

A QUICKER ASSESSMENT OF INPUT UNCERTAINTY

Eunhye Song
Barry L. Nelson

Department of Industrial Engineering & Management Sciences
Northwestern University
Evanston, IL 60208 USA

ABSTRACT

“Input uncertainty” refers the effect of driving a simulation with input distributions that are based on real-world data. At WSC 2012, Ankenman and Nelson presented a quick-and-easy experiment to assess the overall effect of input uncertainty on simulation output. When their method reveals that input uncertainty is substantial, then the natural follow-up questions are which input distributions contribute the most to input uncertainty, and from which input processes would it be most beneficial to collect more data? To answer these questions Ankenman and Nelson proposed a sequence of additional experiments that are in no sense “quick.” In this paper we provide a follow-up analysis that requires *no additional simulation experiments* beyond the overall assessment, and yet provides more information than Ankenman and Nelson. Numerical illustrations are provided.

1 INTRODUCTION

There is increasing recognition of the error in simulation results due to the input models that drive the simulation. In this context “input models” are fully specified probability distributions that represent uncertainty in the system of interest at its most basic level. Realizations from these input models combine with the simulation logic to produce output performance measures, and it is properties of these outputs (e.g., expected or mean value) that are used to infer what the real-world system performance would be.

Every simulation language measures the statistical error due to sampling from the input models, typically via confidence intervals on the performance properties. However, these confidence intervals do not account for the possible (in fact, likely) misspecification of the input models when they are estimated from real-world data. As shown in many papers (e.g, Barton 2012; Barton et al. 2013 Cheng and Holland 1998; Cheng and Holland 2004; Chick 2001; Freimer and Schruben 2002; Zouaoui and Wilson 2003; and Zouaoui and Wilson 2004) the error due to input uncertainty can overwhelm the simulation sampling error.

Ankenman and Nelson (2012) presented a quick-and-easy bootstrap experiment to assess the overall effect of input uncertainty relative to simulation sampling variability; we review their method below. Unfortunately, their method for identifying which input models contribute the most to input uncertainty requires a sequence of follow-up experiments; in the worst case it requires as many experiments as there are input models. In this paper we provide a new follow-up analysis that requires *no additional simulation experiments* beyond the one to assess the overall effect of input uncertainty. This analysis reveals the relative contributions of each input model and the sensitivity of performance-estimator variance to increasing the quantity of real-world data. Using these results, the analyst can decide whether it is worth the time and expense to collect additional data, and on which input processes to do so.

The paper is organized as follows. Section 2 describes the standard simulation experiment that ignores input uncertainty, followed by a review of the method for assessing input uncertainty of Ankenman and

Nelson (2012) in Section 3. Sections 4–6 introduce our new method. An empirical evaluation is provided in Section 7, and we end with conclusions.

2 THE NOMINAL EXPERIMENT

Suppose that our simulation is driven by L mutually independent input processes characterized by a collection of marginal distributions $\mathbf{F} = \{F_1, F_2, \dots, F_L\}$. To have an example in mind \mathbf{F} could be the interarrival-time and service-time distributions in a queueing simulation. Unknown to us, the true, correct “real-world” distributions are $\mathbf{F}^c = \{F_1^c, F_2^c, \dots, F_L^c\}$. To infer the ℓ th distribution we have a sample of real-world data

$$X_{\ell 1}, X_{\ell 2}, \dots, X_{\ell m_\ell} \sim \text{i.i.d. } F_\ell^c$$

for $\ell = 1, 2, \dots, L$. Let $\widehat{\mathbf{F}} = \{\widehat{F}_1, \widehat{F}_2, \dots, \widehat{F}_L\}$ be the collection of fitted distributions.

Given a collection of input distributions $\widehat{\mathbf{F}}$, the simulation generates performance output $Y_j(\widehat{\mathbf{F}})$ on replication $j = 1, 2, \dots, n$. Our objective in running the simulation is to estimate $E[Y(\mathbf{F}^c)]$ using the sample mean $\bar{Y}(\widehat{\mathbf{F}}) = \sum_{j=1}^n Y_j(\widehat{\mathbf{F}})/n$. *Our goal in this paper is to assess the relative impact of \widehat{F}_ℓ on $\text{Var}[\bar{Y}(\widehat{\mathbf{F}})]$.*

The usual assessment of simulation error (e.g., a confidence interval) captures only $\text{Var}[\bar{Y}(\widehat{\mathbf{F}}) \mid \widehat{\mathbf{F}}]$, the variance of the estimator *given* the input distributions.

3 A QUICK REVIEW OF THE QUICK ASSESSMENT

To simplify the review, suppose that the simulation has a single input distribution on which we have an i.i.d. sample of real-world data X_1, X_2, \dots, X_m . Ankenman and Nelson (2012) represented the output of the simulation on replication j , implemented using estimated input distribution \widehat{F} , as

$$Y_j(\widehat{F}) = E[Y(\widehat{F}) \mid \widehat{F}] + \varepsilon_j \tag{1}$$

where $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ are i.i.d. with mean 0 and variance σ^2 representing the natural variability from replication to replication in the simulation. They then provided a confidence interval for $\gamma = \sqrt{n}\sigma_I/\sigma$, where $\sigma_I^2 = \text{Var}\{E[Y(\widehat{F}) \mid \widehat{F}]\}$ is the variability in the expected value due to the possibility of having observed a different real-world sample of data from F^c . Thus, γ is a measure of how large input uncertainty is relative to the standard error of $\bar{Y}(\widehat{F})$ given \widehat{F} . Their CI is formed using bootstrapping:

Algorithm Quick

1. Given real-world data $\{X_1, X_2, \dots, X_m\}$, do the following:
 2. For b from 1 to B
 - (a) Generate bootstrap sample $X_{b1}^*, X_{b2}^*, \dots, X_{bm}^*$ by sampling m times with replacement from $\{X_1, X_2, \dots, X_m\}$.
 - (b) Fit \widehat{F}_b^* to $X_{b1}^*, X_{b2}^*, \dots, X_{bm}^*$.
(If there is more than one input model, do Steps 2a and 2b for each one.)
 - (c) Simulate R replications $Y_j(\widehat{F}_b^*), j = 1, 2, \dots, R$ using input model(s) \widehat{F}_b^* .
3. Report the point estimate and CI for $\gamma = \sqrt{n}\sigma_I/\sigma$ using the formulas in Ankenman and Nelson (2012).

This experiment is quick and easy to conduct. Ankenman and Nelson (2012) show that the optimal allocation of N total replications to B bootstraps of R replications each so as to minimize the expected width of the CI for γ is $B \approx 10$ and $R = N/B$. They also described an *additional* collection of experiments, using a version of sequential bifurcation (Wan et al. 2006), to identify which individual input distributions

are the biggest contributors to input uncertainty. They did this by decomposing the variance as

$$\text{Var} \left[Y_j(\widehat{\mathbf{F}}) \right] = \sigma_I^2 + \sigma^2 = \sum_{\ell=1}^L \sigma_\ell^2 + \sigma^2. \tag{2}$$

In this paper we propose a more robust method to identify the relative contributions of each distribution using a more rigorously justified decomposition and the same bootstrap data generated to estimate γ .

We will measure relative contribution in two ways: The *relative contribution to variance* of the ℓ th input model is

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

where

$$V_\ell(m_\ell) = \text{Var} \left\{ \text{E} \left[Y(F_1^c, F_2^c, \dots, F_{\ell-1}^c, \widehat{F}_\ell, F_{\ell+1}^c, \dots, F_L^c) \mid \widehat{F}_\ell \right] \right\} \tag{3}$$

is the variance due to input uncertainty if *only* input distribution ℓ was estimated. We use the notation $V_\ell(m_\ell)$ to indicate that the variance contribution of the ℓ th input model depends on the quantity of real-world data. The second measure is the *sample-size sensitivity* with respect to input distribution ℓ

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

Clearly the sample size m_ℓ must be integer, so treating it as continuous-valued and taking a derivative is a convenient approximation for the marginal impact of an additional real-world observation from distribution F_ℓ^c . As we have more samples from \widehat{F}_ℓ the variance contribution of ℓ th input model decreases, hence, the sample-size sensitivity is always negative; the more negative the sensitivity is, the more sensitive the variance is to additional real-world samples. We will scale the sensitivities so that the sensitivity of the input distribution with the most negative sensitivity is -1 .

4 A NEW INPUT-UNCERTAINTY MODEL

For any distribution G let the functionals $\mu(\cdot)$ and $\sigma^2(\cdot)$ be defined as

$$\begin{aligned} \mu(G) &= \text{E}_G(X) \\ \sigma^2(G) &= \text{Var}_G(X) \end{aligned}$$

where $X \sim G$. Rather than decomposing the output variance directly, we first model the relationship between the input models and the simulation output itself as

$$Y_j(\mathbf{F}) = \beta_0 + \sum_{\ell=1}^L \beta_\ell \mu(F_\ell) + \sum_{\ell=1}^L v_\ell \sigma^2(F_\ell) + \varepsilon_j \tag{4}$$

where $\beta_0, \beta_1, \dots, \beta_L$ and v_1, v_2, \dots, v_L are constants, and $\varepsilon_1, \varepsilon_2, \dots$ are i.i.d. $(0, \sigma^2)$ random variables. Model (4) approximates the impact of the ℓ th input model through its mean and variance. This is clearly a first-order approximation, since we include no interactions among the input distributions. However, it does reflect the situation in which both where the input distribution is centered, and how much it varies around that center, have an impact. The variability of the simulation output is represented by ε , whose variance does not depend on \mathbf{F} , also clearly an approximation. Ng and Chick (2001, 2006) used a related model that is a function of the natural parameters of the input distributions; we show the value of parameterizing (4) using the mean and variance of each distribution below.

Under Model (4), we can easily identify the input-uncertainty variance contribution (3) because

$$\begin{aligned} \text{Var} \left[\bar{Y}(\hat{\mathbf{F}}) \right] &= \sum_{\ell=1}^L \left\{ \beta_{\ell}^2 \text{Var} \left[\mu(\hat{F}_{\ell}) \right] + v_{\ell}^2 \text{Var} \left[\sigma^2(\hat{F}_{\ell}) \right] + 2\beta_{\ell}v_{\ell} \text{Cov} \left[\mu(\hat{F}_{\ell}), \sigma^2(\hat{F}_{\ell}) \right] \right\} + \frac{\sigma^2}{n} \quad (5) \\ &= \sum_{\ell=1}^L v_{\ell}(m_{\ell}) + \frac{\sigma^2}{n} \end{aligned}$$

remembering that $\hat{F}_1, \hat{F}_2, \dots, \hat{F}_L$ are independent. This is our variance decomposition.

Cheng and Holland (1998) proposed a method for measuring the impact of input parameter uncertainty when parametric input distributions and maximum likelihood estimators of those parameters are used. Their expressions (9)–(10) for the variance of \bar{Y} reduce to (5) if each distribution has two parameters $(\mu(F_{\ell}), \sigma^2(F_{\ell}))$ —in other words the distributions are parameterized by their means and variances. *One way to describe our contribution is that we provide a much easier way to estimate the terms in this model without requiring parametric input distributions or gradient estimates.*

5 ASSESSING THE CONTRIBUTION

To use Expression (5) we need to estimate the model parameters β_{ℓ} and v_{ℓ} and obtain values for $\text{Var} \left[\mu(\hat{F}_{\ell}) \right]$, $\text{Var} \left[\sigma^2(\hat{F}_{\ell}) \right]$ and $\text{Cov} \left[\mu(\hat{F}_{\ell}), \sigma^2(\hat{F}_{\ell}) \right]$ to plug into it. To do this we will take a bootstrapping approach: We pretend that $\hat{\mathbf{F}} = \mathbf{F}^c$, where now $\hat{\mathbf{F}}$ is the collection of empirical cdfs (ecdfs) of the real-world data. Notice that the nominal experiment to estimate $E[Y(\mathbf{F}^c)]$ can be executed with any sort of fitted distributions that the analyst desires; we use the ecdfs here only to assess input uncertainty.

To keep the distinction clear, let $X_{\ell 1}^*, X_{\ell 2}^*, \dots, X_{\ell m_{\ell}}^*$ represent an i.i.d. sample from the ecdf of the real-world data \hat{F}_{ℓ} (which we treat as fixed); this yields its own ecdf \hat{F}_{ℓ}^* . If we pretend $\hat{\mathbf{F}} = \mathbf{F}^c$, then $X_{\ell 1}^*, X_{\ell 2}^*, \dots, X_{\ell m_{\ell}}^*$ corresponds to a real-world sample from the ℓ th distribution, and \hat{F}_{ℓ}^* corresponds to the fitted distribution that would be used in the simulation. Bootstrapping allows us to simulate the effect of taking B “real-world” samples, which obviously we cannot do in reality.

The first useful insight from this approach is that the terms $\text{Var} \left[\mu(\hat{F}_{\ell}^*) \right]$, $\text{Var} \left[\sigma^2(\hat{F}_{\ell}^*) \right]$ and $\text{Cov} \left[\mu(\hat{F}_{\ell}^*), \sigma^2(\hat{F}_{\ell}^*) \right]$ are directly computable from $\hat{\mathbf{F}}$. They are

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

where M_{ℓ}^k is k th central moment of \hat{F}_{ℓ} : $M_{\ell}^k = m_{\ell}^{-1} \sum_{i=1}^{m_{\ell}} (X_{\ell i} - \bar{X}_{\ell})^k$. *This is one of the primary advantages of using the ecdfs and formulating Model (4) as a function of the mean and variance of each input distribution.*

The second useful insight is that we can fit Model (4) using the data from Ankenman and Nelson’s overall assessment reviewed in Section 3. That is, we fit (4) via least-squares regression using the data

$$Y_j(\hat{\mathbf{F}}_b^*), \left\{ \mu(\hat{F}_{\ell b}^*), \sigma^2(\hat{F}_{\ell b}^*), \ell = 1, 2, \dots, L \right\}$$

for $b = 1, 2, \dots, B$ bootstraps from $\hat{\mathbf{F}}$ and $j = 1, 2, \dots, R$ replications of the simulation per bootstrap.

Therefore, our estimate of the variance contribution from the ℓ th input model is

$$\widehat{V}_\ell(m_\ell) = \widehat{\beta}_\ell^2 \text{Var} \left[\mu(\widehat{F}_\ell^*) \right] + \widehat{v}_\ell^2 \text{Var} \left[\sigma^2(\widehat{F}_\ell^*) \right] + 2\widehat{\beta}_\ell \widehat{v}_\ell \text{Cov} \left[\mu(\widehat{F}_\ell^*), \sigma^2(\widehat{F}_\ell^*) \right] \quad (6)$$

which is easily differentiated with respect to m_ℓ to obtain sensitivities. This also provides an alternative to Ankenman and Nelson’s estimator $\widehat{\gamma}$ of $\gamma = \sqrt{n}\sigma_I/\sigma$: $\widehat{\gamma} = \sqrt{n}\widehat{\sigma}_I/\widehat{\sigma}$ where

$$\widehat{\sigma}_I^2 = \sum_{\ell=1}^L \widehat{V}_\ell(m_\ell)$$

and $\widehat{\sigma}^2$ is the residual variance from fitting the regression Model (4).

6 EXPERIMENT DESIGN

There are three experiment design questions that need to be addressed:

1. Our design points for fitting (4) are $\widehat{\mathbf{F}}_1^*, \widehat{\mathbf{F}}_2^*, \dots, \widehat{\mathbf{F}}_B^*$, which are randomly generated instead of being “designed.” Why not used a designed experiment? There are at least two reasons:
 - (a) In the spirit of a “quick” assessment, random sampling allows us to use the results for estimating $\sqrt{n}\sigma_I/\sigma$ to also estimate V_ℓ .
 - (b) The design points $\widehat{\mathbf{F}}_1^*, \widehat{\mathbf{F}}_2^*, \dots, \widehat{\mathbf{F}}_B^*$ will naturally be clustered in that part of the distribution space that we need to model well: the most likely input distribution deviations from $\widehat{\mathbf{F}}$. This space is not one that we could know a priori. We do not expect, nor need, Model (4) to provide a good global fit; we just need it to be locally correct near $\widehat{\mathbf{F}}$.
2. The bootstrap design points $\widehat{\mathbf{F}}_1^*, \widehat{\mathbf{F}}_2^*, \dots, \widehat{\mathbf{F}}_B^*$ must be sampled independently from $\widehat{\mathbf{F}}$, but should common random numbers (CRN) be assigned to the simulations conducted at each $\widehat{\mathbf{F}}_b^*$? Kleijnen (1988) and others have shown that in least-squares regression CRN tends to reduce the variance of the slope-parameter estimators, which are $\widehat{\beta}_\ell, \widehat{v}_\ell, \ell = 1, 2, \dots, L$ in Model (4). Since the variance of \widehat{V}_ℓ is an increasing function of the variances of $\widehat{\beta}_\ell$ and \widehat{v}_ℓ , it seems clear that using CRN is desirable.
3. Given a budget of N simulation replications, how should it be divided between bootstrap samples (B) and simulation replications per bootstrap (R) so that $RB = N$? Ankenman and Nelson (2012) showed that if N is not too small, then $B \approx 10$ is optimal in terms of minimizing the expected width of the CI for $\sqrt{n}\sigma_I/\sigma$. Is this still a good design choice if the goal is to estimate V_ℓ as well? We address this question below.

For the purpose of analyzing design question 3, we simplify Model (4) to only include the “mean effects” and no common random numbers:

$$Y_j(\widehat{\mathbf{F}}) = \beta_0 + \sum_{\ell=1}^L \beta_\ell \mu(\widehat{F}_\ell) + \varepsilon_j. \quad (7)$$

If we assume that both $\mu(\widehat{F}_\ell)$ and ε_j are normally distributed—which is plausible since $\mu(\widehat{F}_\ell)$ is a sample mean when \widehat{F}_ℓ is the ecdf—and force $RB = N$, then we can show that

$$\text{Var} \left[\widehat{\beta}_\ell \right] = \frac{1}{N} \left(\frac{B}{B-L-2} \right) \frac{\sigma^2}{\text{Var} \left[\mu(\widehat{F}_\ell) \right]}. \quad (8)$$

If we extend this in the natural way to include the “variance effects” as in Model (4), then the L term in the denominator of (8) becomes $2L$, reflecting the incorporation of both mean and variance terms for each input

model. This implies that $B > 2L + 2$ is the minimum number of bootstraps required, and $(B = N, R = 1)$ is optimal in the sense of minimizing the variance of $\hat{\beta}_\ell$ and \hat{v}_ℓ , in turn minimizing the variance of \hat{V}_ℓ . This differs from the recommendation of Ankenman and Nelson (2012) that has B small and R large.

How do we balance these two competing objectives? Figure 1 shows one way to view this problem. We refer to the term $B/(B - 2L - 2)$ of (8) as the “contribution precision factor” on this graph, since it determines how precise the estimation of V_ℓ is. The smaller this factor, the more precise the estimation of V_ℓ , and it converges to 1 as $B \rightarrow \infty$. As Ankenman and Nelson (2012) pointed out, the expected length of the 95% confidence interval for the uncertainty ratio σ_I^2/σ^2 is minimized at $B = 11$. Also plotted is the relative increase from this minimum value of the expected length of the 95% CI for σ_I^2/σ^2 . The plot is for the specific case when there are $L = 5$ input models and $N = 10,000$ replications. Notice that $B = 11$ is not feasible since we need at least $B = 13$ to satisfy $B > 2 \times 5 + 2$.

We observe that the contribution precision factor rapidly converges to its minimum as B increases, whereas the expected length of the 95% CI for the uncertainty ratio increases as B increases in the feasible range of B . Hence, the experiment design involves a trade-off between precise estimation of the uncertainty ratio and that of the individual input model contributions. By using Figure 1 we can decide what B and R values to use depending on what is most important. The choice we made was to allow the expected CI length to increase by 50% relative to its minimum. In Figure 1 this corresponds to assigning $B = 50$ and $R = 200$, which implies a 32% increase in the contribution precision factor from its minimum.

7 EMPIRICAL RESULTS

This section summarizes an empirical study of the proposed method and also an illustration on a realistic problem.

We first evaluate the method of Section 5 using a well-known example, a stochastic activity network (SAN); see Figure 2. The goal is to estimate the mean time to complete the network. We defined a set of correct, real-world input distributions for the activity times $\mathbf{F}^c = \{F_1^c, F_2^c, \dots, F_5^c\}$ and then generated m_ℓ samples from each F_ℓ^c , which we treat as the “real-world” data. Experiments under different settings of sample size and mean and variance of the activity-time distributions were conducted and the variance contribution and sensitivity of each was estimated. To evaluate our method, we precisely estimated the true contributions of the input models from side experiments. These side experiments exploit the fact that the true, correct real-world distributions are known, which is obviously not the case in practice.

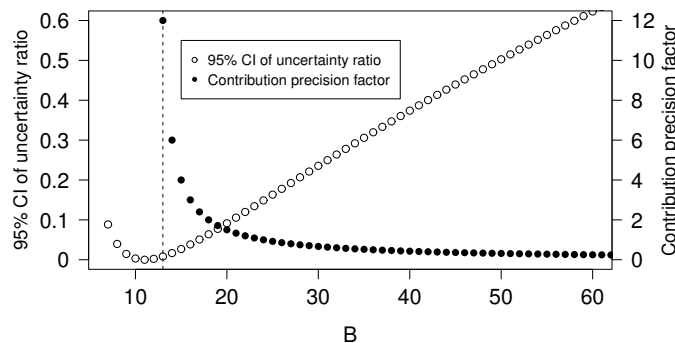


Figure 1: The trade-off between contribution precision factor and the expected width of the 95% CI for the uncertainty ratio as a function of number of bootstraps B when $L = 5$ input models and $N = 10,000$ replications.

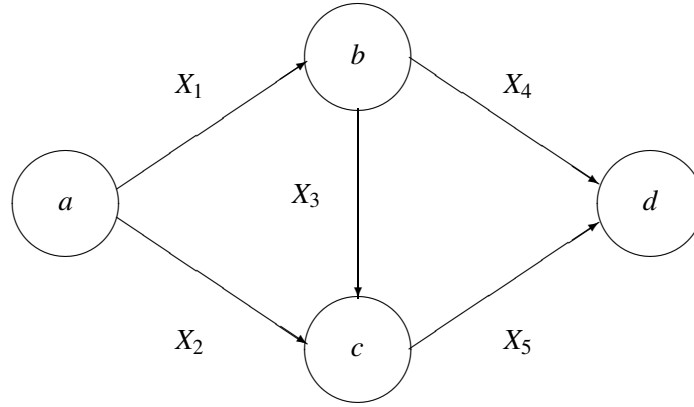


Figure 2: A small stochastic activity network.

We then introduce a more realistic example of simulating a remote order taking system to illustrate the practical application of our method.

7.1 SAN

For each activity time X_ℓ we have a correct real-world distribution F_ℓ^c which we pretend is unknown. The total time to finish the project is $Y = \max\{X_1 + X_4, X_1 + X_3 + X_5, X_2 + X_5\}$ and the purpose of the simulation is to estimate the expected value of Y . In the remainder of the section we describe empirical results for contribution and sensitivity of the input distributions under two experimental settings: (i) different real-world sample sizes, and (ii) different means and variances of the activities. We also compare two different designs for the follow-up experiment, $(B = 50, R = 200)$ and $(B = 10,000, R = 1)$. The latter design does not permit us to calculate the point estimate and confidence interval for $\gamma = \sqrt{n}\sigma_I/\sigma$ suggested by Ankenman and Nelson (2012) since $R = 1$. Instead, we report the point estimate $\tilde{\gamma}$ based on Model (4).

To evaluate how well our method measures the input-uncertainty variance contribution of each input model, a series of side experiments was conducted for each experimental setting:

Side Experiment

1. Given the true distributions $F_1^c, F_2^c, \dots, F_5^c$ for activity times X_1, X_2, \dots, X_5 , do the following:
2. For b from 1 to B
 - (a) Generate m_1 samples from F_1^c and construct ecdf \hat{F}_1 from the samples.
 - (b) Run R replications $Y_j(\hat{F}_1, F_2^c, \dots, F_5^c), j = 1, 2, \dots, R$ using the ecdf \hat{F}_1 and true distributions $F_2^c, F_3^c, \dots, F_5^c$ as the input models.
 - (c) Calculate sample mean $\bar{Y}_b(\hat{F}_1, F_2^c, \dots, F_5^c)$ from the replications.
3. Calculate the sample variance of $\bar{Y}_b(\hat{F}_1, F_2^c, \dots, F_5^c), b = 1, 2, \dots, B$

$$\frac{1}{B-1} \sum_{b=1}^B \left\{ \bar{Y}_b(\hat{F}_1, F_2^c, \dots, F_5^c) - \bar{\bar{Y}}(\hat{F}_1, F_2^c, \dots, F_5^c) \right\}^2 \quad (9)$$

to estimate the true input uncertainty contribution V_1 of X_1 , where

$$\bar{Y}(\hat{F}_1, F_2^c, \dots, F_5^c) = \frac{1}{B} \sum_{b=1}^B \bar{Y}_b(\hat{F}_1, F_2^c, \dots, F_5^c).$$

- Conduct Steps 2–3 for each distribution $F_2^c, F_3^c, \dots, F_5^c$ in turn.

A key point is that we make R large enough that

$$\text{Var} \left[\bar{Y}_b(\hat{F}_1, F_2^c, \dots, F_5^c) \mid \hat{F}_1 \right] \approx 0.$$

This implies that for each \hat{F}_1 , $\bar{Y}_b(\hat{F}_1, F_2^c, \dots, F_5^c) \approx E \left[\bar{Y}_b(\hat{F}_1, F_2^c, \dots, F_5^c) \mid \hat{F}_1 \right]$ and we can treat the sample variance (9) as an estimate of the variance due to different real-world samples; i.e., input uncertainty only. We used $(B = 100, R = 5,000)$ which gives relative errors of less than 2%. Notice that B and R in these side experiments have nothing to do with the choices we make for uncertainty analysis.

7.1.1 Exponential SAN

In this experiment setting all activity times have exponential real-world distributions with mean 1. Under this setting, the second path, $X_1 + X_3 + X_5$, is likely to be the longest. To see the effect of different real-world sample sizes on \hat{V}_ℓ , we conducted simulations for two cases: (a) $m_\ell = 100, \ell = 1, 2, \dots, 5$, and (b) $m_\ell = 100, \ell = 1, 2, 4, 5$, and $m_3 = 50$.

Figure 3 displays the simulation results when $(B = 50, R = 200)$ averaged over 1,000 macro replications to provide a relative error of less than 3%. Plotted are the relative contributions $\hat{V}_\ell(m_\ell) / \sum_{i=1}^L \hat{V}_i(m_i)$, and the sensitivities scaled so that the greatest sensitivity has value -1 . In case (a), X_1 and X_5 have the largest contributions and are more sensitive to additional real-world data since these two activity times are involved in two out of three paths on the network. On the other hand, X_2 and X_4 have the smallest contributions and sensitivities since they are only involved in one path that is likely to be dominated by $X_1 + X_3 + X_5$. The higher contribution/sensitivity of X_3 compared to X_2 and X_4 can be explained by the same reasoning. This trend changes slightly when m_3 decreases to 50 in case (b): since there are fewer real-world samples for X_3 in this case, X_3 shows the largest contribution and sensitivity.

Table 1 provides estimates of the uncertainty ratio γ for case (a) under two different designs for the follow-up experiment: $(B = 50, R = 200)$ and $(B = 10,000, R = 1)$. In the $(B = 50, R = 200)$ case, $\tilde{\gamma}$ is

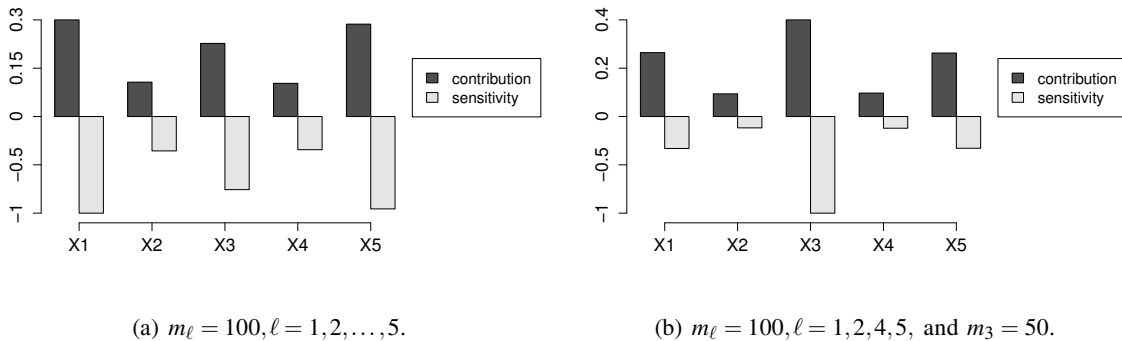


Figure 3: Scaled contributions (positive) and sensitivities (negative) of input models for the exponential SAN.

Table 1: Comparison of simulation results for case (a) when $n = 10,000$ under two experimental designs using 100 macro replications. “AN” refers to Ankenman and Nelson’s (2012) $\hat{\gamma}$ and corresponding CI = [LCL, UCL], while “SN” refers to our estimator $\tilde{\gamma}$.

(B, R)	Contribution					Uncertainty Ratio			
	X1	X2	X3	X4	X5	AN $\hat{\gamma}$	LCL	UCL	SN $\tilde{\gamma}$
(50, 200)									
average	0.00901	0.00334	0.00739	0.00276	0.00855	9.778	7.146	13.298	10.392
stderr	0.00045	0.00030	0.00043	0.00021	0.00047	0.145	0.138	0.167	0.115
(10,000, 1)									
average	0.00859	0.00290	0.00644	0.00312	0.00901	-	-	-	10.243
stderr	0.00036	0.00022	0.00039	0.00023	0.00041	-	-	-	0.092

contained in the confidence interval of Ankenman and Nelson (2012). Notice that the standard errors of the contribution and sensitivity estimates are not always reduced by using the $(B = 10,000, R = 1)$ design. This shows that using the optimal choice $(B = 10,000, R = 1)$ does not have much effect on the precision, and under this design we cannot provide the confidence interval for the uncertainty ratio. Therefore, we conclude that the gain from setting $B = N$ is not significant enough to give up providing the confidence interval for γ in this particular example.

Figure 4 compares the estimated input contributions from our method and the side experiments for cases (a) and (b). Notice that the contributions are not scaled in this graph. Clearly the estimated contributions from the two approaches are close, which implies Model (4) successfully represents the exponential SAN.

7.1.2 Gamma SAN

In this section, Gamma distributions are the true distributions for the activity times. Two sets of experiments were conducted to investigate the effect of different mean and variance values on the contribution of X_3 . In the first set of experiments, we fixed the variance of all activity times to 0.5 and set X_1, X_2, X_4, X_5 to have mean 1 and X_3 to have mean 10. In the second set, the means of all activities are fixed to 10, but X_1, X_2, X_4, X_5 have variance 1 while X_3 has variance 5. In all cases $m_\ell = 100$ “real-world” samples were collected for each activity time. For the follow-up experiment we used $(B = 50, R = 200)$ and the results are averaged over 1,000 macroreplications.

Figure 5 shows the experiment results. Observe that in the first set the trend in contribution/sensitivity is not much different from case (a) in Section 7.1.1. This might not seem intuitive since one might think

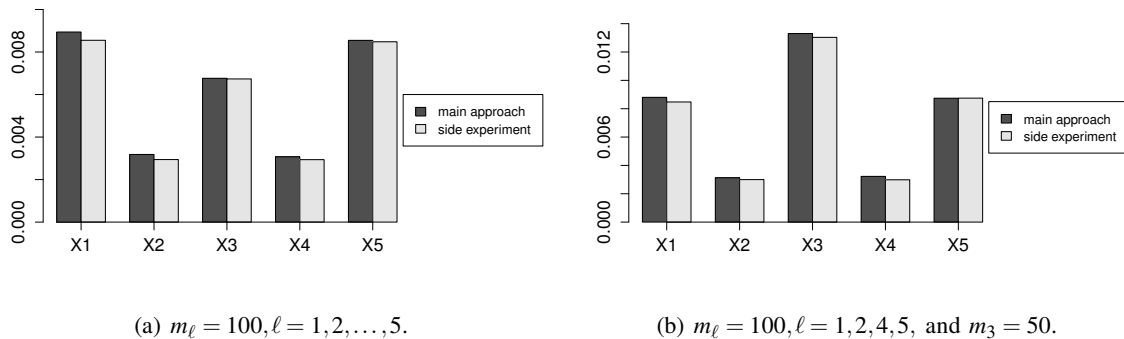
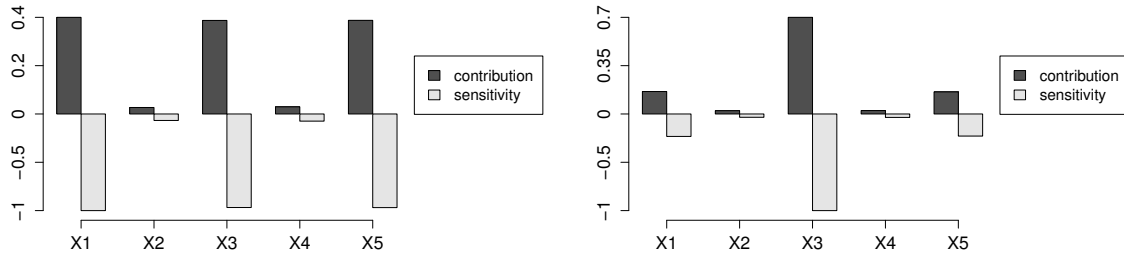


Figure 4: Comparison of estimated input model contributions and the true contributions for the exponential SAN.



(a) $X_\ell \sim \text{Gamma}(2, 2), \ell = 1, 2, 4, 5$ and $X_3 \sim \text{Gamma}(200, 20)$. (b) $X_\ell \sim \text{Gamma}(100, 10), \ell = 1, 2, 4, 5$ and $X_3 \sim \text{Gamma}(20, 2)$.

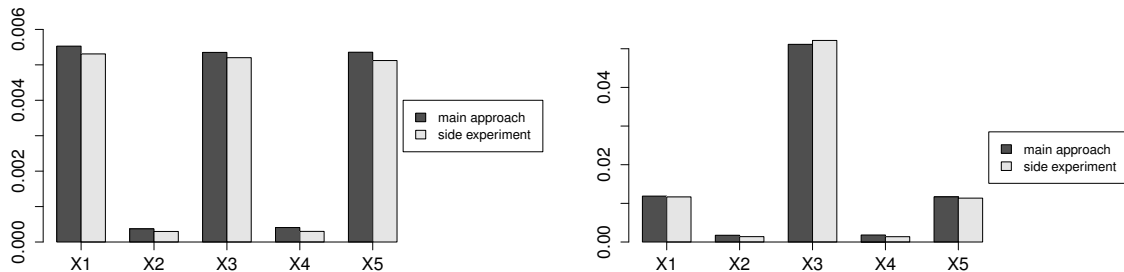
Figure 5: Scaled contributions (positive) and sensitivities (negative) of input models for the gamma SAN.

that the large mean value of X_3 would increase its contribution. However, a large mean for X_3 only makes the path $X_1 + X_3 + X_5$ more likely to be dominant and, therefore, the input uncertainty due to X_1 and X_5 still has a significant impact on output variability since they are also included in other paths. In the second set, X_3 has the largest contribution and is more sensitive to additional data collection, which can be explained by its relatively large variance.

Figure 6 presents the estimated input contributions from the side experiments and compares them to the results from our method. As in the exponential SAN case in Section 7.1.1, the two estimated values are close in both cases (a) and (b).

7.2 Illustration: Remote Order-Taking System

In this section we apply our method to a more realistic simulation. The problem, which is taken from Nelson (2013), is to evaluate replacing the current drive-through order windows for a chain of fast-food restaurants with the equivalent of a call center. The current store design has two fully staffed windows, one for order taking and another for food delivery. The proposal is to replace the first window with a remote order-taking service in which agents communicate with customers through the electronic order board and then relay the order to the store. The fast-food chain has high standards for customer service and requires that the average waiting time for a customer to be greeted by an agent once they reach the order board be



(a) $X_\ell \sim \text{Gamma}(2, 2), \ell = 1, 2, 4, 5$ and $X_3 \sim \text{Gamma}(200, 20)$. (b) $X_\ell \sim \text{Gamma}(100, 10), \ell = 1, 2, 4, 5$ and $X_3 \sim \text{Gamma}(20, 2)$.

Figure 6: Comparison of estimated input model contributions and the true contributions for the gamma SAN.

Table 2: Scaled contribution and sensitivity results for the remote order-taking system.

ℓ	Input data	m_ℓ	Average	Contribution	Sensitivity
1	Interarrival 1	351	307.0	0.01558	-0.00717
2	Interarrival 2	153	701.3	0.02079	-0.02195
3	Interarrival 3	421	256.1	0.01139	-0.00437
4	Interarrival 4	261	413.0	0.00691	-0.00428
5	Interarrival 5	342	308.9	0.00262	-0.00124
6	Interarrival 6	354	304.4	0.00868	-0.00396
7	Interarrival 7	472	228.5	0.00209	-0.00071
8	Order taking	150	84.8	0.92882	-1.00000
9	Moving	70	5.4	0.00312	-0.00719

less than 3 seconds. The proof-of-concept simulation uses data from 7 stores to investigate whether the call center can meet this standard with substantially fewer agents than the number of stores served.

There are 9 sets of real-world data: customer interarrival times from the 7 stores; order-taking times; and the time for a car to pull away from the order board and the next one in line (if there is one) to pull up to the order board. Interarrival times were collected during the busiest 3-hour period over 10 days. The numbers of observations of order-taking time and car moving time available were 150 and 70, respectively. To evaluate busy-period performance, steady-state simulations were conducted using a replication-deletion experiment design employing $n = 1,000$ replications.

To analyze input uncertainty and the contribution from each data source, we generated $B = 50$ bootstrap samples from each of the $L = 9$ real-world data sets and ran $R = 200$ replications for each, using the ecdfs from the bootstrap samples as the driving input distributions. Below we report results from the simulation with 4 agents, which seemed to be adequate to achieve the desired standard.

Using the method of Ankenman and Nelson (2012) a 95% CI for $\gamma = \sqrt{n}\sigma_I/\sigma$ is [15.5, 19.4]. Thus, input uncertainty is significantly larger than simulation uncertainty, so we are interested in which distributions are the largest contributors.

Table 2 displays the estimated scaled contribution and sensitivity of each input model, as well as its real-world sample size and its sample mean (in seconds). Clearly the input model for order-taking time makes the largest contribution to input uncertainty, and it is the most sensitive to collecting additional data. This makes sense: every customer requires an order-taking time, and the real-world sample size for this input is relatively small. If it is possible to collect additional input data, this is the distribution for which we would achieve the most benefit.

Notice that the rank order of contribution and sensitivity do not always coincide. For instance, the interarrival time distribution of the 4th store has a larger relative contribution, but smaller sensitivity, than the moving-time distribution. This can be explained by the real-world sample sizes: Although the contribution of the moving-time input model is smaller, the sample size is also smaller; hence, one additional real-world sample of moving time can reduce the variability in the simulation result by a relatively large amount.

8 CONCLUSIONS

In this paper we presented a method to quantify the overall impact of input-model uncertainty on simulation-based performance estimates, to identify which input models make the largest contribution to this uncertainty, and to identify the input data sources from which additional observations would lead to the greatest reduction in input uncertainty. Our approach builds on Ankenman and Nelson (2012) by obtaining the contribution and sensitivity results from the same experiment that they used to measure overall input uncertainty.

Open questions remain about the design of the follow-up experiment, and in particular the total budget N that should be expended on this experiment to obtain reliable results.

REFERENCES

- Ankenman, B. E. and B. L. Nelson. 2012. A quick assessment of input uncertainty. In C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose and A. M. Uhrmacher (Eds.), *Proceedings of the 2012 Winter Simulation Conference*. Piscataway, N.J.: Institute of Electrical and Electronics Engineers.
- Barton, R. R. 2012. Tutorial: Input uncertainty in output analysis. In C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose and A. M. Uhrmacher (Eds.), *Proceedings of the 2012 Winter Simulation Conference*. Piscataway, N.J.: Institute of Electrical and Electronics Engineers.
- Barton, R. R., B. L. Nelson and W. Xie. 2013. Quantifying input uncertainty via simulation confidence intervals. *INFORMS Journal on Computing*, in press.
- 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.
- Chick, S. E. 2001. Input distribution selection for simulation experiments: Accounting for input uncertainty. *Operations Research* 49: 744–758.
- 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 B. A. Peters, J. S. Smith, D. J. Medeiros, and M. W. Rohrer, (Eds.), *Proceedings of the 2001 Winter Simulation Conference*, 364–371. Piscataway, N.J.: Institute of Electrical and Electronics Engineers.
- Ng, S. H. and S. E. Chick. 2006. Reducing parameter uncertainty for stochastic systems. *ACM Transactions on Modeling and Computer Simulation* 16: 26–51.
- Freimer, M. and L. W. Schruben. 2002. Collecting data and estimating parameters for input distributions. In E. Yücesan, C. Chen, J. L. Snowdon, and J. M. Charnes (Eds.), *Proceedings of the 2002 Winter Simulation Conference*, 392–399. Piscataway, N.J.: Institute of Electrical and Electronics Engineers.
- Kleijnen, J. P. C. 1988. Analyzing simulation experiments with common random numbers. *Management Science* 34: 65–74.
- Wan, H., B. E. Ankenman and B. L. Nelson. 2006. Controlled sequential bifurcation: A new factor-screening method for discrete-event simulation. *Operations Research* 54: 743–755.
- 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.

ACKNOWLEDGMENTS

This research was partially supported by National Science Foundation Grant CMMI-1068473. We thank Russell Barton for helpful discussion.

AUTHOR BIOGRAPHIES

EUNHYE SONG is a Ph.D. student of the Department of Industrial Engineering and Management Sciences at Northwestern University. Her research interests are 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 and web addresses are nelsonb@northwestern.edu and www.iems.northwestern.edu/~nelsonb.