

SUBSAMPLING VARIANCE FOR INPUT UNCERTAINTY QUANTIFICATION

Henry Lam
Huajie Qian

Department of Industrial Engineering and Operations Research
Columbia University
500 W. 120th Street
New York, NY 10027, USA

ABSTRACT

In stochastic simulation, input uncertainty refers to the output variability arising from the statistical noise in specifying the input models. This uncertainty can be measured by a variance contribution in the output, which is estimated commonly via the bootstrap. However, due to the convolution of the simulation noise and the input noise, the computation effort required in the existing bootstrap schemes are typically substantial. This paper investigates a subsampling framework as a computation saver. We demonstrate the strengths of our subsampled bootstrap in terms of theoretical computation requirements, and substantiate them with numerical illustrations.

1 INTRODUCTION

Stochastic simulation is one of the most widely used analytic tools in operations research. It provides a flexible mean to approximate complex models and subsequently to inform decisions. See, for instance, Law et al. (1991) for applications in manufacturing, revenue management, service and operations systems etc. In practice, the simulation platform relies on input models that are typically observed or calibrated from data. These statistical noises can propagate to the output analysis, leading to significant errors and suboptimal decision-making. In the literature, this problem is commonly known as input uncertainty or extrinsic uncertainty.

This paper concerns the quantification of input uncertainty. In conventional simulation output analysis where the input model is completely pre-specified, the statistical errors come solely from the Monte Carlo noises, and it suffices to account for such noises in analyzing the output variability. When input uncertainty is present, such an analysis will undermine the actual variability. One common approach to quantify the additional uncertainty is to estimate the variance in the output that is contributed from the input noises (Song et al. 2014); for convenience, we call this the *input variance*. Moreover, the input variance can be used to identify models that are overly ambiguous and inform more data collection. They also collectively provide a building block to construct valid output confidence intervals (CIs) that account for combined input and simulation errors (Cheng and Holland 2004).

Bootstrap resampling is a common approach to estimate input variances. This applies most prominently in the nonparametric regime, namely when no assumptions are placed on the input parametric family. But it could also be used in the parametric case (where more alternatives are available). For example, Cheng and Holland (1997) proposes the variance bootstrap, and Song and Nelson (2015) studies the consistency of this strategy on a random-effect model that describes the uncertainty propagation. A bottleneck with using bootstrap resampling in estimating input variances, however, is the need to “outwash” the simulation noise, which often places substantial burden on the required simulation effort. More precisely, with both the input and the simulation noises, the bootstrap procedure typically comprises a two-layer sampling that

first resamples the input data (i.e., *outer* sampling), followed by running simulation replications using each resample (i.e., *inner* replications). On the other hand, the magnitude of the input variance typically scales with the amount of input data. It turns out that, relative to the data size, the procedure will require either a large outer bootstrap size or a large inner replication size in order to extinguish the effect of simulation noises. This requirement is often ignored in the existing literature by assuming the power to generate plenty of bootstrap samples (there are exceptions, e.g., the recent work of Song and Nelson (2018)). When simulating high-fidelity models, where each simulation takes hours or days, one may not afford to have such computation resources.

The main goal of this paper is to investigate *subsampling* as a simulation saver for input variance estimation. In particular, we show how a judicious use of subsampling can lead to estimation errors that depend much more favorably on the outer and inner sample sizes. This approach leverages the relation between the structure of the input variance in terms of the data size and the estimation error incurred in the required two-layer sampling. In the statistics literature, subsampling has been proposed to tackle situations where the conventional bootstrap fails, due to a lack of uniform convergence required for consistency, or in time series where data are non-i.i.d. (Politis et al. 1999). It typically involves sampling without replacement so that each resample consists of truthful observations. In contrary, our subsampling approach is introduced to reduce the simulation effort faced by the two-layer sampling. It serves to more efficiently deconvolute the effect of the simulation noise from the input noise. We will illustrate how to incorporate our scheme in the variance bootstrap that was studied in Cheng and Holland (1997) and Song and Nelson (2015), and will analyze the involved simulation complexity (i.e., minimally required total simulation replication size) and allocation rules.

We close this introduction with a brief review of other related work in input uncertainty. In the nonparametric case, Barton and Schruben (1993) and Barton and Schruben (2001) use the basic bootstrap to construct CIs, where the CI limits are determined from the quantiles of the bootstrap distributions. Lam and Qian (2016), Lam and Qian (2017) study the use of empirical likelihood, and Xie et al. (2018) studies nonparametric Bayesian methods. In the parametric case, Barton et al. (2013) studies the percentile bootstrap with a metamodel built in advance, a technique known as the metamodel-assisted bootstrap. Cheng and Holland (1997) studies the delta method, and Cheng and Holland (1998), Cheng and Holland (2004) reduce its computation burden via the so-called two-point method. Lin et al. (2015) and Song and Nelson (2018) study regression approaches to estimate sensitivity coefficients which are used to apply the delta method. Finally, Chick (2001), Zouaoui and Wilson (2004) and Xie et al. (2014) study variance estimation and interval construction from a Bayesian perspective. For general surveys on input uncertainty, readers are referred to Barton et al. (2002), Henderson (2003), Barton (2012), Song et al. (2014) and Lam (2016).

Due to space limit, this paper only gives a concise discussion on some of our results, and leaves further details and extensions, including a related approach that uses subsampling on a bootstrap aggregating scheme, to a full journal paper.

The remainder of the paper is as follows. Section 2 introduces the input uncertainty problem and explains the computation bottleneck in the existing bootstrap scheme. Section 3 presents our subsampling idea, procedures and the main statistical results. Section 4 reports some numerical experiments.

2 PROBLEM MOTIVATION

We describe the problem and our motivation. Section 2.1 first describes the input uncertainty problem, Section 2.2 discusses the existing bootstrap approach, and Section 2.3 discusses its computational barrier, thus motivating our subsampling investigation.

2.1 The Input Uncertainty Problem

Suppose there are m independent input processes driven by input distributions F_1, F_2, \dots, F_m . We consider a generic performance measure $\psi(F_1, \dots, F_m)$ that is simulable, i.e., given the input distributions, independent unbiased replications of ψ can be generated in a computer. As a primary example, think of F_1 and F_2 as the interarrival and service time distributions in a queue, and ψ is some output measure such as the mean queue length averaged over a time horizon.

The input uncertainty problem arises in situations where the input distributions F_1, \dots, F_m are unknown but real-world data are available. One then has to use their estimates $\hat{F}_1, \dots, \hat{F}_m$ to drive the simulation. Denote a point estimate of $\psi(F_1, \dots, F_m)$ as $\bar{\psi}(\hat{F}_1, \dots, \hat{F}_m)$, where typically we take

$$\bar{\psi}(\hat{F}_1, \dots, \hat{F}_m) = \frac{1}{R} \sum_{r=1}^R \hat{\psi}_r(\hat{F}_1, \dots, \hat{F}_m)$$

with $\hat{\psi}_r(\hat{F}_1, \dots, \hat{F}_m)$ being a conditionally unbiased simulation replication driven by $\hat{F}_1, \dots, \hat{F}_m$. This point estimate is affected by both the input statistical noises and the simulation noises. By conditioning on the estimated inputs (or viewing the point estimate as a random effect model with uncorrelated input and simulation noises), the variance of $\bar{\psi}(\hat{F}_1, \dots, \hat{F}_m)$ can be expressed as

$$\text{Var}[\bar{\psi}(\hat{F}_1, \dots, \hat{F}_m)] = \sigma_I^2 + \sigma_S^2$$

where

$$\sigma_I^2 = \text{Var}[\psi(\hat{F}_1, \dots, \hat{F}_m)] \tag{1}$$

is interpreted as the overall input variance, and

$$\sigma_S^2 = \frac{\mathbb{E}[\text{Var}[\hat{\psi}_r(\hat{F}_1, \dots, \hat{F}_m) | \hat{F}_1, \dots, \hat{F}_m]]}{R}$$

as the variance contributed from the simulation noises. Assuming that the estimates \hat{F}_i 's are consistent in estimating F_i 's, then, as n_i grows, σ_S^2 is approximately $\text{Var}[\hat{\psi}_r(F_1, \dots, F_m)]/R$ and can be estimated by taking the sample variance of all simulation replications (see, e.g., Cheng and Holland 1997). The key and the challenge in quantifying input uncertainty is to estimate σ_I^2 .

To this end, suppose further that for each input model i , we have n_i i.i.d. data $\{X_{i,1}, \dots, X_{i,n_i}\}$ generated from the distribution F_i . Typically, the overall input variance is decomposable, when n_i 's are large, into

$$\sigma_I^2 \approx \sum_{i=1}^m \frac{\sigma_i^2}{n_i} \tag{2}$$

where σ_i^2/n_i is the variance contributed from the data noise for model i . In the parametric case where \hat{F}_i comes from a parametric family containing the estimated parameters, this decomposition is well known from the delta method (Asmussen and Glynn 2007, Chapter 3). Here, σ_i^2/n_i is typically $\nabla_i \psi' \Sigma_i \nabla_i \psi$, where $\nabla_i \psi$ is the collection of sensitivity coefficients, i.e., the gradient, with respect to the parameters in model i , and Σ_i is the asymptotic estimation variance of the point estimates of these parameters (scaled reciprocally with n_i). In the nonparametric case where the empirical distribution $\hat{F}_i(x) := \sum_{j=1}^{n_i} \mathbf{1}\{X_{i,j} \leq x\}/n_i$ is used, (2) still holds under mild conditions (e.g., Assumption 1 in the sequel). The quantity σ_i^2 will be equal to $\text{Var}_{F_i}[g_i(X_i)]$, where $g_i(\cdot)$ is the influence function (Hampel 1974) of ψ with respect to the distribution F_i , defined on the value space of the input variate X_i , and $\text{Var}_{F_i}[\cdot]$ denotes the variance under F_i . The influence function can be viewed as a Gateaux derivative taken with respect to the probability distributions F_i 's (see Serfling 2009, Chapter 6), and dictates the first-order asymptotic variance of the plug-in estimate of ψ . Related derivations can be found in, e.g., Lam and Qian (2017).

Under further regularity conditions, a Gaussian approximation holds for $\bar{\psi}(\hat{F}_1, \dots, \hat{F}_m)$ so that

$$\bar{\psi}(\hat{F}_1, \dots, \hat{F}_m) \pm z_{1-\alpha/2} \sqrt{\sigma_F^2 + \sigma_S^2} \quad (3)$$

is an asymptotically tight $(1 - \alpha)$ -level CI for $\psi(F_1, \dots, F_m)$, where $z_{1-\alpha/2}$ is the standard normal $1 - \alpha/2$ quantile. This CI, which provides a bound-based alternative to quantify input uncertainty, again requires a statistically valid estimate of σ_F^2 (and σ_S^2).

We will investigate how to estimate $\sum_{i=1}^m \sigma_i^2/n_i$. We focus primarily on the nonparametric case. The most common estimation technique is bootstrap resampling, which we discuss next.

2.2 Bootstrap Resampling

Let \hat{F}_i^* represent the empirical distribution constructed using a bootstrap resample from the original data $\{X_{i,1}, \dots, X_{i,n_i}\}$ for input F_i , i.e., n_i points drawn by uniformly sampling with replacement from $\{X_{i,1}, \dots, X_{i,n_i}\}$. The bootstrap variance estimator is $\text{Var}_*[\psi(\hat{F}_1^*, \dots, \hat{F}_m^*)]$, where $\text{Var}_*[\cdot]$ denotes the variance over the bootstrap resamples from the data, under $\hat{F}_1, \dots, \hat{F}_m$.

The principle of bootstrap entails that $\text{Var}_*[\psi(\hat{F}_1^*, \dots, \hat{F}_m^*)] \approx \text{Var}[\psi(\hat{F}_1, \dots, \hat{F}_m)]$ and consistently approximates the asymptotic variance σ_F^2 under mild conditions. Here $\text{Var}_*[\psi(\hat{F}_1^*, \dots, \hat{F}_m^*)]$ is obtained from a (hypothetical) infinite number of bootstrap resamples and simulation runs per resample. In practice, however, one would need to use a finite bootstrap size and a finite simulation size. This comprises B conditionally independent bootstrap resamples of $\{\hat{F}_1^*, \dots, \hat{F}_m^*\}$, and R simulation replications driven by each realization of resampled input distributions. This generally incurs two layers of Monte Carlo errors.

Denote $\hat{\psi}_r(\hat{F}_1^b, \dots, \hat{F}_m^b)$ as the r -th simulation run driven by the b -th bootstrap resample $\{\hat{F}_1^b, \dots, \hat{F}_m^b\}$. Denote $\bar{\psi}^b$ as the average of the R simulation runs driven by the b -th resample, and $\bar{\psi}$ as the grand sample average from all the BR runs. An unbiased estimator for $\text{Var}_*[\psi(\hat{F}_1^*, \dots, \hat{F}_m^*)]$ is given by

$$\frac{1}{B-1} \sum_{b=1}^B (\bar{\psi}^b - \bar{\psi})^2 - \frac{V}{R} \quad (4)$$

where

$$V = \frac{1}{B(R-1)} \sum_{b=1}^B \sum_{r=1}^R (\hat{\psi}_r(\hat{F}_1^b, \dots, \hat{F}_m^b) - \bar{\psi}^b)^2.$$

To explain, the first term in (4) is an unbiased estimate of the variance of $\bar{\psi}^b$, which is $\text{Var}_*[\psi(\hat{F}_1^*, \dots, \hat{F}_m^*)] + (1/R)E_*[\text{Var}[\hat{\psi}_r(\hat{F}_1^*, \dots, \hat{F}_m^*)|\hat{F}_1^*, \dots, \hat{F}_m^*]]$ (where $E_*[\cdot]$ denotes the expectation on \hat{F}_i^* 's under \hat{F}_i 's) since $\bar{\psi}^b$ incurs both the bootstrap noise and the simulation noise. In other words, the variance of $\bar{\psi}^b$ is upward biased for $\text{Var}_*[\psi(\hat{F}_1^*, \dots, \hat{F}_m^*)]$. The second term in (4), namely V/R , removes this bias. This bias adjustment can be viewed as arising from $\text{Var}_*[\psi(\hat{F}_1^*, \dots, \hat{F}_m^*)]$ being the variance of a conditional expectation. Alternately, $\psi(\hat{F}_1^*, \dots, \hat{F}_m^*)$ can be viewed as a random effect model where each “group” corresponds to each realization of $\hat{F}_1^*, \dots, \hat{F}_m^*$, and (4) estimates the “between-group” variance in an analysis-of-variance (ANOVA). Formula (4) appeared in the input uncertainty literature, e.g., Cheng and Holland (1997), Song and Nelson (2015), Lin et al. (2015), and also in Zouaoui and Wilson (2004) in the Bayesian context. Algorithm 1 summarizes the procedure.

2.3 A Complexity Barrier

We explain intuitively the total number of simulation runs needed to ensure that the variance bootstrap depicted above can meaningfully estimate the input variance. For convenience, we call this number the *simulation complexity*. This turns out to be a lot bigger than the data size.

Algorithm 1 ANOVA-based Variance Bootstrap

Parameters: $B \geq 2, R \geq 2$

for $b = 1$ **to** B **do**

 Draw a resample $\{X_{i,1}^b, \dots, X_{i,n_i}^b\}$ for all i to obtain bootstrapped empirical distributions \widehat{F}_i^b

for $r = 1$ **to** R **do**

 Draw a $\widehat{\psi}_r(\widehat{F}_1^b, \dots, \widehat{F}_m^b)$

end for

 Compute $\bar{\psi}^b = \frac{1}{R} \sum_{r=1}^R \widehat{\psi}_r(\widehat{F}_1^b, \dots, \widehat{F}_m^b)$

end for

Compute $V = \frac{1}{B(R-1)} \sum_{b=1}^B \sum_{r=1}^R (\widehat{\psi}_r(\widehat{F}_1^b, \dots, \widehat{F}_m^b) - \bar{\psi}^b)^2$ and $\bar{\bar{\psi}} = \frac{1}{B} \sum_{b=1}^B \bar{\psi}^b$

Output $\widehat{\sigma}_{VB}^2 = \frac{1}{B-1} \sum_{b=1}^B (\bar{\psi}^b - \bar{\bar{\psi}})^2 - \frac{V}{R}$

To facilitate discussion, denote n as a scaling of the data size (i.e., assume n_i all grow linearly with n). Note that $\widehat{\sigma}_{VB}^2$ is unbiased for $\text{Var}_*[\psi(\widehat{F}_1^*, \dots, \widehat{F}_m^*)]$, which is approximately σ_I^2 . Thus, to analyze the error of $\widehat{\sigma}_{VB}^2$, it suffices to focus on its variance. An ANOVA-type analysis as in Sun et al. (2011) gives

$$\text{Var}_*[\widehat{\sigma}_{VB}^2] = \Theta_p \left(\frac{1}{Bn^2} + \frac{1}{BR^2} \right) \quad (5)$$

where Θ_p denotes stochastic boundedness (similarly, o_p will denote stochastic asymptotic negligibility). Stochastic boundedness is used here, as well as in later sections, since $\text{Var}_*[\widehat{\sigma}_{VB}^2]$ depends on the input data and hence is random. The two terms in (5) correspond to the variance coming from the bootstrap resampling and from the simulation noise respectively. From the $\Theta(1/n)$ magnitude of σ_I^2 and in turn $\text{Var}_*[\psi(\widehat{F}_1^*, \dots, \widehat{F}_m^*)]$, consistency requires $\widehat{\sigma}_{VB}^2$ to have a variance $o(1/n^2)$. This implies unfortunately that the total number of simulation runs, BR , must be $\omega(n)$, i.e., of order higher than the data size, because only then the second term in the bound (5) can be $o(1/n^2)$. We summarize with the following result.

Theorem 1 Under the assumptions in Section 3.2.1, the required simulation budget $N = BR$ to achieve (relative) consistency in estimating σ_I^2 by Algorithm 1, i.e., $\widehat{\sigma}_{VB}^2/\sigma_I^2 \xrightarrow{P} 1$, is $N = \omega(n)$.

Finally, we note that such a computational barrier does not only occur in variance bootstrap. For instance, the basic bootstrap studied in Barton and Schruben (1993), Barton and Schruben (2001) appears to also require a dominating inner replication size than the data size in order to obtain valid quantile estimates (the authors actually used one inner replication, but Barton (2012) commented that more is needed). Similarly, the regression approach in Song and Nelson (2018) requires simulation effort of order n^γ where $1 < \gamma < 2$. Likewise, the empirical likelihood framework studied in Lam and Qian (2017) also requires a higher order of simulation runs to estimate the influence function. Nonetheless, in this paper we focus only on how to reduce computation load in variance estimation.

3 PROCEDURES AND GUARANTEES IN THE SUBSAMPLING FRAMEWORK

This section presents our methodologies and results on subsampling. Section 3.1 first discusses how to incorporate subsampling in the variance bootstrap estimator depicted in Algorithm 1. Section 3.2 then provides theoretical guarantees of the subsampled procedure.

3.1 Proportionate Subsampled Variance Bootstrap

As explained before, the reason why the $\widehat{\sigma}_{VB}^2$ in Algorithm 1 requires a huge simulation effort, as implied by its mean square error (5), lies in the small scale of the input variance. In general, in order to estimate a quantity that is of order $1/n$, one must use a sample size more than n so that the estimation standard deviation relatively vanishes. This requirement manifests in the inner replication size in constructing $\widehat{\sigma}_{VB}^2$.

To reduce the inner replication size, we leverage the relation between the form of the input variance and the estimation variance depicted in (5) as follows. The approximate input variance contributed from model i , with data size n_i , has the form σ_i^2/n_i . If we use the variance bootstrap directly as in Algorithm 1, then we need an order more than n_i total simulation runs due to (5). Now, pretend that we have fewer data, say s_i , then the input variance will be σ_i^2/s_i , and the required simulation runs is now only of order higher than s_i . An estimate of σ_i^2/s_i , however, already gives us enough information in estimating σ_i^2/n_i , because we can merely rescale our estimate of σ_i^2/s_i by s_i/n_i to get an estimate of σ_i^2/n_i . This estimation can be done by subsampling the input distribution with size s_i . Thus, by subsampling to estimate σ_i^2/s_i , we can both use fewer simulation runs and also retain correct estimation via multiplying by a s_i/n_i factor.

To make the above argument more transparent, the bootstrap principle and the asymptotic approximation of the input variance imply that

$$\text{Var}_*[\psi(\widehat{F}_1^*, \dots, \widehat{F}_m^*)] = \sum_{i=1}^m \frac{\sigma_i^2}{n_i} (1 + o_p(1)).$$

The subsampling approach builds on the observation that a similar relation holds for

$$\text{Var}_*[\psi(\widehat{F}_{1,s_1}^*, \dots, \widehat{F}_{m,s_m}^*)] = \sum_{i=1}^m \frac{\sigma_i^2}{s_i} (1 + o_p(1))$$

where \widehat{F}_{i,s_i}^* denotes a bootstrapped distribution of size s_i (i.e., the empirical distribution that is uniformly sampled with replacement from $\{X_{i,1}, \dots, X_{i,n_i}\}$). If we let $s_i = \lfloor \theta n_i \rfloor$ for some $\theta > 0$ so that $s_i \rightarrow \infty$, then we have

$$\text{Var}_*[\psi(\widehat{F}_{1,\lfloor \theta n_1 \rfloor}^*, \dots, \widehat{F}_{m,\lfloor \theta n_m \rfloor}^*)] = \sum_{i=1}^m \frac{\sigma_i^2}{\theta n_i} (1 + o_p(1)).$$

Multiplying both sides with θ , we get

$$\theta \text{Var}_*[\psi(\widehat{F}_{1,\lfloor \theta n_1 \rfloor}^*, \dots, \widehat{F}_{m,\lfloor \theta n_m \rfloor}^*)] = \sum_{i=1}^m \frac{\sigma_i^2}{n_i} (1 + o_p(1)).$$

Note that the right hand side above is the original input variance of interest. This leads to our *proportionate subsampled variance bootstrap* summarized in Algorithm 2.

Algorithm 2 Proportionate Subsampled Variance Bootstrap

Parameters: $B \geq 2, R \geq 2, 0 < \theta \leq 1$

Compute $s_i = \lfloor \theta n_i \rfloor$ for all i

for $b = 1$ **to** B **do**

Draw a uniform subsample $\{X_{i,1}^b, \dots, X_{i,s_i}^b\}$ with replacement from data for all i , which induces the empirical distribution \widehat{F}_{i,s_i}^b

for $r = 1$ **to** R **do**

Draw a $\widehat{\psi}_r(\widehat{F}_{1,s_1}^b, \dots, \widehat{F}_{m,s_m}^b)$

end for

Compute $\bar{\psi}^b = \frac{1}{R} \sum_{r=1}^R \widehat{\psi}_r(\widehat{F}_{1,s_1}^b, \dots, \widehat{F}_{m,s_m}^b)$

end for

Compute $V = \frac{1}{B(R-1)} \sum_{b=1}^B \sum_{r=1}^R (\widehat{\psi}_r(\widehat{F}_{1,s_1}^b, \dots, \widehat{F}_{m,s_m}^b) - \bar{\psi}^b)^2$ and $\bar{\psi} = \frac{1}{B} \sum_{b=1}^B \bar{\psi}^b$

Output $\widehat{\sigma}_{SVB}^2 = \theta \left(\frac{1}{B-1} \sum_{b=1}^B (\bar{\psi}^b - \bar{\psi})^2 - \frac{V}{R} \right)$

3.2 Statistical Guarantees

We present our key theoretical guarantees for Algorithm 2. Section 3.2.1 first states our assumptions on the performance measure. Sections 3.2.2 then presents the theories on estimation consistency, simulation complexity and budget allocation for the proportionate subsampled variance bootstrap.

3.2.1 Regularity Assumptions

We assume the data sizes of different input models are of the same order. To be specific, $\max_i n_i / \min_i n_i \leq c$ for some finite constant $c > 0$ as $n_i \rightarrow \infty$. We first make an assumption on the smoothness of the performance measure:

Assumption 1 (Smoothness at the true input distributions) The performance measure ψ , as a functional of the input distributions, has first order influence functions $g_i : \mathbb{R} \rightarrow \mathbb{R}, i = 1, \dots, m$ at F_1, \dots, F_m such that $\mathbb{E}_{F_i}[g_i(X_i)] = 0, 0 < \text{Var}_{F_i}[g_i(X_i)] < \infty$ for all i and the plug-in estimator $\psi(\widehat{F}_1, \dots, \widehat{F}_m)$ admits the Taylor expansion

$$\psi(\widehat{F}_1, \dots, \widehat{F}_m) = \psi(F_1, \dots, F_m) + \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} g_i(X_{i,j}) + \varepsilon \quad (6)$$

where the remainder ε satisfies $\mathbb{E}[\varepsilon^2] = o(n^{-1})$.

Assumption 1 entails that the performance measure is differentiable and the error of the linear approximation is negligible in an asymptotic sense. Specifically, the linear term in (6) is $\Theta_p(n^{-1/2})$ by the central limit theorem, whereas the error ε is $o_p(n^{-1/2})$. Hence the variance of the linear term contributes dominantly to the overall input variance as n_i 's are large, which gives rise to the additive form of the input variance thanks to the independence among the input models. Thus we have:

Proposition 1 Under Assumption 1, the input variance σ_I^2 defined in (1) takes the form $\sigma_I^2 = \sum_{i=1}^m \sigma_i^2 / n_i + o(n^{-1})$, where each $\sigma_i^2 = \text{Var}_{F_i}[g_i(X_i)]$ is the variance of the i -th influence function.

The next assumption concerns the bootstrap principle under subsampling:

Assumption 2 (Smoothness at the empirical input distributions) The performance measure ψ has first order influence functions $\hat{g}_i : \mathbb{R} \rightarrow \mathbb{R}, i = 1, \dots, m$ at $\widehat{F}_1, \dots, \widehat{F}_m$ such that

$$\frac{1}{n_i} \sum_{j=1}^{n_i} \hat{g}_i(X_{i,j}) = 0, \hat{\sigma}_i^2 := \frac{1}{n_i} \sum_{j=1}^{n_i} \hat{g}_i^2(X_{i,j}) \xrightarrow{p} \sigma_i^2, \frac{1}{n_i} \sum_{j=1}^{n_i} \hat{g}_i^4(X_{i,j}) = O_p(1) \text{ for all } i \quad (7)$$

and the bootstrapped performance measure with subsample size s_i admits the Taylor expansion

$$\psi(\widehat{F}_{1,s_1}^*, \dots, \widehat{F}_{m,s_m}^*) = \psi(\widehat{F}_1, \dots, \widehat{F}_m) + \sum_{i=1}^m \frac{1}{s_i} \sum_{k=1}^{s_i} \hat{g}_i(X_{i,k}^*) + \varepsilon^*$$

where the remainder ε^* satisfies

$$\mathbb{E}_*[(\varepsilon^* - \mathbb{E}_*[\varepsilon^*])^4] = o_p\left(\left(\sum_{i=1}^m \frac{1}{s_i}\right)^3\right).$$

The \hat{g}_i 's in Assumption 2 are the empirical influence functions and are expected to approach the true influence functions g_i 's in Assumption 1. This leads to the first and second conditions in (7). The fourth moment conditions on the empirical influence functions, i.e., the last condition in (7), and on the remainder ε^* are needed for controlling the variance of our variance estimator.

To introduce our assumptions on the simulation noise, denote by τ^2 the simulation variance of $\hat{\psi}$ under the true input models, i.e.,

$$\tau^2 = \mathbb{E}_{F_1, \dots, F_m}[(\hat{\psi}(F_1, \dots, F_m) - \psi(F_1, \dots, F_m))^2].$$

Similarly, we use $\hat{\tau}^2$ and $\hat{\tau}_*^2$ to represent the variance of $\hat{\psi}$ conditional on the empirical input models and the bootstrapped input models

$$\hat{\tau}^2 = \mathbb{E}_{\hat{F}_1, \dots, \hat{F}_m} [(\hat{\psi}(\hat{F}_1, \dots, \hat{F}_m) - \psi(\hat{F}_1, \dots, \hat{F}_m))^2], \quad \hat{\tau}_*^2 = \mathbb{E}_{\hat{F}_{1,s_1}^*, \dots, \hat{F}_{m,s_m}^*} [(\hat{\psi}(\hat{F}_{1,s_1}^*, \dots, \hat{F}_{m,s_m}^*) - \psi(\hat{F}_{1,s_1}^*, \dots, \hat{F}_{m,s_m}^*))^2]$$

and use $\hat{\mu}_4^*$ to represent the central fourth moment of $\hat{\psi}$ under the bootstrapped input models

$$\hat{\mu}_4^* = \mathbb{E}_{\hat{F}_{1,s_1}^*, \dots, \hat{F}_{m,s_m}^*} [(\hat{\psi}(\hat{F}_{1,s_1}^*, \dots, \hat{F}_{m,s_m}^*) - \psi(\hat{F}_{1,s_1}^*, \dots, \hat{F}_{m,s_m}^*))^4].$$

The assumptions on the simulation noise are:

Assumption 3 (Convergence of variance) $\hat{\tau}^2 \xrightarrow{P} \tau^2$ and $\mathbb{E}_*[(\hat{\tau}_*^2 - \hat{\tau}^2)^2] = o_p(1)$.

Assumption 4 (Boundedness of the fourth moment of $\hat{\psi}$) $\mathbb{E}_*[\hat{\mu}_4^*] = O_p(1)$.

Assumption 3 stipulates that the variance of the Monte Carlo estimator $\hat{\psi}$ as a functional of the underlying input models is smooth enough in the inputs. Assumption 4 is again a fourth moment condition used to control the variance of our variance estimator.

Although the above assumptions may look complicated, one can verify that all of them hold for a wide class of performance measures, namely those over finite time horizons in the form

$$\psi(F_1, \dots, F_m) = \mathbb{E}_{F_1, \dots, F_m} [h(\mathbf{X}_1, \dots, \mathbf{X}_m)] \quad (8)$$

where each T_i is a deterministic time, the i -th input process $\mathbf{X}_i = (X_i(1), \dots, X_i(T_i))$ consists of T_i i.i.d. variables distributed under F_i , and h is some performance function. An unbiased simulation estimate $\hat{\psi}_r$ of the performance measure simply outputs a copy of $h(\mathbf{X}_1, \dots, \mathbf{X}_m)$. Assume the performance function h satisfies the following:

Assumption 5 For each i , $0 < \text{Var}_{F_i}[\sum_{t=1}^{T_i} \mathbb{E}_{F_1, \dots, F_m} [h(\mathbf{X}_1, \dots, \mathbf{X}_m) | X_i(t) = X_i]] < \infty$.

Assumption 6 For each i let $I_i = (I_i(1), \dots, I_i(T_i))$ be a sequence of indices such that $1 \leq I_i(t) \leq T_i$, and $\mathbf{X}_{i,I_i} = (X_i(I_i(1)), \dots, X_i(I_i(T_i)))$. Assume $\max_{I_1, \dots, I_m} \mathbb{E}_{F_1, \dots, F_m} [|h(\mathbf{X}_{1,I_1}, \dots, \mathbf{X}_{m,I_m})|^4] < \infty$.

Then we have:

Proposition 2 Under Assumptions 5 and 6, we have Assumptions 1-4 hold for the finite-horizon performance measure ψ given by (8).

3.2.2 Simulation Complexity and Optimal Allocation

Recall from Section 3.1 that the proportionate subsampled variance takes the form

$$\sigma_{SVB}^2 = \theta \text{Var}_* [\psi(\hat{F}_{[\theta n_1], 1}^*, \dots, \hat{F}_{[\theta n_m], m}^*)] \quad (9)$$

where $0 < \theta \leq 1$ is the subsample ratio, and $[\cdot]$ outputs the integer part of a number. In Section 3.1 we have presented a heuristic argument on consistency of (9). Here we give a formal statement:

Theorem 2 Under Assumptions 1 and 2, if the subsample ratio $\theta = \omega(n^{-1})$, then the subsampled variance (9) is consistent as $n_i \rightarrow \infty$, i.e. $\sigma_{SVB}^2 / \sigma_I^2 \xrightarrow{P} 1$.

The requirement $\theta = \omega(n^{-1})$ implies that $s_i \rightarrow \infty$, which is natural as one needs minimally an increasing subsample size to ensure the consistency of our estimator. It turns out that this minimal requirement is enough to ensure consistency even relative to the magnitude of σ_I^2 .

Now we turn to the discussion on the Monte Carlo error, from both the outer bootstrapping and the inner simulation, for $\hat{\sigma}_{SVB}^2$ in Algorithm 2. The following lemma characterizes the level of Monte Carlo noise in terms of the conditional mean square error.

Lemma 1 Under Assumptions 1-4, if $B = \omega(1)$, $\theta = \omega(n^{-1})$, then the mean square error conditional on the input data is

$$\mathbb{E}_*[(\hat{\sigma}_{SVB}^2 - \sigma_{SVB}^2)^2] = \frac{2}{B} \left(\sum_{i=1}^m \frac{\sigma_i^2}{n_i} + \frac{\tau^2 \theta}{R} \right)^2 (1 + o_p(1)). \quad (10)$$

The target quantity σ_I^2 here is of order n^{-1} by Proposition 1, and hence the Monte Carlo noise of our variance estimate has to vanish faster than n^{-1} in order to achieve consistency. This leads to the following simulation complexity of Algorithm 2:

Theorem 3 Under Assumptions 1-4, the variance estimate $\hat{\sigma}_{SVB}^2$ is consistent, i.e. $\hat{\sigma}_{SVB}^2/\sigma_I^2 \xrightarrow{P} 1$ if and only if the parameters B, R, θ of Algorithm 2 are chosen such that

$$B = \omega(1), BR^2 = \omega((\theta n)^2), \theta = \omega\left(\frac{1}{n}\right). \quad (11)$$

In particular, the minimum requirement of total simulation budget is $N = BR = \omega(1)$ by choosing $\theta = \omega(n^{-1})$.

Theorem 3 tells us what values of B, R would work in general. For a given subsample ratio θ , the first two conditions in (11) implies a simulation budget of $\omega(\theta n)$ for consistent estimation, which recovers the result in Theorem 1 by setting $\theta = \Theta(1)$. Next, we establish the optimal configuration of B, R, θ to minimize the Monte Carlo error:

Theorem 4 Under Assumptions 1-4, given a simulation budget N and a subsample ratio θ such that $N = \omega(\theta n)$ and $\theta = \omega(n^{-1})$, the optimal outer and inner sizes that minimize the conditional mean square error (10) are

$$B^* = \frac{N}{R^*} (1 + o_p(1)), R^* = \frac{\theta \tau^2}{\sigma_I^2} (1 + o_p(1))$$

giving a conditional mean square error $\mathbb{E}_*[(\hat{\sigma}_{SVB}^2 - \sigma_{SVB}^2)^2] = (1 + o_p(1)) 8\theta \tau^2 \sigma_I^2 / N$.

Since $\sigma_I^2 = \Theta(n^{-1})$, the theorem indicates that the optimal choice of inner size R is of order $\Theta(\theta n)$, the same as the subsample size, and the outer size B is then chosen accordingly.

Lastly, we show a result on the overall errors in using $\hat{\sigma}_{SVB}^2$ to estimate σ_I^2 . This comprises of both the Monte Carlo error in approximating σ_{SVB}^2 , and the input error between σ_{SVB}^2 and σ_I^2 . The latter is not affected by the choices of B, R , but by the subsample ratio θ . The optimal choice of θ in minimizing the gross error balances a trade-off between the above two sources of errors. Due to space limit, we skip the full details here. We show a result specialized to the case of finite-horizon performance measures:

Theorem 5 Suppose ψ is a finite-horizon performance measure (8), and Assumptions 5 and 6 hold. The optimal θ, R, B given simulation budget $N = \omega(1)$ in minimizing the gross error $\hat{\sigma}_{SVB}^2 - \sigma_I^2$ are respectively

$$\begin{cases} \theta^* = \Theta(N^{1/3} n^{-1}) & \text{if } N \leq n^{3/2} \\ \Theta(n^{-1/2}) \leq \theta^* \leq \Theta(N n^{-2} \wedge 1) & \text{if } N > n^{3/2}, R^* = \Theta(\theta^* n), B^* = \frac{N}{R^*} \end{cases}$$

giving the minimum mean square error $\mathbb{E}[(\hat{\sigma}_{SVB} - \sigma_I^2)^2] = \Theta(1/((N^{2/3} \wedge n)n^2))$.

4 NUMERICAL EXPERIMENTS

We consider an M/M/1 queuing system with arrival rate 0.5 and service rate 1. Suppose the system is empty at time zero. The performance measure of interest is the probability that the waiting time of the 20-th arrival exceeds 2 units of time, whose true value is approximately 0.182. Specifically, the system has two input distributions, i.e. the inter-arrival time distribution $F_1 = \text{Exp}(0.5)$ and the service time distribution $F_2 = \text{Exp}(1)$, for which we have n_1 and n_2 i.i.d. data available respectively. If A_t is the inter-arrival time between the t -th and $(t + 1)$ -th arrivals, and S_t is the service time for the t -th arrival, then the system output

$$\psi(F_1, F_2) = \mathbb{E}_{F_1, F_2}[\mathbf{1}\{W_{20} > 2\}]$$

where the waiting time W_{20} is calculated by the Lindley recursion $W_1 = 0, W_{t+1} = \max\{W_t + S_t - A_t, 0\}$.

We test performance of Algorithms 1 and 2 under different input data sizes. For each input data size, 1000 95%-level CIs in the form of (3) are constructed, each from an independently generated input data set. We allocate a total number of 1000 simulation runs to estimate σ_I^2 using either Algorithm 1 or 2, and another 500 simulation runs driven by the empirical input distributions to compute the point estimator $\bar{\psi}(\hat{F}_1, \dots, \hat{F}_m)$. The simulation variance σ_S^2 is calculated as $(V/2 + \bar{\tau}^2/2)/500$, where V is the within-group variance from Algorithm 1 or 2 and $\bar{\tau}^2$ is the sample variance of the 500 simulation replications.

Tables 1, 2 and 3 show, for different algorithms and settings, the coverage, mean and standard deviation of the CI lengths, the estimated ratio between input and simulation standard deviations, and the number of times that the algorithm outputs a negative variance estimate (because of the debiasing, the estimate is not guaranteed to be always positive). The ratio between the input and simulation variance under each setting demonstrates that the input uncertainty is not negligible relative to the stochastic noise.

We make a few observations. First, the use of subsampling reduces the variability of the variance estimate when compared with standard variance bootstrap, and in turn gives rise to more accurate CIs. This is supported by the large counts of negative variance estimates in Table 1 versus negligible counts in Table 2 and 3, and the larger standard deviations of the CI lengths in Table 1 (7 times larger than in Table 2 and 14 times larger than in Table 3 when $n_1 = 4000, n_2 = 2000$). There are two cases in Table 1 that standard bootstrap has 90.7% and 94.0% coverage. Although they look close to 95%, this could be because we have reset the variance estimate to 0 whenever it is negative (which happens in almost half of the cases) and such an artificial correction makes the variance estimate upward biased, leading to a larger but possibly incorrect coverage probability. Second, with the simulation budget fixed, the coverage probabilities in both Tables 2 and 3 increase from around 85% to the nominal level 95% as the data size grows, validating Theorem 3. Thirdly, as shown in Table 3, under the subsample ratio $(N^{1/3} \wedge n^{1/2})n^{-1}$ as suggested by Theorem 5 our subsampled bootstrap exhibits reasonably good performance. However, the hidden constant there can be difficult to estimate in general. Table 2 indicates that simply using a subsample size around 30 seems a good enough compromise for this particular example.

Table 1: Algorithm 1 with $B = 50, R = 20$.

input data sizes	coverage estimate	mean CI length	std. CI length	$\frac{\sigma_I}{\sigma_S}$	# neg var
$n_1 = 60, n_2 = 30$	84.3%	0.419	0.180	6.96	5
$n_1 = 200, n_2 = 100$	87.5%	0.245	0.093	3.77	43
$n_1 = 600, n_2 = 300$	84.5%	0.150	0.070	2.26	236
$n_1 = 2000, n_2 = 1000$	90.7%	0.122	0.061	1.76	409
$n_1 = 4000, n_2 = 2000$	94.0%	0.113	0.055	1.56	458

Table 2: Algorithm 2 with $B = 50, R = 20, \theta = \frac{30}{n_2}$.

input data sizes	coverage estimate	mean CI length	std. CI length	$\frac{\sigma_I}{\sigma_S}$	# neg var
$n_1 = 60, n_2 = 30$	85.2%	0.424	0.183	7.02	12
$n_1 = 200, n_2 = 100$	89.9%	0.249	0.069	3.78	1
$n_1 = 600, n_2 = 300$	94.7%	0.158	0.028	2.18	0
$n_1 = 2000, n_2 = 1000$	93.8%	0.103	0.012	1.19	0
$n_1 = 4000, n_2 = 2000$	94.9%	0.087	0.008	0.84	0

Table 3: Algorithm 2 with $B = 100$, $R = 10$, $\theta = (10 \wedge \sqrt{(n_1 + n_2)/2}) / ((n_1 + n_2)/2)$.

input data sizes	coverage estimate	mean CI length	std. CI length	$\frac{\sigma_l}{\sigma_s}$	# neg var
$n_1 = 60, n_2 = 30$	87.3%	0.420	0.123	6.85	0
$n_1 = 200, n_2 = 100$	92.0%	0.240	0.043	3.73	0
$n_1 = 600, n_2 = 300$	94.4%	0.150	0.017	2.16	0
$n_1 = 2000, n_2 = 1000$	94.5%	0.098	0.007	1.18	0
$n_1 = 4000, n_2 = 2000$	94.0%	0.083	0.004	0.84	0

5 CONCLUSION

We have explained how estimating input variances in stochastic simulation can require large computation effort when using conventional bootstrapping. This arises as the bootstrap involves a two-layer sampling, which adds up to a total effort of larger order than the data size in order to achieve relative consistency. To alleviate this issue, we have proposed a subsampling method that leverages the relation between the structure of input variance and the estimation error from the two-layer sampling, so that the resulting total effort depends much less on the data size. We have presented some theoretical and numerical results to support our claims. Future work comprises a more comprehensive investigation of our subsampling scheme, including its generalization to a bootstrap aggregating approach, and more extensive numerical studies.

ACKNOWLEDGMENTS

We gratefully acknowledge support from the National Science Foundation under grants CMMI-1542020, CMMI-1523453 and CAREER CMMI-1653339/1834710.

REFERENCES

- Asmussen, S., and P. W. Glynn. 2007. *Stochastic Simulation: Algorithms and Analysis*, Volume 57. Springer Science & Business Media.
- Barton, R. R. 2012. "Tutorial: Input Uncertainty in Output Analysis". In *Proceedings of the 2012 Winter Simulation Conference*, edited by C. Laroque et al., 1–12. Piscataway, New Jersey: IEEE.
- 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 et al., 503–508. Piscataway, New Jersey: IEEE.
- 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 et al., 372–378. Piscataway, New Jersey: IEEE.
- Barton, R. R., S. E. Chick, R. C. Cheng, S. G. Henderson, A. M. Law, B. W. Schmeiser, L. M. Leemis, L. W. Schruben, and J. R. Wilson. 2002. "Panel Discussion on Current Issues in Input Modeling". In *Proceedings of the 2002 Winter Simulation Conference*, edited by E. Yücesan et al., 353–369. Piscataway, New Jersey: IEEE.
- Barton, R. R., B. L. Nelson, and W. Xie. 2013. "Quantifying Input Uncertainty via Simulation Confidence Intervals". *INFORMS Journal on Computing* 26(1):74–87.
- Cheng, R. C., 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.
- Cheng, R. C., and W. Holland. 1998. "Two-Point Methods for Assessing Variability in Simulation Output". *Journal of Statistical Computation Simulation* 60(3):183–205.
- Cheng, R. C., and W. Holland. 2004. "Calculation of Confidence Intervals for Simulation Output". *ACM Transactions on Modeling and Computer Simulation* 14(4):344–362.

- Chick, S. E. 2001. "Input Distribution Selection for Simulation Experiments: Accounting for Input Uncertainty". *Operations Research* 49(5):744–758.
- Hampel, F. R. 1974. "The Influence Curve and Its Role in Robust Estimation". *Journal of the American Statistical Association* 69(346):383–393.
- Henderson, S. G. 2003. "Input Modeling: 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 et al., 90–100. Piscataway, New Jersey: IEEE.
- 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 et al., 178–192. Piscataway, New Jersey: IEEE.
- 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 et al., 791–802. Piscataway, New Jersey: IEEE.
- Lam, H., and H. Qian. 2017. "Optimization-Based Quantification of Simulation Input Uncertainty via Empirical Likelihood". *arXiv preprint arXiv:1707.05917*.
- Law, A. M., W. D. Kelton, and W. D. Kelton. 1991. *Simulation Modeling and Analysis*, Volume 2. McGraw-Hill New York.
- Lin, Y., E. Song, and B. Nelson. 2015. "Single-Experiment Input Uncertainty". *Journal of Simulation* 9(3):249–259.
- Politis, D. N., J. P. Romano, and M. Wolf. 1999. *Subsampling*. Springer.
- Serfling, R. J. 2009. *Approximation Theorems of Mathematical Statistics*, Volume 162. John Wiley & Sons.
- Song, E., and B. L. Nelson. 2015. "Quickly Assessing Contributions to Input Uncertainty". *IIE Transactions* 47(9):893–909.
- Song, E., and B. L. Nelson. 2018. "Input-Output Uncertainty Comparisons for Optimization via Simulation". *Accepted in Operations Research*.
- 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 et al., 162–176. Piscataway, New Jersey: IEEE.
- 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.
- 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.
- Xie, W., C. Li, and P. Zhang. 2018. "A Bayesian Nonparametric Hierarchical Framework for Uncertainty Quantification in Simulation". *Working paper, available at homepages.rpi.edu*.
- 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

HENRY LAM is an Associate Professor in the Department of Industrial Engineering and Operations Research at Columbia University. His research focuses on Monte Carlo simulation, risk and uncertainty quantification, and stochastic optimization. His email address is kh12114@columbia.edu.

HUAJIE QIAN is a Ph.D. student in the Department of Industrial Engineering and Operations Research at Columbia University. His research interest lies in simulation uncertainty quantification, data-driven simulation analysis, and stochastic optimization. His email address is h.qian@columbia.edu.