). His research interests include stochastic simulation design and analysis, optimization via simulation, nonlinear optimization, and financial and actuarial applications of simulation and optimization methodologies. His e-mail address is ben.feng@uwaterloo.ca and his website is <http://www.math.uwaterloo.ca/~mbfeng/>.

EUNHYE SONG is Harold and Inge Marcus Early Career Assistant Professor in the Department of Industrial and Manufacturing Engineering at the Penn State University. Her research interests include simulation design of experiments, simulation uncertainty and risk quantification, optimization via simulation under input model risk and large-scale discrete optimization via simulation. Her email address is eus358@psu.edu.