

ADVANCED TUTORIAL: INPUT UNCERTAINTY QUANTIFICATION

Eunhye Song
Barry L. Nelson

Dept. of Ind. Engr. & Mgmt. Sci.
Northwestern University
Evanston, IL 60208 USA

C. Dennis Pegden

Simio LLC
504 Beaver Street
Sewickley, PA 15143 USA

ABSTRACT

“Input uncertainty” refers to the (often unmeasured) effect of not knowing the true, correct distributions of the basic stochastic processes that drive the simulation. These include, for instance, interarrival-time and service-time distributions in queueing models; bed-occupancy distributions in health care models; distributions for the values of underlying assets in financial models; and time-to-failure and time-to-repair distributions in reliability models. When the input distributions are obtained by fitting to observed real-world data, then it is possible to quantify the impact of input uncertainty on the output results. In this tutorial we carefully define input uncertainty, describe various proposals for measuring it, contrast input uncertainty with input sensitivity, and provide and illustrate a practical approach for quantifying overall input uncertainty and the relative contribution of each input model to overall input uncertainty.

1 INTRODUCTION

The “stochastic” in *stochastic simulation* is produced by input models that drive the simulation output. In the simplest case the input models are fully specified univariate probability distributions from which the simulation generates independent and identically distributed (i.i.d.) random variates. More generally, the input models are stochastic processes described probabilistically (e.g., their joint distribution), rather than logically (customers are served in the order they arrive). The fidelity with which the input models capture the uncertainty in the system of interest directly affects the fidelity of the conclusions that can be drawn from the simulation experiment.

In this tutorial we focus on input models that are “fit” to representative samples of real-world data. Because the quantity of data is necessarily finite, the resulting input models are almost surely imperfect representations of reality. Our interest is in quantifying this uncertainty in the input models that is propagated to the simulation performance estimates.

Let F denote the collection of input models in the simulation (we will be more specific about this collection later). We will represent the simulation output on replication j as

$$Y_j(F) = \eta(F) + \varepsilon_j(F)$$

where $\eta(F) = E[Y_j(F)]$ is the expected value of the simulation output random variable when we use input distributions F , and $\varepsilon_j(F)$ is a mean 0 random variable representing the stochastic noise that is a consequence of the input models. Since the output could be an average, an indicator variable, a sample variance, a sample quantile or any number of other quantities, this is a very general representation that emphasizes the dependence of the system performance measure $\eta(F)$ and the simulation output $Y_j(F)$ on the input models F .

To execute a simulation experiment we need a specific choice or choices for F , and the natural approach when we have real-world input data is to use a fitted distribution \hat{F} , which is a stand-in for the unknown

true real-world distribution denoted by F^c (the “c” indicates “correct”). Therefore, the simulation results we observe are

$$Y_j(\widehat{F}) = \eta(\widehat{F}) + \varepsilon_j(\widehat{F}).$$

Given the choice \widehat{F} , there is a vast literature on simulation design and analysis that provides guidance on how many replications we need, and what estimator to use, to estimate $\eta(\widehat{F})$. Clearly this analysis is *conditional* on \widehat{F} , and therefore ignores a potentially large source of error due to using $\widehat{F} \neq F^c$.

The *input uncertainty problem* is to quantify how much uncertainty in the simulation estimator of $\eta(F^c)$ is caused by the uncertainty in estimating F^c . In this tutorial we formally define input uncertainty, describe various proposals for measuring it, contrast input uncertainty with input sensitivity, and provide and illustrate a practical approach for quantifying overall input uncertainty and the relative contribution of each input model to it. We start with a high-level discussion of input modeling itself.

2 INPUT MODELING

For ease of exposition, suppose (temporarily) that the system of interest has only one source of randomness represented by a univariate input distribution F^c . Later we will extend to the case of multiple input distributions. Generally F^c is unknown; however, we can observe a realization $\mathbf{z}_m = \{z_1, z_2, \dots, z_m\}$ of the i.i.d. input process $Z_1, Z_2, \dots, Z_m \sim F^c$. Notice that we use capital letters such as Z to denote a random variable, and lowercase letters such as z for a realization. We will also use Z for real-world input data from F^c , and X for simulation-generated input data from \widehat{F} . The estimated distribution \widehat{F} depends on any prior information we have about F^c and the observed data \mathbf{z}_m . Choosing \widehat{F} is called *input modeling*.

What constitutes good input modeling? The goal of the simulation experiment is to precisely estimate the true system performance measure, which is $\eta(F^c)$ in our set up. Therefore, estimating F^c precisely is a sufficient, but not a necessary condition; to achieve our goal we need $\eta(\widehat{F})$ to be close to $\eta(F^c)$, not $\widehat{F} = F^c$. In some specialized situations we can figure out what specific properties of \widehat{F} need to be close to the corresponding properties of F^c for $\eta(\widehat{F})$ to be close to $\eta(F^c)$, but this is certainly not possible in general because $\eta(\cdot)$ is unknown. Therefore, the focus of input modeling is typically on the fidelity of \widehat{F} with respect to F^c .

The two main philosophies of input modeling are frequentist and Bayesian, and while certainly different, they share some common features when it comes to input uncertainty.

Frequentists assume that there exists a true distribution F^c , and the goal is to infer F^c from the real-world data using methods that have good performance when averaged over the possible samples Z_1, Z_2, \dots, Z_m . For instance, if F^c is believed to have a known distribution family but unknown parameters θ^c —that is, $F^c(x) = F(x|\theta^c)$ with F known—then the parameters may be estimated using maximum likelihood estimators (MLEs), least squares, generalized method of moments, or other methods. Typically these estimators are asymptotically consistent for θ^c as $m \rightarrow \infty$. Prior information such as the physical basis of the input process may influence the choice of distribution family. There is commercial input modeling software that fits the parameters of a large collection of distribution families and also provides ways to assess the fit, including goodness-of-fit tests such as Chi-square, Anderson-Darling and Kolmogorov-Smirnov. Errors in fitting are characterized by the sampling distribution of the parameter estimator $\widehat{\theta}$; since this distribution is often hard to derive, approximations based on large-sample asymptotics or bootstrapping are often used.

Bayesians are less interested in discovering the true distribution F^c and more interested in the likelihood of possible choices given the data. Bayesian inference starts by capturing any prior knowledge available about F^c in the form of a prior probability distribution; this distribution may be on the family of distributions, the parameters of a specific family of distributions, or both. For instance, we might believe that F^c is one of the families in the set $\{\mathcal{F}_1, \mathcal{F}_2, \dots, \mathcal{F}_k\}$ with prior probabilities $\pi_1, \pi_2, \dots, \pi_k$. More often, we assume that the family of F^c is known and provide a prior $\pi_{\Theta}(\theta)$ on the value of its parameter vector Θ , which is treated as a random vector.

After collecting the real-world data \mathbf{z}_m , the prior distribution $\pi_{\Theta}(\theta)$ is updated to a posterior distribution $p_{\Theta}(\theta|\mathbf{z}_m)$ according to Bayes' rule

$$p_{\Theta}(\theta|\mathbf{z}_m) \propto L(\mathbf{z}_m; \theta)\pi_{\Theta}(\theta),$$

where $L(\mathbf{z}_m; \theta)$ is the likelihood function of \mathbf{z}_m given $\Theta = \theta$. Notice that the posterior distribution is conditional on the particular realization \mathbf{z}_m that was obtained. The posterior distribution $p_{\Theta}(\theta|\mathbf{z}_m)$ plays the role for Bayesians that the sampling distribution of $\hat{\theta}$ plays for frequentists by quantifying the uncertainty about θ^c . Under some regularity conditions, the posterior distribution of Θ converges to a degenerate distribution at θ^c independent of the choice of the prior distribution as the sample size $m \rightarrow \infty$; this is the Bayesian concept of consistency.

Returning to the main purpose of the simulation study—to estimate $\eta(F^c)$ —both frequentist and Bayesian input modelers face a similar problem: to actually execute a simulation experiment requires a specific choice of distribution and distribution parameters. When only the parameters are unknown, we often use the single “best” choice, such as the MLE for frequentists or the maximum a posteriori (MAP) estimator for Bayesians, but a single choice assumes away the very real uncertainty about the input model. However, since $\eta(\cdot)$ is unknown, characterizing uncertainty about the input model (via a sampling distribution or posterior) is not enough since the goal is to measure the uncertainty in estimating $\eta(F^c)$. Therefore, to quantify input uncertainty, some way to propagate the uncertainty about F^c through to the estimate of $\eta(F^c)$ is required.

In the remainder of the tutorial we define some specific measures of input uncertainty and illustrate one practical approach to estimate them.

3 MEASURES OF INPUT UNCERTAINTY

One difficulty in describing input uncertainty is that there are many types of input models that may arise in a computer simulation. These include univariate inputs such as customer service times; time-series inputs such as the weekly demands for milk; multivariate inputs such as the age, income, gender and occupation of a customer; and time-dependent inputs such as the arrival times of calls to a call center. For this tutorial we will assume that the input models for the simulation consist of L independent, univariate distributions $\mathbf{F} = \{F_1, F_2, \dots, F_L\}$ each of whose role is to model an i.i.d. input process. For instance, in a stationary queueing system with server failures we might have $\mathbf{F} = \{F_1, F_2, F_3, F_4\}$, the distributions of interarrival times, service times, time until server failure, and server repair time, respectively. Let $Y(\mathbf{F})$ be a random variable that represents the generic output of the simulation—such as the average customer delay in the queueing system with failures—when the input distributions are \mathbf{F} .

We assume that there are unknown, but “correct,” real-world distributions $\mathbf{F}^c = \{F_1^c, F_2^c, \dots, F_L^c\}$. From the ℓ th distribution we are able to observe m_{ℓ} i.i.d. real-world observations $Z_{\ell 1}, Z_{\ell 2}, \dots, Z_{\ell m_{\ell}}$ from which to “fit” F_{ℓ}^c , giving the collection of fitted input models $\hat{\mathbf{F}} = \{\hat{F}_1, \hat{F}_2, \dots, \hat{F}_L\}$. We then run the simulation and obtain outputs $Y_1(\hat{\mathbf{F}}), Y_2(\hat{\mathbf{F}}), \dots, Y_n(\hat{\mathbf{F}})$ across n i.i.d. replications. *The goal of the simulation experiment is to estimate $\eta(\mathbf{F}^c) = E[Y(\mathbf{F}^c)]$.* For ease of exposition here we assume that the $\eta(\mathbf{F}^c)$ is estimated by the sample mean across n replications

$$\begin{aligned} \bar{Y}(\hat{\mathbf{F}}) &= \frac{1}{n} \sum_{j=1}^n Y_j(\hat{\mathbf{F}}) \\ &= \eta(\hat{\mathbf{F}}) + \frac{1}{n} \sum_{j=1}^n \varepsilon_j(\hat{\mathbf{F}}). \end{aligned}$$

The variance of the point estimator is

$$\begin{aligned} \text{Var}[\bar{Y}(\hat{\mathbf{F}})] &= \text{Var} \left[\text{E}(\bar{Y}(\hat{\mathbf{F}})|\hat{\mathbf{F}}) \right] + \text{E} \left[\text{Var}(\bar{Y}(\hat{\mathbf{F}})|\hat{\mathbf{F}}) \right] \\ &= \text{Var} \left[\eta(\hat{\mathbf{F}}) \right] + \frac{1}{n} \text{E} \left[\text{Var}(\varepsilon_1(\hat{\mathbf{F}})|\hat{\mathbf{F}}) \right], \end{aligned} \quad (1)$$

where outer variance and expectation on the right hand side is with respect to the sampling distribution of $\hat{\mathbf{F}}$ (frequentist) or the posterior distribution of $\hat{\mathbf{F}}$ (Bayesian). Notice that the second term in (1) accounts for the inherent variability in the simulation error (given a choice of input distributions), and it can be driven to 0 by increasing the simulation effort n . The size of the first term, however, depends on a complex interaction between the real-world sample sizes $\mathbf{m} = \{m_1, m_2, \dots, m_L\}$ and the structure of $\eta(\cdot)$.

Our formal definition of input uncertainty is $\sigma_I^2 = \text{Var} \left[\eta(\hat{\mathbf{F}}) \right]$, the variance in the system mean due to having estimated \mathbf{F}^c . Although this definition is straightforward, how to estimate it is not immediately clear as it requires 1) the sampling distribution of $\hat{\mathbf{F}}$ (frequentist case) and 2) the functional form of $\eta(\cdot)$, neither of which is known in general. This illustrates why assessing input uncertainty is difficult and motivates the methods that have been suggested in the literature and are reviewed below.

Conceptually, if we had B independent real-world samples of input data of size \mathbf{m} , yielding B independent fitted input distributions $\hat{\mathbf{F}}_1, \hat{\mathbf{F}}_2, \dots, \hat{\mathbf{F}}_B$, and we knew the true mean response functional $\eta(\cdot)$, then we could estimate σ_I^2 via the unbiased estimator

$$\hat{\sigma}_I^2 = \frac{1}{B-1} \sum_{b=1}^B (\eta(\hat{\mathbf{F}}_b) - \bar{\eta})^2, \text{ where } \bar{\eta} = \frac{1}{B} \sum_{b=1}^B \eta(\hat{\mathbf{F}}_b).$$

However, if we actually had multiple real-world data sets then we would certainly pool them to obtain a better estimator of \mathbf{F}^c ; thus, no matter how much data we have it is best to treat it as one sample. In addition, $\eta(\cdot)$ is unknown or we would not be simulating. In the next two subsections we consider these problems (single sample, unknown response function) in turn, both from frequentist and Bayesian points of view.

It is worth mentioning that one other aspect of input uncertainty that is rarely noted, even in the input-uncertainty literature, is bias. Specifically,

$$\text{E} \left[Y(\hat{\mathbf{F}}) - \eta(\mathbf{F}^c) \right] = \text{E} \left[\eta(\hat{\mathbf{F}}) \right] - \eta(\mathbf{F}^c) \neq 0$$

in general. This is the case even when $\text{E}[\hat{\mathbf{F}}] = \mathbf{F}^c$ because $\eta(\cdot)$ is typically a nonlinear transformation. We might hope that $\eta(\hat{\mathbf{F}}) \rightarrow \eta(\mathbf{F}^c)$ in some sense as the real-world sample sizes $m_\ell \rightarrow \infty$, $\ell = 1, 2, \dots, L$, but this is the most we can hope.

3.1 Variability of $\hat{\mathbf{F}}$

For a frequentist, the variability of $\hat{\mathbf{F}}$ as an estimator of \mathbf{F}^c is quantified by its sampling distribution; that is, we try to infer the distribution of possible $\hat{\mathbf{F}}$'s that could be realized from the single sample of data that we actually have. There have been two primary approaches in the literature: large-sample asymptotic distributions as in Cheng and Holland (1998) and bootstrapping as in Barton and Schruben (2001), Ankenman and Nelson (2012), and Song and Nelson (2013).

To simplify the exposition, suppose again that there is only $L = 1$ input distribution, and we believe $F^c(z) = F(z|\theta^c)$ is a parametric distribution with known family (e.g., Poisson), but unknown parameter vector θ^c . The family might be “known” because of process physics (e.g., an arrival process consisting of the superposition of many customers making independent but infrequent decisions about when to arrive tends to be Poisson), or from fitting a large number of possible parametric families and ranking their goodness of fit.

For the parametric case there is a well-developed theory for parameter estimators $\hat{\theta} = \hat{\theta}(Z_1, Z_2, \dots, Z_m)$ such as the MLE. The sampling distributions for these estimators are known, at least asymptotically as $m \rightarrow \infty$. For instance, MLEs are often asymptotically normally distributed. Thus, for parametric distributions, input *model* uncertainty becomes input *parameter* uncertainty which is quantified by the sampling distribution of $\hat{\theta}$. Cheng and Holland (1998) provide an expression for σ_I^2 by applying a first-order Taylor series approximation of $\eta(\hat{\theta})$ to obtain

$$\sigma_I^2 = \text{Var}[\eta(\hat{\theta})] \approx \nabla(\eta(\theta^c))^\top \text{Var}[\hat{\theta}] \nabla(\eta(\theta^c)) \quad (2)$$

where $\text{Var}[\hat{\theta}]$ is the variance-covariance matrix of $\hat{\theta}$ and $\nabla(\eta(\theta^c))$ is the gradient of $\eta(\cdot)$ evaluated at θ^c . The approximation (2) shows that input uncertainty depends both on how well the input model has been estimated, and how sensitive the system response is to the input model. The sampling distribution (or an approximation to it) can provide an estimator of $\text{Var}[\hat{\theta}]$, but that still leaves the question of how to estimate $\nabla(\eta(\theta^c))$.

In bootstrapping, we assume that the fitted distribution \hat{F} is a good stand-in for F^c . An empirical cumulative distribution function (ecdf) is often used as the fitted distribution, although this is not required. Then instead of gathering B distinct real-world samples of input data, we simulate B bootstrap samples from the empirical distribution:

$$X_1^{*(b)}, X_2^{*(b)}, \dots, X_m^{*(b)} \stackrel{i.i.d.}{\sim} \hat{F} \text{ for } b = 1, 2, \dots, B.$$

Each bootstrap sample yields a fitted input distribution $\hat{F}^{*(b)}$. Therefore, an estimator of σ_I^2 is

$$\hat{\sigma}_I^{2*} = \frac{1}{B-1} \sum_{b=1}^B (\eta(\hat{F}^{*(b)}) - \bar{\eta}^*)^2$$

where $\bar{\eta}^* = \sum_{b=1}^B \eta(\hat{F}^{*(b)})/B$. The validity of the bootstrap can also be established as $m \rightarrow \infty$. However, this estimator requires a stand-in for $\eta(\cdot)$.

Bayesian input modeling typically emphasizes the choice of parameters for a parametric distribution, although as noted earlier it is possible to extend the definition of “parameter” to include the distribution family as well. The posterior distribution of the parameters $p_{\Theta}(\theta|\mathbf{z}_m)$ plays the role of the sampling distribution of $\hat{\theta}$ for the frequentist. However, in a Bayesian treatment there is no need for an asymptotic approximation because the posterior is valid for any quantity of real-world data. In fact, the posterior explicitly accounts for the quantity of data that is available: the more data there is, the more concentrated the posterior distribution will be.

From a Bayesian perspective, the corresponding overall measure of input uncertainty is $\sigma_I^2 = \text{Var}[\eta(\Theta)]$ where the variance is with respect to $\Theta \sim p_{\Theta}(\theta|\mathbf{z}_m)$. One difficulty in computing this quantity arises when the posterior distribution is not simple and can only be evaluated approximately via simulation methods, such as Markov chain Monte Carlo. And, of course, $\eta(\cdot)$ is not known.

Instead of using the posterior distribution of Θ , Ng and Chick (2001, 2006) employ the asymptotic normal distribution of the MAP estimator $\tilde{\Theta}_m$. Similar to Approximation (2), they combine the asymptotic normal approximation with a first-order Taylor series expansion of $\eta(\tilde{\Theta}_m)$ to derive σ_I^2 :

$$\hat{\sigma}_I^2 \approx \nabla(\eta(\tilde{\theta}_m))^\top \Sigma_m \nabla(\eta(\tilde{\theta}_m)),$$

where $\tilde{\theta}_m$ is the MAP of Θ and Σ_m is the inverse of the Bayesian observed information given \mathbf{z}_m . Likewise, they need an assumption on the structure of the unknown $\eta(\cdot)$ to propagate the uncertainty in Θ to the simulation estimator.

3.2 Propagating Input Model Uncertainty

The second reason that $\hat{\sigma}_I^2$ is not realizable is that $\eta(\cdot)$ is unknown, and it is this mapping that propagates the input model uncertainty to the simulation performance measure uncertainty. Whether frequentist or Bayesian, there is really no choice but to estimate $\eta(\cdot)$ using simulation, and we can only estimate $\eta(\mathbf{F})$ given specific choices of \mathbf{F} . The key questions are, what should we estimate, and how should we expend the simulation effort to do it?

Direct simulation chooses a collection of distributions $\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_B$, runs simulations at each choice, and estimates $\eta(\mathbf{F}_i)$ by the average of the simulation results $\bar{Y}(\mathbf{F}_i)$. The distributions $\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_B$ are usually randomly chosen from the sampling distribution of $\hat{\theta}$ (frequentist) or the posterior distribution of Θ (Bayesian) so as to represent the input model uncertainty.

1. For $b = 1$ to B
 - (a) If frequentist:
 - i. Generate $\hat{\theta}_b^*$ from the sampling distribution of $\hat{\theta}$ (e.g., using the asymptotic distribution or bootstrapping).
 - ii. Using $\hat{\theta}_b^*$, run n replications to obtain the sample mean $\bar{Y}_b(\hat{\theta}_b^*)$.
 - (b) Else if Bayesian:
 - i. Generate $\Theta_b \sim p_{\Theta}(\theta|\mathbf{z}_m)$
 - ii. Using Θ_b , run n replications to obtain the sample mean $\bar{Y}_b(\Theta_b)$.
- Next b
2. Use the data $\bar{Y}_b(\hat{\theta}_b^*)$ or $\bar{Y}_b(\Theta_b)$ for $b = 1, 2, \dots, B$ to estimate measures of input uncertainty.

Notice that in either case input uncertainty is confounded with stochastic simulation noise so they may need to be separated, which is the challenge.

Metamodeling chooses distributions $\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_d$ to cover a relevant design space, and then uses the simulation results to fit a metamodel $\hat{\eta}(\cdot)$ to stand in for $\eta(\cdot)$. Barton et al. (2014) take a frequentist approach to input modeling and use Gaussian processes to build a metamodel at a chosen set of parameters. Given the metamodel, $\hat{\sigma}_I^2$ or a number of other measures can be estimated inexpensively. But there is still a need to separate input uncertainty from metamodel error.

Gradients of $\eta(\theta)$ are needed if we use parametric input models and approximate $\text{Var}[\eta(\hat{\theta})]$ using a Taylor series expansion of $\eta(\hat{\theta})$ around θ^c as in (2). Since $\eta(\theta)$ is unknown, its gradients are also unknown. Cheng and Holland (2004) suggest several methods to estimate $\nabla(\eta(\theta^c))$. The expansion in (2) facilitates propagating uncertainty about θ (as represented by the sampling distribution of $\hat{\theta}$ or the posterior distribution of Θ) to the performance measure $\eta(\hat{\theta})$. Clearly this is another form of metamodeling.

3.3 Additional Measures of Input Uncertainty

Although we focused on estimating σ_I^2 , there are other useful measures depending on the objective.

In (1) let $\sigma^2 = \text{E} \left[\text{Var}(\varepsilon_1(\hat{\mathbf{F}})|\hat{\mathbf{F}}) \right]$, the simulation output variance. Then Ankenman and Nelson (2012) standardize σ_I^2 as

$$\gamma = \frac{\sigma_I}{\sigma/\sqrt{n}}.$$

Thus, the ratio γ is the standard deviation due to input uncertainty in units of the standard error of the simulation point estimator. For instance, $\gamma = 0.5$ means that the error due to input uncertainty is half as large as that due to estimating $\eta(\hat{\mathbf{F}})$ via simulation. If the replication number n is sufficiently large, small γ indicates that the input uncertainty is insignificant.

Interval estimates are also valuable. The typical simulation confidence interval (CI) guarantees $(1 - \alpha)$ 100% coverage of $\eta(\hat{\mathbf{F}})$. Instead, an interval $[C_L, C_U]$ that guarantees

$$\Pr\{\eta(\mathbf{F}^c) \in [C_L, C_U]\} \approx 1 - \alpha$$

accounts for input model uncertainty, simulation error and bias. The corresponding Bayesian credible interval (CrI) would assert

$$\Pr\{\eta(\Theta) \in [Q_L, Q_U]\} = 1 - \alpha.$$

This interval contains $1 - \alpha$ of the probability content of the induced posterior distribution on $\eta(\Theta)$ when $\Theta \sim p_{\Theta}(\theta|\mathbf{z}_m)$. Many papers have proposed generating CIs or CrIs from direct simulation, as described above, using empirical quantiles of $\bar{Y}_b(\hat{\theta}_b^*)$ or $\bar{Y}_b(\Theta_b)$, respectively, for $b = 1, 2, \dots, B$. See, for instance, Xie et al. (2014) and Zouaoui and Wilson (2003, 2004).

So far we have only considered the measures of input uncertainty as a function of $\hat{\mathbf{F}}$, the collection of all input models used in the simulation. We next turn our attention to assessing which input models among $\hat{\mathbf{F}}$ make the greatest contributions to input uncertainty, and from which would we most benefit from collecting more real-world data.

3.4 Measures of Contribution and Sample-Size Sensitivity

Contribution measures try to decompose σ_I^2 in a meaningful way. This is difficult, and in fact is currently an area of active research. The ultimate goal is to identify from which distributions it would be most beneficial to reduce uncertainty further, or even to specify how to spend a budget for additional data collection.

Clearly, the contribution of ℓ th input model \hat{F}_ℓ depends on the real-world sample size m_ℓ ; as m_ℓ becomes infinitely large, \hat{F}_ℓ converges to F_ℓ^c (frequentist) or to a degenerate posterior distribution (Bayesian). Therefore, small m_ℓ causes bigger contribution. However, the contribution also depends on how sensitive $\eta(\cdot)$ is to F_ℓ . To make the point clear, assume that F_1 is in fact a dummy distribution so that the functional $\eta(\cdot)$ does not depend on F_1 . Then, no matter how small m_1 is, the contribution of \hat{F}_1 should be 0. On the other hand, if $\eta(\cdot)$ is very sensitive to F_1 , the uncertainty in \hat{F}_1 is amplified as it is propagated to the uncertainty in $\eta(\hat{\mathbf{F}})$.

Song and Nelson (2013) defined the contribution of \hat{F}_ℓ to input uncertainty as

$$V_\ell(m_\ell) \equiv \text{Var} \left[\text{E}[Y(F_1^c, F_2^c, \dots, F_{\ell-1}^c, \hat{F}_\ell, F_{\ell+1}^c, \dots, F_L^c) | \hat{F}_\ell] \right]. \quad (3)$$

In words, $V_\ell(m_\ell)$ is the variability in the simulation's expected value when all of the true input distributions except F_ℓ^c are known and F_ℓ^c is estimated by \hat{F}_ℓ . Notice that V_ℓ is a function of the sample size m_ℓ ; the larger m_ℓ is, the smaller $V_\ell(m_\ell)$ becomes as \hat{F}_ℓ approaches F_ℓ^c .

Unfortunately, $\sum_{\ell=1}^L V_\ell(m_\ell) \neq \sigma_I^2$ in general. Song and Nelson (2013) suggest looking instead at the relative contribution of the ℓ th input model,

$$\frac{V_\ell(m_\ell)}{\sum_{i=1}^L V_i(m_i)}.$$

A reasonable heuristic to reduce σ_I^2 is to obtain more data for the distributions with the largest relative contribution.

Song and Nelson (2013) propose a measure that does not decompose σ_I^2 , but instead quantifies the impact of additional data for each input distribution: the sample-size sensitivity of $\text{Var}[\bar{Y}(\hat{\mathbf{F}})]$ with respect to ℓ th input model. If we approximate m_ℓ as real valued then this is

$$\left. \frac{\partial \text{Var}[\bar{Y}(\hat{\mathbf{F}})]}{\partial m'_\ell} \right|_{m'_\ell = m_\ell}. \quad (4)$$

The sample-size sensitivity quantifies how much the estimator variance would be reduced by observing one more real-world sample from the ℓ th input process given we already have m_ℓ observations. The input distributions with the largest (most negative) sensitivities are targets for more real-world data. We illustrate one method for estimating the relative contribution and sample-size sensitivity in Section 4.

Unfortunately, sample-size sensitivities are only local gradients, and therefore are not ideal for optimally allocating a substantial amount of additional effort. Freimer and Schruben (2002) and Ng and Chick (2001, 2006) attempt to go further.

If the budget for real-world data is effectively unlimited, then it makes sense to collect enough data so that $\text{Var}[\bar{Y}(\hat{\mathbf{F}})] \approx \text{E}[\text{Var}(\bar{Y}(\hat{\mathbf{F}})|\hat{\mathbf{F}})]$; i.e., no input uncertainty. Freimer and Schruben (2002) achieve this objective by sequentially adding real-world data until the effect of input model uncertainty is statistically negligible relative to simulation output variability. To formalize this they use a random-effects metamodel where the “random effects” are due to distribution parameter estimates obtained from bootstrapping the available data. In the case of $L = 2$ parameters, their model for the simulation output on replication h is

$$Y_{ijh} = \eta + A_i + C_j + AC_{ij} + \varepsilon_{ijh}, \text{ for } i, j = 1, 2, \dots, B, h = 1, 2, \dots, n \quad (5)$$

where A_i , C_j and AC_{ij} are the random effects of bootstrapped parameters θ_{1i}^* , θ_{2j}^* and their interaction, respectively, B is the number of bootstrap samples, and ε_{ijh} is the stochastic variability of the simulation. They stop collecting input data when the hypothesis that these three variance effects are 0 is no longer rejected.

If the budget for collecting real-world data, and perhaps also for running the simulation itself, is limited, then it makes sense to try to allocate effort to minimize the variance of the point estimator within the budget. This requires an approximation for the variance as a function of \mathbf{m} and n . Ng and Chick (2001, 2006) obtain such an approximation by combining a Bayesian treatment of input model uncertainty with a linear metamodel for the simulation output

$$Y(\Theta) = \eta(\Theta) + \varepsilon \approx \lambda_0 + \sum_{p=1}^P \lambda_p(\Theta_p - \tilde{\theta}_p) + \varepsilon \quad (6)$$

where P is the total number of individual parameters of all input distributions, and $\tilde{\theta}_p$ is the MAP of the p th parameter Θ_p . Estimation of the metamodel parameters is also treated in a Bayesian manner. Using this model they derive an expression for the posterior variance of $\eta(\Theta)$ as a function of \mathbf{m} and n , which they use to optimally allocate the budget for additional input data or additional simulation replications.

3.5 Connections to Sensitivity Analysis

“Sensitivity analysis,” as it is usually construed in stochastic simulation, means estimating terms like $\nabla g(\theta)$ at given values of θ . As we noted in approximation (2), input uncertainty is more than just sensitivity. As illustrated in Section 3.4, input uncertainty is a function of both the size of the real-world data set from which we estimate the input distribution parameters and the sensitivity of the response to the parameter.

There is at least a superficial connection between input uncertainty and global or probabilistic sensitivity. Key references include Wagner (1995), Homma and Saltelli (1996), Oakley and O’Hagan (2004), and Plischke et al. (2013). Expressing the global sensitivity problem in our notation, it is concerned with a response $y = \eta(\theta)$ where $\eta(\cdot)$ is a function that can be evaluated without noise, but usually at great computational expense. The problem arises because the parameter θ is either unknown or uncontrollable (e.g., it contains environmental factors). Therefore, its value is modeled as a random variable Θ with *known* distribution $F_\Theta(\cdot)$. The response can also be treated a random variable $Y = \eta(\Theta)$ with $\Theta \sim F_\Theta$. Global sensitivity analysis tries to decompose or explain $\text{Var}[Y]$ by attributing portions of it to each component parameter of $\Theta = (\Theta_1, \Theta_2, \dots, \Theta_L)$. Research is driven by the fact that direct evaluation of $\eta(\cdot)$ is too expensive to do for many values of θ , rather than by lack of knowledge of F_Θ . Our input uncertainty problem, on the other hand, is characterized by lack of knowledge of the input distributions themselves and the presence of simulation noise in the evaluation of $\eta(\cdot)$.

4 A PRACTICAL APPROACH AND ILLUSTRATION

In this section, we summarize a method suggested by Song and Nelson (2014) to estimate the input uncertainty measures introduced in Section 3, along with an illustration of its application on a practical problem. This approach takes a frequentist's view of input modeling and input uncertainty quantification. We continue to assume that the collection of input models \mathbf{F} consist of L independent, univariate distributions. The goal of the experiment is to estimate $\eta(\mathbf{F}^c)$.

Song and Nelson (2014) describe a sequence of experiments: *nominal*, *diagnostic* and *follow-up*. Typically, a simulation analyst performs a nominal experiment by fitting the input models $\widehat{\mathbf{F}} = \{\widehat{F}_1, \widehat{F}_2, \dots, \widehat{F}_L\}$ from a real-world sample $\mathbf{z}_m = \{\mathbf{z}_{m_1}, \mathbf{z}_{m_2}, \dots, \mathbf{z}_{m_L}\}$ and conducting multiple, say n , simulation replications to obtain a point estimator $\bar{Y}(\widehat{\mathbf{F}})$ and its CI conditional on $\widehat{\mathbf{F}}$. As noted earlier, this estimator and CI do not take into account input uncertainty caused by having estimated \mathbf{F}^c by $\widehat{\mathbf{F}}$. To address this problem, Song and Nelson (2014) introduce the diagnostic and follow-up experiments, described in the following subsection.

4.1 The diagnostic and follow-up experiments

The purpose of the diagnostic experiment is to quantify the input uncertainty in the estimator from the nominal experiment using the measures defined in Section 3. Song and Nelson (2014) employ the following *mean-variance effects metamodel* as a stand-in for $Y_j(\mathbf{F})$:

$$Y_j(\mathbf{F}) = \eta(\mathbf{F}) + \varepsilon_j = \beta_0 + \sum_{\ell=1}^L \beta_\ell \mu(F_\ell) + \sum_{\ell=1}^L v_\ell \sigma^2(F_\ell) + \varepsilon_j, \quad (7)$$

where $\mu(F_\ell)$ and $\sigma^2(F_\ell)$ denote the mean and variance, respectively, of F_ℓ ; β_ℓ and v_ℓ are constant coefficients; and ε_j is an independent mean-zero random variable with constant variance σ^2 . The idea behind Model (7) is that the effect of F_ℓ on the mean response can be captured by some summary properties of F_ℓ ; in particular, the realized center (mean) and spread (variance) of F_ℓ . Also, β_ℓ and v_ℓ represent how sensitive $\eta(\cdot)$ is to the ℓ th input distribution. We do not expect Model (7) to fit well over the entire sampling space of $\widehat{\mathbf{F}}$; rather, it is expected to have a good local fit for $\widehat{\mathbf{F}}$ near \mathbf{F}^c .

Under Model (7), the simulation uncertainty σ_f^2 can be represented as

$$\sigma_f^2 = \text{Var}[\eta(\widehat{\mathbf{F}})] = \sum_{\ell=1}^L \left\{ \beta_\ell^2 \text{Var}[\mu(\widehat{F}_\ell)] + v_\ell^2 \text{Var}[\sigma^2(\widehat{F}_\ell)] + 2\beta_\ell v_\ell \text{Cov}[\mu(\widehat{F}_\ell), \sigma^2(\widehat{F}_\ell)] \right\}. \quad (8)$$

Likewise, the contribution of \widehat{F}_ℓ to the input uncertainty can be derived from the definition in (3) as

$$\begin{aligned} V_\ell(m_\ell) &= \text{Var} \left[\beta_0 + \sum_{i=1, i \neq \ell}^L \beta_i \mu(F_i^c) + \sum_{i=1, i \neq \ell}^L v_i \sigma^2(F_i^c) + \beta_\ell \mu(\widehat{F}_\ell) + v_\ell \sigma^2(\widehat{F}_\ell) \right] \\ &= \beta_\ell^2 \text{Var}[\mu(\widehat{F}_\ell)] + v_\ell^2 \text{Var}[\sigma^2(\widehat{F}_\ell)] + 2\beta_\ell v_\ell \text{Cov}[\mu(\widehat{F}_\ell), \sigma^2(\widehat{F}_\ell)]. \end{aligned} \quad (9)$$

Notice that under Model (7) the overall input uncertainty is the sum of each distribution's contribution, i.e., $\sigma_f^2 = \sum_{\ell=1}^L V_\ell(m_\ell)$. Hence, the sample-size sensitivity defined as in (4) becomes

$$\left. \frac{\partial \text{Var}[\bar{Y}(\widehat{\mathbf{F}})]}{\partial m'_\ell} \right|_{m'_\ell = m_\ell} = \left. \frac{\partial V_\ell(m'_\ell)}{\partial m'_\ell} \right|_{m'_\ell = m_\ell}.$$

To estimate the contributions from (9), we need to estimate the parameters β_ℓ and v_ℓ , and the variance components of the sample statistics, $\text{Var}[\mu(\widehat{F}_\ell)]$, $\text{Var}[\sigma^2(\widehat{F}_\ell)]$ and $\text{Cov}[\mu(\widehat{F}_\ell), \sigma^2(\widehat{F}_\ell)]$, for $\ell = 1, 2, \dots, L$. If we had B collections of real-world data, then the parameters β_ℓ and v_ℓ could be obtained by fitting

Model (7) using the corresponding fitted distribution $\widehat{\mathbf{F}}_1, \widehat{\mathbf{F}}_2, \dots, \widehat{\mathbf{F}}_B$. However, as discussed in Section 3.1, we prefer to pool all real-world data to obtain a better estimate of \mathbf{F}^c . Instead, Song and Nelson (2014) bootstrap $\widehat{\mathbf{F}}$ to obtain ecdfs $\widehat{\mathbf{F}}_1^*, \widehat{\mathbf{F}}_2^*, \dots, \widehat{\mathbf{F}}_B^*$ from bootstrapped samples and fit Model (7) via least-squares regression to estimate the β_ℓ 's and ν_ℓ 's.

Another advantage of the bootstrap approximation is that it yields simple expressions for the variance components in (9). Bootstrapping implies that we assume $\text{Var}[\mu(\widehat{F}_\ell^*)|\widehat{F}_\ell]$, $\text{Var}[\sigma^2(\widehat{F}_\ell^*)|\widehat{F}_\ell]$, and $\text{Cov}[\mu(\widehat{F}_\ell^*), \sigma^2(\widehat{F}_\ell^*)|\widehat{F}_\ell]$ are stand-ins for $\text{Var}[\mu(\widehat{F}_\ell)]$, $\text{Var}[\sigma^2(\widehat{F}_\ell)]$, and $\text{Cov}[\mu(\widehat{F}_\ell), \sigma^2(\widehat{F}_\ell)]$. Song and Nelson (2014) derive

$$\begin{aligned} \text{Var}[\mu(\widehat{F}_\ell^*)|\widehat{F}_\ell] &= \frac{M_\ell^2}{m_\ell} \\ \text{Var}[\sigma^2(\widehat{F}_\ell^*)|\widehat{F}_\ell] &= \frac{(m_\ell - 1)^2}{m_\ell^3} M_\ell^4 - \frac{(m_\ell - 3)(m_\ell - 1)}{m_\ell^3} (M_\ell^2)^2 \approx \frac{M_\ell^4 - (M_\ell^2)^2}{m_\ell} \\ \text{Cov}[\mu(\widehat{F}_\ell^*), \sigma^2(\widehat{F}_\ell^*)|\widehat{F}_\ell] &= \frac{(m_\ell - 1)^2}{m_\ell^3} M_\ell^3 \approx \frac{M_\ell^3}{m_\ell} \end{aligned}$$

where M_ℓ^k is k th central moment of \widehat{F}_ℓ . Therefore, the contribution and sample-size sensitivity estimator can be written as

$$\begin{aligned} \widehat{V}_\ell(m_\ell) &= \frac{1}{m_\ell} \left\{ \widehat{\beta}_\ell^2 M_\ell^2 + \widehat{\nu}_\ell^2 (M_\ell^4 - (M_\ell^2)^2) + 2\widehat{\beta}_\ell \widehat{\nu}_\ell M_\ell^3 \right\}, \\ \left. \frac{\partial \widehat{V}_\ell(m'_\ell)}{\partial m'_\ell} \right|_{m'_\ell = m_\ell} &= -\frac{\widehat{V}_\ell(m_\ell)}{m_\ell}. \end{aligned} \tag{10}$$

The algorithm for the diagnostic experiment is as follows:

Diagnostic Experiment

1. Given the estimated input models $\widehat{\mathbf{F}}$ (either parametric or non-parametric), do the following:
 2. For bootstrap sample $b = 1$ to B
 - (a) For input model $\ell = 1$ to L
 - i. Generate $X_{\ell 1}^{*(b)}, X_{\ell 2}^{*(b)}, \dots, X_{\ell m_\ell}^{*(b)}$ by sampling m_ℓ times from \widehat{F}_ℓ .
 - ii. Let $\widehat{F}_\ell^{*(b)}$ be the ecdf of $X_{\ell 1}^{*(b)}, X_{\ell 2}^{*(b)}, \dots, X_{\ell m_\ell}^{*(b)}$ and calculate the mean $\mu(\widehat{F}_\ell^{*(b)})$ and variance $\sigma^2(\widehat{F}_\ell^{*(b)})$.
 - Next ℓ
 - (b) Using input models $\widehat{\mathbf{F}}^{*(b)} = \{\widehat{F}_1^{*(b)}, \widehat{F}_2^{*(b)}, \dots, \widehat{F}_L^{*(b)}\}$, simulate R i.i.d. replications $Y_j(\widehat{\mathbf{F}}^{*(b)})$, $j = 1, 2, \dots, R$ and calculate the sample mean $\bar{Y}(\widehat{\mathbf{F}}^{*(b)})$.
 - Next b
 3. Fit the model

$$Y_j(\widehat{\mathbf{F}}^*) = \beta_0 + \sum_{\ell=1}^L \beta_\ell \mu(\widehat{F}_\ell^*) + \sum_{\ell=1}^L \nu_\ell \sigma^2(\widehat{F}_\ell^*) + \varepsilon_j \tag{11}$$

via least-squares regression with dependent variable $\bar{Y}(\widehat{\mathbf{F}}^{*(b)})$ from Step 2b and independent variables $\mu(\widehat{F}_1^{*(b)}), \mu(\widehat{F}_2^{*(b)}), \dots, \mu(\widehat{F}_L^{*(b)})$ and $\sigma^2(\widehat{F}_1^{*(b)}), \sigma^2(\widehat{F}_2^{*(b)}), \dots, \sigma^2(\widehat{F}_L^{*(b)})$ from Step 2a for $b = 1, 2, \dots, B$ to estimate the coefficients $\beta_0, \beta_1, \dots, \beta_L, \nu_1, \nu_2, \dots, \nu_L$.

4. Estimate the contribution $\widehat{V}_\ell(m_\ell)$ and the sensitivity for $\ell = 1, 2, \dots, L$ using (10).

5. Estimate the overall input uncertainty $\hat{\sigma}_I^2 = \sum_{\ell=1}^L \hat{V}_\ell(m_\ell)$ and the estimated ratio

$$\hat{\gamma} = \frac{\hat{\sigma}_I}{\hat{\sigma}/\sqrt{n}} = \frac{\sqrt{n \sum_{\ell=1}^L \hat{V}_\ell(m_\ell)}}{\hat{\sigma}}.$$

Notice that $\hat{\sigma}^2$ in Step 5 is an estimator of the simulation error variance from the nominal experiment and n is the number of replications in the nominal experiment.

Because the diagnostic experiment requires additional simulation effort, Song and Nelson (2014) also study how to break the given simulation budget $N = BR$ into B bootstraps of R replications each. For the estimation quality, large B is beneficial because it reduces the variance of the estimated $\hat{\beta}_\ell$'s and \hat{v}_ℓ 's. However, from a computational perspective, small B is preferred. Therefore, finding a balance between these two is the key to the experiment design. One suggestion made by Song and Nelson (2014) is to allocate N according to

$$B > 2L + 2 + \sqrt{\frac{2L + 2}{\delta}} \quad (12)$$

with $R = \lfloor N/B \rfloor$. This choice makes B just as big enough so that the incremental variance reduction of increasing B by 1 is less than $\delta 100\%$, where they recommend $\delta = 0.01$.

As mentioned in Section 3.3, the estimated ratio $\hat{\gamma}$ can be used to decide the significance of the input uncertainty in the simulation estimator provided that the number of replications n for the nominal experiment is sufficiently large. If $\hat{\gamma}$ is large, e.g., $\hat{\gamma} = 20$, then we can conclude that there is substantial input uncertainty and move on to the problem of deciding from which input processes we should collect additional real-world data (if possible). This question can be answered by the sample-size sensitivities of the input models. See Section 4.2 for details.

Song and Nelson (2014) also propose a heuristic way to adjust the nominal CI using $\hat{\gamma}$ to include the input uncertainty σ_I^2 . Roughly, the adjusted $100(1 - \alpha)\%$ CI is given as

$$\bar{Y}(\hat{\mathbf{F}}) \pm q_{\alpha/2} \frac{\hat{\sigma}}{\sqrt{n}} \sqrt{1 + \hat{\gamma}^2}, \quad (13)$$

where $q_{\alpha/2}$ is the $100(\alpha/2)$ -th quantile of the standard normal distribution.

Even if $\hat{\gamma}$ is large, it may be infeasible to collect additional data from any of the input processes (for instance, the system no longer exists). In this case, Song and Nelson (2014) suggest to improve the estimator from the nominal experiment by combining it with the replications obtained from the diagnostic experiment. The combined estimator \tilde{Y} is

$$\tilde{Y} = \omega \bar{Y}(\hat{\mathbf{F}}) + (1 - \omega) \bar{\bar{Y}}(\hat{\mathbf{F}}),$$

where $\bar{\bar{Y}}(\hat{\mathbf{F}}) = \sum_{b=1}^B \bar{Y}(\hat{\mathbf{F}}^{*(b)})/B$ and $\omega \in [0, 1]$. The optimal weight ω^* given $\hat{\mathbf{F}}$ can be obtained by minimizing the conditional mean squared error (MSE) of \tilde{Y} given $\hat{\mathbf{F}}$; an estimator of the optimal weight is

$$\hat{\omega}^* = \frac{(\hat{\mathbf{b}}^*)^2 + \hat{\sigma}_I^2/B + \hat{\sigma}^2/BR}{(\hat{\mathbf{b}}^*)^2 + \hat{\sigma}_I^2/B + \hat{\sigma}^2/BR + \hat{\sigma}^2/n},$$

where $\hat{\mathbf{b}}^* = \bar{\bar{Y}}(\hat{\mathbf{F}}) - \bar{Y}(\hat{\mathbf{F}})$ is an estimator of bias. By combining $\bar{Y}(\hat{\mathbf{F}})$ and $\bar{\bar{Y}}(\hat{\mathbf{F}})$, the bias is increased as bootstrapped distributions add more bias to the estimator $\bar{Y}(\hat{\mathbf{F}})$, but the variance is decreased because we have more replications. The optimal weight ω^* balances these two.

When input uncertainty is significant, and it is possible to collect additional real-world data on the most offending input models, then a follow-up experiment is performed using the refined input models formed

from $\mathbf{m}' = \{m'_1, m'_2, \dots, m'_L\}$ observations, where $m'_\ell \geq m_\ell$. Song and Nelson (2014) investigated whether it is worthwhile to combine simulation results from the nominal and follow-up experiments and concluded that typically it is not. Because of the increased real-world sample, $\bar{Y}(\hat{\mathbf{F}}_{\mathbf{m}'})$ is less biased than $\bar{Y}(\hat{\mathbf{F}}_{\mathbf{m}})$. In addition, the variance of $\bar{Y}(\hat{\mathbf{F}}_{\mathbf{m}'})$ is smaller as it has less input uncertainty. Therefore, combining the two estimators is unlikely to reduce the MSE.

4.2 Illustration: Remote Order-taking System

In this section, we summarize an illustration in Song and Nelson (2014)—borrowed from Nelson (2013)—to describe a practical use of the three-phase experiments. All results in this illustration are performed using Simio and its input uncertainty analysis add-in that implements the diagnostic experiment procedure in Section 4.1.

The problem is to evaluate replacing the current order-taking system at the drive-through window of a chain of fast-food restaurants by a remote order-taking system. In the current system, there is a server at each store who takes orders from the incoming customers when they reach the order board. Under the proposed system, the server at the store is replaced by a pool of agents at a remote call center. The hope is that many fewer agents are needed than one per store. The fast food chain has high standards for service quality and in particular wants the average time that a customer waits until being greeted by an agent to be less than 1.5 seconds. A simulation study of this new system using the data from 7 stores is undertaken.

Table 1 shows 9 input processes used in the simulation: inter-arrival times of customers at 7 stores; order-taking times for the remote agents; and the time for a car to pull up to the order board after the previous car (if there is one) departs. The inter-arrival times were collected during the busiest 3-hour period over 10 days. With these data, Song and Nelson (2014) fit parametric distributions to perform the nominal experiment. Although these “real-world” data were actually generated from exponential distributions, they did not take advantage of knowing the families of distributions or parameters. The fitted distributions were exponentials, Weibulls, and Gammas. To obtain an estimator of the mean customer waiting time, they performed 1,000 replications of a steady-state simulation adopting a replication-deletion experiment design. The results showed that with 4 agents the 95% CI for mean customer waiting time was 0.99 ± 0.04 seconds, which is clearly less than 1.5 seconds. However, as mentioned earlier, this nominal CI does not take into account the input uncertainty. Therefore, to make a more robust management decision, one can conduct a diagnostic experiment.

Given the budget $N = 10,000$ for the diagnostic experiment, $(B = 80, R = 125)$ is chosen according to (12) where $\delta = 0.01$ is used. The diagnostic experiment gave $\hat{\gamma} = 17.48$, indicating that there is significant input uncertainty in the simulation estimator from the nominal experiment. The heuristic adjustment (13) expands the CI half width to $0.04 \times \sqrt{1 + 17.48^2} \approx 0.63$. The resulting CI $(0.36, 1.62)$ contains the critical value, 1.5 seconds.

To further investigate which input model contributes the most to input uncertainty, Table 1 shows the estimated relative contributions and sample-size sensitivities of input models using (10). Notice that the contributions are normalized to sum to 1 and sensitivities are scaled relative to the most sensitive input model (i.e., order taking time). Based on these results it makes sense to collect additional order-taking times to reduce input uncertainty, since it has the largest (most negative) sensitivity.

To see the effect of having additional observations, Song and Nelson (2014) conducted follow-up/diagnostic experiments with increasing numbers of observations of order-taking time (the “observations” are generated from an exponential distribution with mean 90 seconds). Figure 1 shows the point estimate and adjusted CI. The trend in $\hat{\gamma}$, indicated by \bullet , and the adjusted 95% CIs, are presented for different total amounts of data. The number of simulation replications for the follow-up experiments and the number of agents at the remote order-taking system remain $n = 1,000$ and 4, respectively. We observe that $\hat{\gamma}$ decreases as the amount of real-world data increases and it also becomes more clear that the adjusted CI includes 1.5. Therefore, one might consider increasing the number of agents from 4 to 5, which would be a different final decision than what would have been made by performing the nominal experiment only.

Table 1: Scaled contribution and sample-size sensitivity results for the remote order taking system.

ℓ	Input data	m_ℓ	Average	Contribution	Sensitivity
1	Interarrival 1	351	307.0	0.014	-0.007
2	Interarrival 2	153	701.3	0.011	-0.012
3	Interarrival 3	421	256.1	0.021	-0.008
4	Interarrival 4	261	413.0	0.008	-0.005
5	Interarrival 5	342	308.9	0.018	-0.008
6	Interarrival 6	354	304.4	0.006	-0.003
7	Interarrival 7	472	228.5	0.006	-0.002
8	Order taking	150	84.8	0.914	-1.000
9	Moving	70	5.4	0.003	-0.006

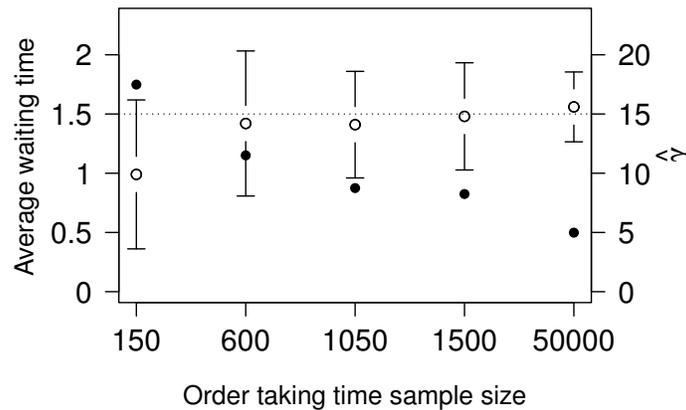


Figure 1: The adjusted confidence interval for expected customer waiting time along with $\hat{\gamma}$ (indicated by ●) for different real-world sample sizes for the order-taking time.

5 CONCLUSIONS

In this tutorial, we described the input uncertainty problem and investigated different measures of input uncertainty. See Barton (2012) for a more comprehensive list of the relevant literature. We put particular emphasis on contribution and sensitivity measures proposed by Song and Nelson (2014) that can be used to decide from which input processes to collect more data. While useful, challenges arise when we have multivariate input models where two or more input random variables are correlated; see Biller and Corlu (2011). Also, questions remain as to how to quantify input uncertainty when we have time series input models or input model parameters that vary over time (e.g., a non-homogeneous Poisson process).

Although input uncertainty quantification for point estimation has been an active area of research, little has been done on the implications of input uncertainty for other analysis problems. An exception is Corlu and Biller (2013) who consider a subset selection problem in the presence of input uncertainty. We expect that many “solved” problems in stochastic simulation output analysis will be revisited taking input uncertainty into account.

ACKNOWLEDGMENTS

This tutorial is based upon work supported by the National Science Foundation under Grant No. CMMI-1068473.

REFERENCES

- Ankenman, B. E. and B. L. Nelson. 2012. "A Quick Assessment of Input Uncertainty." In *Proceedings of the 2012 Winter Simulation Conference*, edited by C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose and A. M. Uhrmacher, 241–250. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Barton, R. R. 2012. "Tutorial: Input "Uncertainty in Output Analysis." In *Proceedings of the 2012 Winter Simulation Conference*, edited by C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A. M. Uhrmacher, 67–78. 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* 26:74–87.
- 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.
- Biller, B. and C. G. Corlu. 2011. "Accounting for Parameter Uncertainty in Large-scale Stochastic Simulations with Correlated Inputs." *Operations Research* 59:661–673.
- Cheng, R. C. H. and W. Holland. 1998. "Two-point Methods for Assessing Variability in Simulation Output." *Journal of Statistical Computation and Simulation* 60:183–205.
- Cheng, R. C. H. and W. Holland. 2004. "Calculation of Confidence Intervals for Simulation Output." *ACM Transactions on Modeling and Computer Simulation* 14:344–362.
- Corlu, C. G. and B. Biller. 2013. "A Subset Selection Procedure under Input Parameter Uncertainty." In *Proceedings of the 2013 Winter Simulation Conference*, edited by R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, 464–473. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Freimer, M. and L. W. Schruben. 2002. "Collecting Data and Estimating Parameters for Input Distributions." In *Proceedings of the 2002 Winter Simulation Conference*, edited by E. Yücesan, C. Chen, J. L. Snowdon, and J. M. Charnes, 392–399. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Homma, T. and A. Saltelli. 1996. "Importance Measures in Global Sensitivity Analysis of Model Output." *Reliability Engineering and System Safety* 52:1–17.
- Nelson, B. L. 2013. *Foundations and Methods of Stochastic Simulation: A First Course*. New York: Springer.
- Ng, S. H. and S. E. Chick. 2001. "Reducing Input Parameter Uncertainty for Simulations." In *Proceedings of the 2001 Winter Simulation Conference*, edited by B. A. Peters, J. S. Smith, D. J. Medeiros, and M. W. Rohrer, 364–371. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Ng, S. H. and S. E. Chick. 2006. "Reducing Parameter Uncertainty for Stochastic Systems." *ACM Transactions on Modeling and Computer Simulation* 16:26–51.
- Oakley, J. E. and A. O'Hagan. 2004. "Probabilistic Sensitivity Analysis of Complex Models: A Bayesian Approach." *Journal of the Royal Statistical Society* 66:751–769.
- Plischke, E., E. Borgonovo, and C. L. Smith. 2013. "Global Sensitivity Measures from Given Data." *European Journal of Operations Research* 226:536–550.

- Song, E., and B. L. Nelson. 2013. "A Quicker Assessment of Input Uncertainty." In *Proceedings of the 2013 Winter Simulation Conference*, edited by R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, 474–485. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Song, E. and B. L. Nelson. 2014. "Quickly Assessing Contributions to Input Uncertainty." Working Paper, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL, USA.
- Wagner, H. M. 1995. "Global Sensitivity Analysis." *Operations Research* 43:948–969.
- Xie, W., B. L. Nelson, and R. R. Barton. 2014. "A Bayesian Framework for Quantifying Uncertainty in Stochastic Simulation." Working Paper, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL, USA.
- Zouaoui, F. and J. R. Wilson. 2003. "Accounting for Parameter Uncertainty in Simulation Input Modeling." *IIE Transactions* 35:781–792.
- Zouaoui, F. and J. R. Wilson. 2004. "Accounting for Input-model and Input-parameter Uncertainties in Simulation." *IIE Transactions* 36:1135–1151.

AUTHOR BIOGRAPHIES

EUNHYE SONG is a Ph.D. candidate of the Department of Industrial Engineering and Management Sciences at Northwestern University. Her research interests are input uncertainty quantification, design of simulation experiments and simulation optimization. Her e-mail address is eunhyesong2016@u.northwestern.edu.

BARRY L. NELSON is the Walter P. Murphy Professor and Chair of the Department of Industrial Engineering and Management Sciences at Northwestern University. He is a Fellow of INFORMS and IIE. His research centers on the design and analysis of computer simulation experiments on models of stochastic systems, and he is the author of *Foundations and Methods of Stochastic Simulation: A First Course*, from Springer. His e-mail address is nelsonb@northwestern.edu.

C. DENNIS PEGDEN is the founder and CEO of Simio LLC. He was the founder and CEO of Systems Modeling Corporation, now part of Rockwell Software, Inc. He has held faculty positions at the University of Alabama in Huntsville and The Pennsylvania State University. He led in the development of the SLAM, SIMAN, Arena, and Simio simulation tools. He is the author/co-author of three textbooks in simulation and has published papers in a number of fields including mathematical programming, queuing, computer arithmetic, scheduling, and simulation. His email is cdpegden@simio.com.