Proceedings of the 2022 Winter Simulation Conference
B. Feng, G. Pedrielli, Y. Peng, S. Shashaani, E. Song, C.G. Corlu, L.H. Lee, E.P. Chew, T. Roeder, and P. Lendermann, eds.

CHEAP BOOTSTRAP FOR INPUT UNCERTAINTY QUANTIFICATION

Henry Lam

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

ABSTRACT

When a simulation model contains input distributions that need to be calibrated from external data, proper simulation output analysis needs to account for not only the noises from the Monte Carlo sample generation, but also the statistical noises from these input data. The latter issue is known commonly as the input uncertainty in the literature. An array of methods have been proposed to address input uncertainty, but one recurrent challenge faced by many of these methods is the demanding simulation load. In this paper, we present a method, based on a sort of modified bootstrap, to handle input uncertainty with multiplicatively less computation than these existing methods. In particular, this "cheap" bootstrap is able to construct confidence intervals that account for both input data and Monte Carlo noise efficiently by substantially reducing the number of outer samples in a nested procedure.

1 INTRODUCTION

Stochastic simulation models often contain parameters or input distributions that are unknown in advance and need to be calibrated from external data. The statistical errors from these calibrations could propagate to the simulation outputs and impact output analysis, on top of the Monte Carlo noises incurred from running the simulation itself. A statistically valid output analysis therefore needs to account for both the input data and the Monte Carlo noises. This problem, which is beyond classical output analysis that only focuses on the Monte Carlo noises, is known commonly as the input uncertainty problem in the simulation literature. It has gathered growing interests over recent years. See, e.g., the surveys Henderson (2003), Barton (2012), Song et al. (2014), Lam (2016), Corlu et al. (2020), Barton et al. (2022), and Nelson (2013) Chapter 7.

To address input uncertainty, an array of methods have been proposed, including various bootstrap methods (Barton and Schruben 1993; Barton and Schruben 2001; Cheng and Holloand 1997; Song and Nelson 2015; Lam and Qian 2022), delta method or infinitesimal jackknife (Cheng and Holland 1998; Cheng and Holland 2004; Lin et al. 2015; Song and Nelson 2019; Lam and Qian 2019; Lam et al. 2022), batching or sectioning (Glynn and Lam 2018), and Bayesian approaches (Chick 2001; Zouaoui and Wilson 2003; Zouaoui and Wilson 2004; Xie et al. 2014) that are conceptually related to the bootstrap. Despite the growing literature, a recurrent challenge faced by many of these methods is the demanding simulation effort. This challenge is arguably fundamental in the input uncertainty problem, and in fact is what distinguishes the problem from standard statistical inference: Roughly speaking, the two sources of errors, from the input data and Monte Carlo respectively, are entangled in a convoluted fashion and, in order to utilize standard statistical procedures, one needs to use a large amount of simulation effort to wash away the Monte Carlo error so that addressing input uncertainty reduces essentially to a standard inference problem. As we will explain later, this overwhelming simulation effort is typically manifested as the requirement of an overall simulation size comparable or of higher order than the input data size. In

practice, the latter means the required simulation size can be "a lot" more than the input data size, and thus can be costly if the simulation model is large-scale and expensive to run.

The goal of this paper is to introduce a sort of modified bootstrap procedure to handle input uncertainty with substantially less simulation effort than many of the existing methods. In particular, compared to the bootstrap methods used in the input uncertainty literature, the reduction in simulation effort is *multiplicative*. To explain in more detail, the task of addressing input uncertainty can be cast as the construction of output confidence intervals (CIs) that are statistically valid under both input and Monte Carlo noises, or the closely related task of estimating the output variance. Due to the two sources of noises, using the bootstrap to construct a CI or estimate the variance would require a nested simulation. Namely, in the outer layer, we resample the input data many times to obtain resample input distributions. Then, we use each resample input distribution to drive many simulation replications to obtain the outputs, which constitutes the inner layer. A CI is then constructed by properly aggregating the averages of outputs in the inner layer. The leveraging of bootstrap approaches from statistics, including quantile-based methods such as the basic bootstrap and the percentile bootstrap (Efron and Tibshirani 1994; Davison and Hinkley 1997), and also the standard error or variance bootstrap (Efron 1992), would require adequate sample sizes in both the outer and inner layers, thus leading to a multiplicative demand of simulation. The crux of the method proposed in this paper is to *reduce the outer-layer sample size to a very low number, while retaining desirable statistical guarantees, thus effectively removing the multiplicative computation effort*.

Our method uses, and can be viewed as a multi-input generalization, of the cheap bootstrap method proposed in Lam (2022). This latter method considers a twist of the standard bootstrap principle that allows the use of very few, i.e., as low as one, resample to construct statistically valid CIs in estimating statistical functionals. To explain, standard bootstrap methods are based on the principle that the resample distribution approximates the sampling distribution of the original estimate. The key is that while the latter is difficult to obtain, the former can be estimated readily by using Monte Carlo to generate many resamples, thus forming an approximation of the sampling distribution which then can be used for inference. The cheap bootstrap instead utilizes the independence between the original estimate and all resample estimates and, coupled with asymptotic normality, it can construct valid CIs via a pivotal t-statistic using as few as one resample. Our proposed method here takes this cheap bootstrap approach further to consider a functional of multiple input distributions, with simulation noises. This adaption allows us to use, in principle, as few as one resample in the outer layer.

In the following, we first introduce in detail the input uncertainty problem (Section 2) and describe the computational challenges (Section 3). We then introduce the cheap bootstrap (Section 4) and its use in handling input uncertainty (Section 5). We contrast our method to some recent approaches that also aim to reduce computation for input uncertainty quantification (Section 6). Finally, we show some numerics to support the strengths of our method (Section 7).

Throughout the paper, "$\Rightarrow$" denotes convergence in distribution, "$\stackrel{d}{=}$" denotes equality in distribution, and "$\stackrel{d}{\approx}$" stands (heuristically) for "approximately equal in distribution". For two positive sequences a_k and b_k, we say $a_k = O(b_k)$ if $a_k/b_k \le C$ for all k for some constant $C > 0$, $a_k = o(b_k)$ if $a_k/b_k \to 0$ as $k \to \infty$, $a_k = \omega(b_k)$ if $a_k/b_k \to \infty$ as $k \to \infty$, and $a_k = \Theta(b_k)$ if $\underline{C} \le a_k/b_k \le \overline{C}$ for all k for some constants $\underline{C}, \overline{C} > 0$.

2 THE INPUT UNCERTAINTY PROBLEM

Consider the estimation of a quantity, say $\psi(\mathbf{P}) \in \mathbb{R}$, where $\mathbf{P} = (P_1, \ldots, P_m)$ is a list of independent input probability distributions and ψ is a function that can be simulated. That is, given a list of distributions $\mathbf{Q}$, we have the capability to generate unbiased samples of $\psi(\mathbf{Q})$, say $\hat{\psi}_r(\mathbf{Q})$. As a basic example, think of $\mathbf{P} = (P_1, P_2)$ as the interarrival time and service time distributions of a queueing system, and $\psi(\mathbf{P})$ is the expected average waiting time over a certain number of customers.

Suppose the input distributions $\mathbf{P}$ in the simulation model are unknown, but input data are available. More specifically, suppose we have i.i.d. data $X_{j1}, \ldots, X_{jn_j}$ for input model P_j, and n_j is its sample size. In the case

where we do not assume or utilize parametric information, then, to obtain a point estimate of $\psi(\mathbf{P})$, a natural approach is to construct the empirical distributions $\hat{\mathbf{P}} = (\hat{P}_1, \ldots, \hat{P}_m)$, where $\hat{P}_j(\cdot) = (1/n_j)\sum_{i=1}^{n_j} I(X_{ji} \in \cdot)$ with $I(\cdot)$ denoting the indicator function. Then, we use these empirical input distributions to run a number of simulation runs, say R runs $\hat{\psi}_r(\hat{\mathbf{P}}), r = 1, \ldots, R$, and take average to obtain a point estimate $\hat{\psi}(\hat{\mathbf{P}})$.

The goal is to construct a CI for $\psi(\mathbf{P})$. Alternately, we can also cast our goal as to estimate the output variance, and more specifically the noise contribution coming from each input data source and the Monte Carlo in the variance decomposition. These two goals are closely related, as the variance estimate can be used directly to construct a CI. To this end, to facilitate the formulation of a limit theorem, we introduce a scale parameter n, and impose the scaling that $n_j/n \to q_j$ and $R/n \to q$. Then, under mild conditions (Glynn and Lam 2018), we have the central limit theorem (CLT)

$$\sqrt{n}(\hat{\psi}(\hat{\mathbf{P}}) - \psi(\mathbf{P})) \Rightarrow N\left(0, \sum_{j=1}^{m} \frac{\sigma_j^2}{q_j} + \frac{\tau^2}{q}\right) \tag{1}$$

as $n \to \infty$. In other words, we have

$$\hat{\psi}(\hat{\mathbf{P}}) - \psi(\mathbf{P}) \overset{d}{\approx} N\left(0, \sum_{j=1}^{m} \frac{\sigma_j^2}{n_j} + \frac{\tau^2}{R}\right). \tag{2}$$

Here $\sum_{j=1}^{m} \sigma_j^2/n_j$ represents the variance contributed from the input data noise, or what we would simply call the input variance, and each σ_j^2/n_j is the contribution from the noise in estimating input model j. The term τ^2/R is the variance from the Monte Carlo noise. Note that our point estimate has two "hats", one above ψ and one above $\mathbf{P}$, which signifies that it contains two sources of noises and hence the CLT entails two components in the variance.

The constants in the variance can be written in further mathematical detail. The constant in the Monte Carlo noise $\tau^2 = Var(\hat{\psi}_r(\mathbf{P}))$ which is the variance of one simulation run under the true input distributions $\mathbf{P}$. The constants in the input variance $\sigma_j^2 = Var_{P_j}(IF_j(X_j; \mathbf{P}))$, where X_j is a random variable governed by P_j and Var_{P_j} denotes the variance under P_j, $IF(\cdot; \mathbf{P})$ is the so-called influence function which can be interpreted as a Gateaux derivative of $\psi(\mathbf{P})$ with respect to the distribution P_j, i.e.,

$$\lim_{\varepsilon \to 0_+} \frac{\psi((1-\varepsilon)\mathbf{P} + \varepsilon(0, \ldots, 0, \delta_x, 0, \ldots, 0)) - \psi(\mathbf{P})}{\varepsilon} = IF_j(x; \mathbf{P}) \tag{3}$$

where δ_x is the delta mass at x. The influence function $IF_j(\cdot; \mathbf{P})$ has a mean-zero property that $E_{P_j}[IF_j(X_j; \mathbf{P})] = 0$ where E_{P_j} denotes the expectation under P_j that governs X_j.

In terms of variance estimation, we can decompose

$$Var(\hat{\psi}(\hat{\mathbf{P}})) = Var(\psi(\hat{\mathbf{P}})) + E[Var(\hat{\psi}(\hat{\mathbf{P}})|\hat{\mathbf{P}})] = Var(\psi(\hat{\mathbf{P}})) + \frac{E[Var(\hat{\psi}_r(\hat{\mathbf{P}})|\hat{\mathbf{P}})]}{R} \tag{4}$$

where

$$Var(\psi(\hat{\mathbf{P}})) \approx \sum_{j=1}^{m} \frac{\sigma_j^2}{n_j} \tag{5}$$

and

$$E[Var(\hat{\psi}_r(\hat{\mathbf{P}})|\hat{\mathbf{P}})] \approx \tau^2.$$

Thanks to the CLT (1) or (2), we can construct a $(1-\alpha)$-level normality CI

$$\left[\hat{\psi}(\hat{\mathbf{P}}) - z_{1-\alpha/2}\sqrt{\sum_{j=1}^{m} \frac{\sigma_j^2}{n_j} + \frac{\tau^2}{R}}, \hat{\psi}(\hat{\mathbf{P}}) + z_{1-\alpha/2}\sqrt{\sum_{j=1}^{m} \frac{\sigma_j^2}{n_j} + \frac{\tau^2}{R}}\right] \tag{6}$$

where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$-th quantile of the standard normal distribution. The question then becomes how to estimate σ_j^2 and τ^2, or any other approaches to construct CI without directly using the CLT.

3 EXISTING APPROACHES AND COMPUTATIONAL CHALLENGES

We focus on the CI construction problem. In the literature, the two major approaches are bootstrap resampling and the direct use of the delta method. We present some details and explain their computational challenges.

3.1 Quantile-Based Bootstrap

The idea of the bootstrap, as described in the introduction, is to use the resample distribution to approximate the sampling distribution of the original estimate. For concreteness, suppose we are interested in estimating $\psi(P)$ for an input distribution P, and given i.i.d. data $\mathscr{X} = \{X_1, \ldots, X_n\}$, we use $\psi(\hat{P})$ as a point estimate (suppose for now that $\psi(\cdot)$ can be evaluated exactly and there is only one input distribution). Then the bootstrap principle stipulates the following. For a resample, i.e., sampling with replacement from $\mathscr{X}$ n times to get $X_1^*, \ldots, X_n^*$, we construct the resample empirical distribution $P^*(\cdot) = (1/n)\sum_{i=1}^n I(X_i^* \in \cdot)$. Then the distribution of $\psi(P^*) - \psi(\hat{P})$, conditional on $\hat{P}$, is approximately the same as that of $\psi(\hat{P}) - \psi(P)$. Under regularity conditions, this approximation can be mathematically formulated as the CLT

$$\sqrt{n}(\psi(\hat{P}) - \psi(P)) \Rightarrow N(0, \sigma^2) \tag{7}$$

which is guaranteed by the delta method, and

$$\sqrt{n}(\psi(P^*) - \psi(\hat{P})) \Rightarrow N(0, \sigma^2) \tag{8}$$

conditional on $X_1, X_2, \ldots$, where $N(0, \sigma^2)$ is the same limiting variable as the CLT for the original estimate. While $\psi(\hat{P}) - \psi(P)$ is generally unobtainable, we can resample many times, say B times, to obtain $P^{*b}, b = 1, \ldots, B$, and get an accurate approximation of the distribution of $\psi(P^*) - \psi(\hat{P})$.

To construct CI, note that if we could find the $\alpha/2$-th and $(1-\alpha/2)$-th quantiles of $\psi(\hat{P}) - \psi(P)$, say $q_{\alpha/2}$ and $q_{1-\alpha/2}$, then $[\psi(\hat{P}) - q_{1-\alpha/2}, \psi(\hat{P}) - q_{\alpha/2}]$ would give a $(1-\alpha)$ CI. The basic bootstrap uses the $\alpha/2$-th and $(1-\alpha/2)$-th quantiles of the simulated $\psi(P^*) - \psi(\hat{P})$, say $q^*_{\alpha/2}$ and $q^*_{1-\alpha/2}$, to approximate $q_{\alpha/2}$ and $q_{1-\alpha/2}$, giving an implementable interval $[\psi(\hat{P}) - q^*_{1-\alpha/2}, \psi(\hat{P}) - q^*_{\alpha/2}]$. The percentile bootstrap simply uses $[q^*_{\alpha/2}, q^*_{1-\alpha/2}]$, by exploiting the symmetry of the limiting distribution in the CLT. In the input uncertainty setting, note that $\psi(\cdot)$ cannot be exactly evaluated but requires noisy simulation, and thus to approximate $q^*_{\alpha/2}$ and $q^*_{1-\alpha/2}$ we need to generate $\hat{\psi}(P^*) - \hat{\psi}(\hat{P})$, where $\hat{\psi}(\cdot)$ is obtained by simulating many, say R, replications and taking average, which thus amounts to a total of BR simulation runs.

To dive further into the above issue, note that as long as the resamples correctly approximate the quantiles of $\hat{\psi}(\hat{P}) - \psi(P)$, where $\hat{\psi}(\hat{P})$ is the original point estimate obtained from averaging many simulation runs driven by $\hat{P}$, then the bootstrap method is valid. Unfortunately, the resample quantity $\hat{\psi}(P^*) - \hat{\psi}(\hat{P})$ has two "hats" above the ψ on both P^* and $\hat{P}$, or in other words the "centering" of $\hat{\psi}(P^*)$, which is $\hat{\psi}(\hat{P})$, contains simulation noise. Because of this there is a systematic mismatch of the quantiles of the resample and the quantiles of the original estimate. To address this, one can use an $\omega(n)$ amount of simulation runs, i.e., larger order than the sample size, to obtain the original estimate $\hat{\psi}(\hat{P})$ which is also used as the "centering" of the resample estimate. If we further use an $\omega(n)$ amount of simulation runs to obtain each resample estimate $\hat{\psi}(\hat{P}^{*b})$, then essentially the simulation noises become negligible and the basic or percentile bootstrap becomes statistically valid. However, computationally this is very heavy. An alternative is to use only $O(n)$ amount of simulation runs for each resample estimate, which gives rise to a larger variability of $\hat{\psi}(P^*) - \hat{\psi}(\hat{P})$ than $\hat{\psi}(\hat{P}) - \psi(P)$, and properly "shrink" the estimates or quantiles back to the latter level (Song et al. 2022). Nonetheless, the overall simulation efforts in all these schemes are $\omega(n)$, i.e., the total simulation effort is of larger order than the sample size.

3.2 Variance-Based Bootstrap

Besides quantile-based bootstraps described above, we can also use the bootstrap to estimate the variances inside the normality CI (6) to construct a CI. By (5), we need to estimate $Var(\psi(\hat{P}))$ (keeping one input distribution here for simplicity) and $E[Var(\hat{\psi}_r(\hat{P})|\hat{P})]$. The latter can be readily estimated by using the sample variance of all simulation runs and is relatively easy to handle. For the input variance $Var(\psi(\hat{P}))$, we again invoke the bootstrap principle to use $Var_*(\psi(P^*))$ as an approximation of $Var(\psi(\hat{P}))$, where Var_* denotes the variance of the resample, conditional on the original data. In the input uncertainty context, $\psi(\cdot)$ can only be noisily evaluated and thus we need a nested simulation that first resample to get $P^{*b}, b = 1,\ldots,B$, and for each P^{*b} we run R simulation runs to obtain $\hat{\psi}(P^{*b})$. Then an estimate of $Var(\psi(\hat{P}))$ can be obtained by an analysis-of-variance (ANOVA) estimator

$$\frac{1}{B-1}\sum_{b=1}^{B}(\hat{\psi}(P^{*b})-\overline{\hat{\psi}})^2-\frac{1}{BR(R-1)}\sum_{b=1}^{B}\sum_{r=1}^{R}(\hat{\psi}_r(P^{*b})-\hat{\psi}(P^{*b}))^2 \tag{9}$$

where $\overline{\hat{\psi}} = (1/B)\sum_{b=1}^{B}\hat{\psi}(P^{*b})$. The second term in (9) is a bias-reducing adjustment, i.e., the second term in the last expression of (4). Estimator (9) can be viewed as estimating the between-group variance in a random effect model.

It turns out that the variance of the natural estimator (9) for $Var_*(\psi(P^*))$ is of order

$$O\left(\frac{1}{Bn^2}+\frac{1}{BR^2}\right) \tag{10}$$

(note that variance is the only source of error for the estimation of $Var_*(\psi(P^*))$ here since this estimator is unbiased). Since the input variance $Var(\psi(\hat{P}))$ is of order $1/n$, its estimator needs to have an error $o(1/n)$ in order to be accurate enough for CI construction purpose. To achieve this using (9), we argue that the total simulation effort BR must be necessarily $\omega(n)$, i.e., of larger order than the data size. First, we must choose $B = \omega(1)$ so that the first term in (10) is $o(1/n^2)$. Second, suppose we choose $BR = O(n)$, then R must be $o(n)$ and the second term in (10) is $\omega(1/n^2)$ which makes the overall error of the estimator of larger order than $1/n^2$, leading to a contradiction. Lam and Qian (2022) calls this $\omega(n)$ overall computational need a "complexity barrier" in using the natural ANOVA estimator for running the variance bootstrap.

3.3 Delta Method

An alternative approach to construct CI is to use the delta method, which gives rise to (6) where the constants in the standard error, namely σ_i^2 and τ^2, are directly estimated by using their mathematical forms. As discussed before, while τ^2 can be readily estimated, σ_i^2's are more difficult as they involve the influence function, which is a functional derivative that, in general, can only be observed noisily. In fact, when the simulation model is black-box, i.e., only unbiased output is available but not gradient information, then estimating the influence function requires finite-difference or zeroth-order derivative estimation, which is well-known to possess slow convergence rate. Moreover, the influence function is the derivative with respect to the input distribution, which is infinite-dimensional in nature and adds to the sophistication.

To make the discussion above more precise, suppose we only have one input distribution, which gives an input variance σ^2. This variance can be estimated directly using

$$\frac{1}{n}\sum_{i=1}^{n}IF(X_i;\hat{P})^2$$

where X_i's are data points for P, and we have used the fact that $E_{\hat{P}}[IF(X_i;\hat{P})] = 0$. By using (3), we can estimate $IF(X_i;\hat{P})$ via

$$\frac{\hat{\psi}((1-\varepsilon)\mathbf{P}+\varepsilon(0,\ldots,0,\delta_{X_i},0,\ldots,0))-\hat{\psi}(\mathbf{P})}{\varepsilon}$$

for some ε that scales with n. We can further reduce the estimation bias via an appropriate U-statistic construction. Nonetheless, it turns out that to ensure the estimated σ^2 is relatively consistent, i.e., the estimation error of σ^2 is $o(1/n)$, the minimum simulation budget is $\Theta(n)$ by choosing $\omega(1/n^{1/4}) \le \varepsilon \le o(1)$ (Lam et al. 2022). This is better than the quantile-based and variance bootstraps discussed before, but it requires some careful procedural tuning to ensure the required simulation load is on par with the data size.

4 CHEAP BOOTSTRAP

The cheap bootstrap method is a resampling-based approach to construct CI that requires as few as *one* resample. It works as follows. For simplicity, let us first consider the case where ψ can be evaluated exactly and there is only one input distribution. Instead of using the standard bootstrap principle which stipulates the approximation of the sampling distribution with the resample distribution, one uses the asymptotic independence among the original estimate and all resample estimates. More concretely, suppose, given the data, the original estimate is $\psi(\hat{P})$ and we resample to obtain B resample estimates $\psi(P^{*b}), b = 1, \ldots, B$. We have

$$(\sqrt{n}(\psi(\hat{P}) - \psi(P)), \sqrt{n}(\psi(P^{*1}) - \psi(\hat{P})), \ldots, \sqrt{n}(\psi(P^{*B}) - \psi(\hat{P}))) \Rightarrow (\sigma Z_0, \sigma Z_1, \ldots, \sigma Z_B) \tag{11}$$

where $Z_i \overset{i.i.d.}{\sim} N(0,1)$. This convergence can be shown to be an implication of (7) and (8) (Lam 2022). Intuitively, since the resample CLT holds universally conditional on any realization of the data sequence, the resample and the original estimates should be independent asymptotically, which gives exactly the statement (11).

With (11), the cheap bootstrap method uses the CI

$$[\psi(\hat{P}) - t_{B,1-\alpha/2} S, \psi(\hat{P}) + t_{B,1-\alpha/2} S] \tag{12}$$

where

$$S^2 = \frac{1}{B} \sum_{b=1}^{B} (\psi(P^{*b}) - \psi(\hat{P}))^2$$

and $t_{B,1-\alpha/2}$ is the $(1-\alpha/2)$-th quantile of t_B, the t-distribution with degree of freedom B. Note that S^2 resembles the sample variance of the resamples $\psi(P^{*b})$, but it does not use $B-1$ in the denominator as in "textbook" sample variance and the centering in the squares is $\psi(\hat{P})$ instead of the sample mean of $\psi(P^{*b})$'s. The critical value is now calibrated by a t-distribution with degree of freedom equal to the resampling effort B. Importantly, (12) is defined and, as we see momentarily, works when B is as small as 1. In particular, when $B = 1$, (12) reduces simply to

$$[\psi(\hat{P}) - t_{1,1-\alpha/2} |\psi(P^*) - \psi(\hat{P})|, \psi(\hat{P}) + t_{B,1-\alpha/2} |\psi(P^*) - \psi(\hat{P})|]$$

where $\psi(P^*)$ is now the single resample estimate. The following appears in Lam (2022):

Theorem 1 Suppose (7) and (8) hold. Then, for any fixed $B \ge 1$, (11) holds and consequently (12) is an asymptotically exact $(1-\alpha)$ CI for $\psi(P)$, i.e.,

$$P(\psi(P) \in [\psi(\hat{P}) - t_{B,1-\alpha/2} S, \psi(\hat{P}) + t_{B,1-\alpha/2} S]) \to 1 - \alpha$$

as $n \to \infty$.

This theorem can be justified by noting that

$$\frac{\psi(\hat{P}) - \psi(P)}{S} \tag{13}$$

is a pivotal statistic where the nuisance parameter, in this case the unknown standard deviation σ, can be canceled out so that the statistic converges to t_B by using elementary properties of normal variables. More precisely, we have

$$\frac{\psi(\hat{P})-\psi(P)}{S}=\frac{\sqrt{n}(\psi(\hat{P})-\psi(P))}{\sqrt{(1/B)\sum_{b=1}^{B}(\sqrt{n}(\psi(P^{*b})-\psi(\hat{P})))^2}}\Rightarrow\frac{\sigma Z_0}{\sqrt{(1/B)\sum_{b=1}^{B}(\sigma Z_b)^2}}=\frac{Z_0}{\sqrt{(1/B)\sum_{b=1}^{B}Z_b^2}}\stackrel{d}{=}t_B.$$

It is worth pointing out that a good CI not only requires a valid coverage, but ought to be short and less variable. Indeed, when B is very small, e.g., $B=1$, then the interval width is large, signaling a lack of extracted statistical information. However, the interval width sharply decreases as B moves away from 1, towards a level that is close to the interval width of the normality CI, which can be viewed as when using $B=\infty$. These observations can be readily justified since the cheap bootstrap uses and follows the behavior of a t-interval, asymptotically as $n\to\infty$. For further details, see Lam (2022).

5 A CHEAP PROCEDURE FOR INPUT UNCERTAINTY QUANTIFICATION

The discussion in Section 4 assumes $\psi(\cdot)$ can be exactly evaluated. When $\psi(\cdot)$ can only be observed noisily, the limit theorem (11) becomes more complex and the construction of pivotal statistic that cancels out the nuisance parameters becomes more involved. To be more precise, suppose we consider the general situation now where we have m input distributions $\mathbf{P}=(P_1,\ldots,P_m)$, each with input data size n_j. We use R simulation runs to estimate each $\hat{\psi}(\hat{\mathbf{P}})$ and $\hat{\psi}(\mathbf{P}^{*b})$. Here, the resample empirical distributions $\mathbf{P}^{*b}$ are obtained by resampling, for each input model j, n_j times via sampling with replacement to obtain $X_{ji}^*, i=1,\ldots,n_j$, and the resamplings for different input models are conducted independently. We use the scale parameter n, with $n_i=nq_i$ and $R=nq$. Then we have

$$\begin{aligned}&(\sqrt{n}(\hat{\psi}(\hat{\mathbf{P}})-\psi(P)),\sqrt{n}(\hat{\psi}(\mathbf{P}^{*1})-\hat{\psi}(\hat{\mathbf{P}})),\ldots,\sqrt{n}(\hat{\psi}(\mathbf{P}^{*B})-\hat{\psi}(\hat{\mathbf{P}})))\\ \Rightarrow\quad&\left(\sqrt{\sum_{j=1}^{m}\frac{\sigma_i^2}{q_i}}Z_0+\frac{\tau}{\sqrt{q}}W_0,\sqrt{\sum_{j=1}^{m}\frac{\sigma_i^2}{q_i}}Z_1+\frac{\tau}{\sqrt{q}}W_1-\frac{\tau}{\sqrt{q}}W_0,\ldots,\sqrt{\sum_{j=1}^{m}\frac{\sigma_i^2}{q_i}}Z_B+\frac{\tau}{\sqrt{q}}W_B-\frac{\tau}{\sqrt{q}}W_0\right)\end{aligned}\tag{14}$$

where $Z_i,W_i,i=0,\ldots,B$ are all i.i.d. $N(0,1)$. The limit in the convergence (14) arises from the combinations of input noises represented by Z_i's and simulation noises represented by W_i's. The convergence (14) can be guaranteed by using a similar argument as that for (11), under additional moment assumptions on the simulation noise.

Thus, in contrast to the asymptotic independence in (11), in the input uncertainty context we have an asymptotic joint distribution structure (14) that is more complex. To proceed, we study two approaches, both of which can be viewed as a multi-input generalization to Lam (2022) that considers only the case of one input model.

The first approach, which we call "centered at original estimate" is to consider the statistic $(\hat{\psi}(\hat{\mathbf{P}})-\psi(\mathbf{P}))/S_O$ where $S_O^2=(1/B)\sum_{b=1}^{B}(\hat{\psi}(\mathbf{P}^{*b})-\hat{\psi}(\hat{\mathbf{P}}))^2$. This statistic is the natural analog of (13). By a similar argument as before, we have

$$\frac{\hat{\psi}(\hat{\mathbf{P}})-\psi(\mathbf{P})}{S_O}\Rightarrow\frac{\sqrt{\sum_{j=1}^{m}\frac{\sigma_i^2}{q_i}}Z_0+\frac{\tau}{\sqrt{q}}W_0}{\sqrt{\frac{1}{B}\sum_{b=1}^{B}\left(\sqrt{\sum_{j=1}^{m}\frac{\sigma_i^2}{q_i}}Z_b+\frac{\tau}{\sqrt{q}}W_b-\frac{\tau}{\sqrt{q}}W_0\right)^2}}$$

which can be written as

$$\frac{\theta Z_0+W_0}{\sqrt{\frac{1}{B}\sum_{b=1}^{B}(\theta Z_b+W_b-W_0)^2}}\tag{15}$$

where $\theta = \sqrt{\sum_{j=1}^m \frac{\sigma_i^2}{n_i}}/(\tau/\sqrt{R})$ is the ratio of the input noise over the simulation noise for the original estimate and each resample estimate, measured by the standard deviation. This quantity is unknown, but we can take a worst-case approach. Define $F(q;\theta)$ as the distribution function of (15). Consider the critical value

$$q_{O,1-\alpha/2} = \min\left\{q : \min_{\theta\geq 0} F(q;\theta) \geq 1-\frac{\alpha}{2}\right\}.$$

Then we construct a $(1-\alpha)$ CI as $[\hat{\psi}(\hat{\mathbf{P}}) - q_{O,1-\alpha/2}S_O, \hat{\psi}(\hat{\mathbf{P}}) + q_{O,1-\alpha/2}S_O]$. Note that even though $q_{O,1-\alpha/2}$ does not have a closed form, it can be readily computed by running many Monte Carlo replications for normal variables – not the simulation model $\psi(\cdot)$ itself. We have

Theorem 2 Suppose (14) holds. Then, for any fixed $B \geq 1$,

$$\liminf_{n\to\infty} P(\psi(\mathbf{P}) \in [\hat{\psi}(\hat{\mathbf{P}}) - q_{O,1-\alpha/2}S_O, \hat{\psi}(\hat{\mathbf{P}}) + q_{O,1-\alpha/2}S_O]) \geq 1-\alpha.$$

Unlike Theorem 1, Theorem 2 guarantees only an asymptotically correct coverage, i.e., at least $1-\alpha$, instead of asymptotically exact coverage. This is due to the conservativeness in determining the critical value $q_{O,1-\alpha/}$ because of the unknown nuisance parameter θ that cannot be straightforwardedly canceled out. Nonetheless, we see that this conservativeness appears negligible in our numerical experiments in Section 7.

The second approach, which we call "centered at resample mean", uses the statistic $(\hat{\psi}(\hat{\mathbf{P}}) - \psi(\mathbf{P}))/S_M$ where $S_M^2 = (1/(B-1))\sum_{b=1}^B (\hat{\psi}(\mathbf{P}^{*b}) - \overline{\hat{\psi}})^2$ and $\overline{\hat{\psi}} = (1/B)\sum_{b=1}^B \hat{\psi}(\mathbf{P}^{*b})$ is the sample mean of the resample estimates. With this statistic, now we have

$$\frac{\hat{\psi}(\hat{\mathbf{P}}) - \psi(\mathbf{P})}{S_M} \Rightarrow \frac{\sqrt{\sum_{j=1}^m \frac{\sigma_i^2}{q_i}} Z_0 + \frac{\tau}{\sqrt{q}} W_0}{\sqrt{\frac{1}{B-1}\sum_{b=1}^B \left(\sqrt{\sum_{j=1}^m \frac{\sigma_i^2}{q_i}} Z_b + \frac{\tau}{\sqrt{q}} W_b - \left(\sqrt{\sum_{j=1}^m \frac{\sigma_i^2}{q_i}} \bar{Z} + \frac{\tau}{\sqrt{q}} \bar{W}\right)\right)^2}} \stackrel{d}{=} t_{B-1}$$

by canceling out the nuisance parameter given by the overall standard deviation $\sqrt{\sum_{j=1}^m \frac{\sigma_i^2}{q_i} + \frac{\tau^2}{q}}$. Thus, we can use the $(1-\alpha)$ CI $[\hat{\psi}(\hat{\mathbf{P}}) - t_{B-1,1-\alpha/2}S_M, \hat{\psi}(\hat{\mathbf{P}}) + t_{B-1,1-\alpha/2}S_M]$ with the guarantee:

Theorem 3 Suppose (14) holds. Then, for any fixed $B \geq 2$,

$$P(\psi(\mathbf{P}) \in [\hat{\psi}(\hat{\mathbf{P}}) - t_{B-1,1-\alpha/2}S_M, \hat{\psi}(\hat{\mathbf{P}}) + t_{B-1,1-\alpha/2}S_M]) \to 1-\alpha$$

as $n \to \infty$.

Unlike the "centered at original estimate" approach, here in Theorem 3 the asymptotic coverage is exact. However, we now need $B \geq 2$ instead of $B \geq 1$ in Theorem 2. Even though this difference may sound immaterial since in practice using $B = 1$ or $B = 2$ would give rise to very wide intervals and is undesirable anyway (even if the coverage is correct), the difference in the degree of freedom B versus $B-1$ does show up when one uses a moderately small B, e.g., $B = 5$, which could be a reasonable choice in practice given the computational demand in addressing input uncertainty. In our experiments, we find that both approaches are comparable in terms of coverage accuracy and interval width.

6 COMPARISONS WITH OTHER RECENT REMEDIES

We discuss and contrast the cheap bootstrap with several recent approaches that aim to reduce computation when addressing input uncertainty. The first is a subsampling approach (Lam and Qian 2022) that leverages the mathematical form of the input-contributed variance which is reciprocal in the input data size. Thanks

to this form, by using subsample, i.e., resample of size smaller than the original data size, in the outer layer, to obtain an estimate of the input variance and rescale, it turns out to require less overall budget than standard variance bootstrap. The second is to use subsampled infinitesimal jackknife (Lam et al. 2022), which applies the delta method directly. It first subsamples a subset of input data from the original data set before carrying out the influence function estimation, which makes the estimation error scale with the subsample size instead of the full sample size, and turns out to reduce overall simulation budget. While both subsampling and subsampled infinitesimal jackknife can achieve similar overall simulation effort as our cheap bootstrap approach, they both require choosing the subsample size as a tuning parameter, while the cheap bootstrap does not utilize subsampling in the procedure.

Another approach that reduces computation in input uncertainty, and arguably the closest to ours, is the batching or sectioning approach (Glynn and Lam 2018). This approach first batches the input data into several batches, and for each batch, uses the input data to construct input distributions which then are used to drive adequate simulation runs. A valid CI is then constructed by aggregating the averages of simulation outputs in the batches via a t-statistic. Note that batching generally faces a tradeoff between the number of batches and the sample size per batch, i.e., for a given total sample size, if the number of batches increases then the sample size per batch would decrease and vice versa, while ideally we would like both the number of batches and the sample size per batch to be large – the former reduces the interval width, while the latter ensures the accuracy of the normal approximation. In contrast, cheap bootstrap is free from this tradeoff, as the number of resamples can grow unboundedly without affecting the sample size in each resample. Because of this, it is expected that the cheap bootstrap would outperform batching unless the input data size is very big. Finally, note that we could also use batched jackknife that alleviates the aforementioned tradeoff faced by batching, but this requires more sophisticated procedural parameter tuning as one would need to carefully cancel out the nuisance parameter, namely the overall variance which consists of both the input data and the Monte Carlo contributions.

7 NUMERICAL EXPERIMENTS

We present two numerical examples and compare our methods to some existing approaches. The first example is an $M/M/1$ queue where our goal is to estimate the expected average waiting time of the first 10 customers. The ground-truth interarrival time and service time distributions are exponential with rates 1 and 1.1 respectively. The queue starts from the empty state and the first customer immediately arrives. We assume we collect i.i.d. data for the interarrival and service times, which we use to construct empirical distributions to drive our simulation. Similar examples have been used in, e.g., Barton et al. (2014), Song and Nelson (2015), Zouaoui and Wilson (2004). Our second example is the computer communication network considered in Cheng and Holloand (1997), Lin et al. (2015), Lam and Qian (2022). We follow the specifications in Section 5 of Lam and Qian (2022) and skip the details here due to space limit. We are interested in estimating the average delay of the first 30 messages that arrive to the system. We consider the situation that all arrival rates and distribution of message lengths are unknown and have an i.i.d. data source for each of them, resulting in the need to calibrate 13 input models in total.

For each example, we vary our number of resamples B from 1 or 2 to 20 and apply our "centered at original estimate" and "centered at resample mean" approaches. We use $R = 50$, thus the overall simulation effort is $50B$. We compare our approaches to 1) the basic bootstrap, where we again use $R = 50$ to evaluate each $\hat{\psi}(\cdot)$; 2) the percentile bootstrap with the same specification of R; 3) "centered at original estimate" but using $t_{B,1-\alpha/2}$ as the critical value instead of $q_{O,1-\alpha/2}$; 4) sectioning input data. For each setting, we repeat our experiments 1000 times to calculate the empirical coverage and average interval width, and we run 1 million simulation runs to approximate the ground truths. For the $M/M/1$ example, we use input data sizes $n_1 = 100$ and $n_2 = 100$. For the computer network example, we use $n_i, i = 1,\ldots,13$ that follows Lam and Qian (2022), with n_i's ranging from 30 to 60. Our goal is to construct a 95% CI.

Tables 1 and 2 show the empirical coverages and average interval widths obtained by all the methods, as we vary the number of resamples B from 1 to 20 for the bootstrap approaches, and the number of batches

Table 1: Comparisons of interval performances for the queueing example.

B	Cheap bootstrap centered at original estimate		Cheap bootstrap centered at resample mean		Basic bootstrap		Percentile bootstrap		Centered at original estimate with t		Sectioning	
	Coverage	Width	Coverage	Width	Coverage	Width	Coverage	Width	Coverage	Width	Coverage	Width
1	0.95	7.53	NA	NA	NA	NA	NA	NA	0.95	7.50	NA	NA
2	0.96	2.77	0.94	6.46	0.31	0.34	0.27	0.34	0.96	2.76	0.95	6.33
3	0.96	2.14	0.94	2.48	0.49	0.52	0.40	0.52	0.96	2.14	0.94	2.38
4	0.96	1.90	0.93	1.91	0.60	0.63	0.47	0.63	0.96	1.90	0.95	1.81
5	0.96	1.78	0.94	1.71	0.68	0.71	0.53	0.71	0.96	1.78	0.94	1.60
6	0.96	1.72	0.94	1.61	0.72	0.78	0.58	0.78	0.96	1.72	0.94	1.48
7	0.96	1.69	0.94	1.57	0.76	0.84	0.61	0.84	0.96	1.69	0.93	1.40
8	0.95	1.66	0.93	1.52	0.78	0.88	0.63	0.88	0.95	1.65	0.94	1.38
9	0.95	1.64	0.93	1.50	0.80	0.91	0.65	0.91	0.95	1.64	0.94	1.36
10	0.95	1.61	0.93	1.47	0.82	0.94	0.67	0.94	0.95	1.61	0.94	1.33
11	0.95	1.60	0.93	1.45	0.84	0.97	0.68	0.97	0.95	1.60	0.94	1.32
12	0.95	1.59	0.93	1.44	0.85	0.99	0.70	0.99	0.95	1.58	0.94	1.32
13	0.96	1.58	0.94	1.43	0.86	1.01	0.71	1.01	0.96	1.58	0.94	1.30
14	0.96	1.57	0.93	1.42	0.87	1.02	0.72	1.02	0.95	1.57	0.93	1.29
15	0.95	1.56	0.93	1.41	0.88	1.04	0.72	1.04	0.95	1.55	0.95	1.29
16	0.95	1.55	0.94	1.40	0.88	1.05	0.73	1.05	0.95	1.55	0.95	1.29
17	0.95	1.54	0.94	1.39	0.89	1.06	0.73	1.06	0.95	1.54	0.94	1.28
18	0.95	1.54	0.93	1.38	0.89	1.07	0.73	1.07	0.95	1.53	0.94	1.28
19	0.96	1.53	0.94	1.37	0.89	1.08	0.73	1.08	0.96	1.52	0.93	1.28
20	0.95	1.52	0.94	1.37	0.90	1.09	0.74	1.09	0.95	1.52	0.94	1.28

in the sectioning approach (where B there represents the number of batches). Note that some methods do not give any outputs when B is 1. These include the cheap bootstrap centered at resample mean which requires $B \geq 2$, quantile-based bootstraps including basic and percentile bootstraps which cannot output two distinct finite numbers for two quantiles when there is only one resample, and sectioning which requires the number of batches to be at least 2.

As we see in the tables, the cheap bootstrap methods consistently give rise to accurate coverages close to 95%, starting from $B = 1$ when centered at original estimate and $B = 2$ when centered at resample mean. The coverages of the former are all $95\% - 97\%$, while the latter are all $93\% - 96\%$. In contrast, quantile-based bootstraps are severely under-covered for small B, and even when $B = 20$, their coverages still bear a significant gap with the nominal 95%. For example, when $B = 2$, the basic bootstrap gives 31% for the first example and 34% for the second example, and the percentile bootstrap gives 27% for the first example and 32% for the second example. When $B = 20$, the basic bootstrap gives 90% for both examples, and the percentile bootstrap gives 74% for the first example and 85% for the second example. These obervations are unsurprising since the quantile-based bootstraps are not designed to work for very small B, but require a sufficiently large B for the resample distribution to well approximate the sampling distribution.

The coverages of centered at original estimate but using a t-quantile to calibrate the critical value are similar to the use of $q_{O,1-\alpha}$, suggesting that the conservativeness brought by the worst-case analysis in determining $q_{O,1-\alpha/2}$ is quite light. Finally, sectioning appears to perform well for the first example, with all coverages ranging in $93\% - 95\%$, but deteriorate as B increases beyond 10 in the second example, dropping to only 38% when $B = 20$. This is because when B increases, the data size per batch decreases, and in the second example when B goes beyond 10, the data size per batch could become too small for normal approximation to work well.

Regarding interval width, cheap bootstrap CI are initially wide. For instance, the average widths are 7.53 and 11.07 for the first and second examples respectively when $B = 1$ for centered at original estimate, and 6.46 and 11.12 respectively when $B = 2$ for centered at resample mean. However, the widths drop sharply as B increases and gradually level off. For instance, the widths decrease from 7.53 to 2.77 and from 11.07 to 4.22 for the first and second examples respectively when B increases from 1 to 2 for centered at original estimate, and from 6.46 to 2.48 and from 11.12 to 4.07 respectively when B increases from

Table 2: Comparisons of interval performances for the computer communication network example.

B	Cheap bootstrap centered at original estimate		Cheap bootstrap centered at resample mean		Basic bootstrap		Percentile bootstrap		Centered at original estimate with t		Sectioning	
	Coverage	Width	Coverage	Width	Coverage	Width	Coverage	Width	Coverage	Width	Coverage	Width
1	0.95	11.07	NA	NA	NA	NA	NA	NA	0.95	11.04	NA	NA
2	0.97	4.22	0.96	11.12	0.34	0.59	0.32	0.59	0.97	4.20	0.95	12.84
3	0.96	3.19	0.95	4.07	0.50	0.86	0.48	0.86	0.96	3.18	0.96	4.64
4	0.96	2.88	0.95	3.15	0.61	1.04	0.57	1.04	0.96	2.87	0.96	3.75
5	0.96	2.68	0.95	2.78	0.68	1.16	0.62	1.16	0.96	2.68	0.96	3.44
6	0.96	2.58	0.95	2.61	0.72	1.26	0.66	1.26	0.96	2.58	0.94	3.33
7	0.95	2.50	0.95	2.48	0.75	1.33	0.70	1.33	0.95	2.50	0.96	3.29
8	0.96	2.45	0.95	2.42	0.77	1.40	0.73	1.40	0.96	2.45	0.94	3.41
9	0.96	2.42	0.95	2.36	0.80	1.44	0.74	1.44	0.96	2.41	0.94	3.56
10	0.96	2.38	0.95	2.32	0.81	1.49	0.77	1.49	0.96	2.38	0.94	3.75
11	0.96	2.36	0.95	2.29	0.82	1.53	0.78	1.53	0.96	2.35	0.91	4.17
12	0.96	2.35	0.95	2.27	0.84	1.56	0.79	1.56	0.96	2.34	0.88	4.49
13	0.96	2.32	0.95	2.25	0.85	1.59	0.80	1.59	0.96	2.32	0.86	4.74
14	0.96	2.32	0.95	2.23	0.86	1.62	0.81	1.62	0.96	2.31	0.79	5.18
15	0.96	2.31	0.95	2.22	0.87	1.65	0.82	1.65	0.96	2.30	0.77	5.49
16	0.96	2.28	0.95	2.21	0.88	1.67	0.83	1.67	0.96	2.28	0.73	6.28
17	0.96	2.27	0.95	2.20	0.88	1.68	0.83	1.68	0.96	2.27	0.62	7.20
18	0.96	2.27	0.95	2.19	0.88	1.70	0.84	1.70	0.96	2.26	0.54	7.50
19	0.96	2.27	0.95	2.18	0.89	1.72	0.85	1.72	0.96	2.26	0.47	8.01
20	0.96	2.25	0.95	2.17	0.90	1.73	0.85	1.73	0.96	2.25	0.38	8.47

2 to 3 for centered at resample mean. The decrease continues to be sharp, though less intensely, when B increases from 2 to 3 for centered at original estimate and from 3 to 4 for centered at resample mean, afterwards the decrease becomes more gradual. In contrast, the quantile-based bootstraps appear to be shorter and have a small increase in interval widths as B increases. Note, however, that these short intervals are statistically invalid in terms of coverage. Centered at original estimate but using a t-quantile to calibrate the critical value gives interval widths similar to using $q_{O,1-\alpha/2}$, much like the coverage performances. Finallly, sectioning, like the cheap bootstraps, also generate wide intervals when the number of batches is small. For instance, the average width is 6.33 for the first example and 12.84 for the second example when $B = 2$. However, while the width follows a trend similar to the cheap bootstraps as B increases, which decreases sharply initially and gradually levels off, for the first example, the width first decreases and bounces back up when B is around 11 in the second example. This suggests a degradation of the normality approximation within each batch in the latter, so that the behavior of the interval widths deviate from the asymptotic t-statistic.

ACKNOWLEDGMENTS

I gratefully acknowledge support from the National Science Foundation under grants CAREER CMMI-1834710 and IIS-1849280.

REFERENCES

Barton, R., H. Lam, and E. Song. 2022. "Input Uncertainty in Stochastic Simulation". *The Palgrave Handbook of Operations Research, Chapter 17*.

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

Barton, R. R., B. L. Nelson, and W. Xie. 2014. "Quantifying Input Uncertainty via Simulation Confidence Intervals". *INFORMS Journal on Computing* 26(1):74–87.

Barton, R. R., and L. W. Schruben. 1993. "Uniform and Bootstrap Resampling of Empirical Distributions". In *Proceedings of the 1993 Winter Simulation Conference*, edited by G. W. Evans, M. Mollaghasemi, E. Russell, 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.

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 (TOMACS)* 14(4):344–362.

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

Chick, S. E. 2001. "Input Distribution Selection for Simulation Experiments: Accounting for Input Uncertainty". *Operations Research* 49(5):744–758.

Corlu, C. G., A. Akcay, and W. Xie. 2020. "Stochastic Simulation under Input Uncertainty: A Review". *Operations Research Perspectives* 7:100162.

Davison, A. C., and D. V. Hinkley. 1997. *Bootstrap Methods and Their Application*. Cambridge, UK: Cambridge University Press.

Efron, B. 1992. "Bootstrap Methods: Another Look at the Jackknife". In *Breakthroughs in Statistics*, 569–593. New York, NY: Springer Series in Statistics.

Efron, B., and R. J. Tibshirani. 1994. *An Introduction to the Bootstrap*. Boca Raton, FL: CRC Press.

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. A. Juan, N. Mustafee, A. Skoogh, J. S., and B. Johansson, 1551–1562. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

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, P. J. Sánchez, D. Ferrin, and D. J. Morrice, 90–100. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

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. 2022. "A Cheap Bootstrap Method for Fast Inference". *arXiv preprint arXiv:2202.00090*.

Lam, H., Z. Liu, and H. Qian. 2022. "Efficient Input Uncertainty Quantification without Nested Simulation: A Subsampled Infinitesimal Jackknife Approach". *Preprint*.

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

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

Lin, Y., E. Song, and B. L. Nelson. 2015. "Single-Experiment Input Uncertainty". *Journal of Simulation* 9(3):249–259.

Nelson, B. 2013. *Foundations and Methods of Stochastic Simulation: A First Course*. New York, NY: Springer Science & Business Media.

Song, E., H. Lam, and R. Barton. 2022. "A Shrinkage Approach to Improve Direct Bootstrap Resampling under Input Uncertainty". *Preprint*.

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. 2019. "Input-Output Uncertainty Comparisons for Discrete Optimization via Simulation". *Operations Research* 67(2):562–576.

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. Diallo, I. Ryzhov, L. Yilmaz, S. Buckley, and J. Miller, 162–176. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

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.

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

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

AUTHOR BIOGRAPHIES

HENRY LAM is an Associate Professor of IEOR at Columbia University. His email address is henry.lam@columbia.edu.