IISE Transactions

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uiie21

Meaningful sensitivities: A new family of simulation sensitivity measures

Xi Jiang, Barry L. Nelson & L. Jeff Hong

To cite this article: Xi Jiang, Barry L. Nelson & L. Jeff Hong (2021): Meaningful sensitivities: A new family of simulation sensitivity measures, IISE Transactions, DOI: 10.1080/24725854.2021.1931571

To link to this article: https://doi.org/10.1080/24725854.2021.1931571

IISE TRANSACTIONS
https://doi.org/10.1080/24725854.2021.1931571

Meaningful sensitivities: A new family of simulation sensitivity measures

Xi Jiang[a], Barry L. Nelson[b], and L. Jeff Hong[c]

[a]SAS Institute, Cary, NC, USA; [b]Department of Industrial Engineering & Management Sciences, Northwestern University, Evanston, IL, USA; [c]Department of Management Science, Fudan University, Shanghai, China

ABSTRACT

Sensitivity analysis quantifies how a model output responds to variations in its inputs. However, the following sensitivity question has never been rigorously answered: How sensitive is the mean or variance of a stochastic simulation output to the mean or variance of a stochastic input distribution? This question does not have a simple answer because there is often more than one way of changing the mean or variance of an input distribution, which leads to correspondingly different impacts on the simulation outputs. In this article we propose a new family of output-property-with-respect-to-input-property sensitivity measures for stochastic simulation. We focus on four useful members of this general family: sensitivity of output mean or variance with respect to input-distribution mean or variance. Based on problem-specific characteristics of the simulation we identify appropriate point and error estimators for these sensitivities that require no additional simulation effort beyond the nominal experiment. Two representative examples are provided to illustrate the family, estimators and interpretation of results.

ARTICLE HISTORY
Received 8 September 2020
Accepted 29 April 2021

KEYWORDS
Local sensitivity analysis; stochastic simulation

1. Introduction

Stochastic simulations can be regarded as functions mapping inputs into outputs. With very high confidence we know that a computer model is an imperfect representation of reality. The field of verification and validation provides best practices for good model building; see for instance Robinson (2014). The field of uncertainty quantification, on the other hand, attempts to deliver numerical measures of model sensitivity and risk that are actionable, as described below.

This article addresses uncertainty quantification of stochastic simulation models via local sensitivity analysis. The value of local sensitivity analysis to the analyst arises in at least three ways:

1. Establishing robustness, or lack of robustness, of the simulation-based estimates in the sense that small changes in some aspects of the model do not, or do, lead to large changes in the simulation outputs.
2. Establishing for which aspects of the model additional research to reduce uncertainty might yield the most benefit in improved model accuracy.
3. Establishing where management effort to alter the real system's inputs might provide the most impact.

The new local sensitivity measures introduced in this article can be used to address all three goals. The term "sensitivity analysis" begs the question: sensitive to what? In industrial engineering stochastic simulation the primary answers are: (i) sensitivity to structural aspects of the model that are directly controllable (e.g., number of servers in a queue); and (ii) sensitivity to the input probability distributions that make it a "stochastic simulation." The former is primarily addressed by design of experiments; the latter is the topic of this article.

Stochastic input distributions may be obtained from subjective judgment ("expert knowledge"), underlying process physics, or fit to historical data; in all cases they are subject to uncertainty that propagates (whether measured or not) through the simulation model to the outputs. That is exactly why the sensitivity to, say, the mean and variance of the input distributions is useful: it shows which measure of which input has the biggest influence on the output performance measure of interest. Further, sensitivity analysis can be used early in a modeling project to help determine the time or money to expend to collect real-world observations for stochastic inputs. Understanding the sensitivity of outputs to inputs facilitates addressing 1 to 3 above, which is why commercial simulation software products support measuring such sensitivity.

What we have in mind is to be able to make statements like the following for, say, a hypothetical semiconductor wafer fab simulation:

> When considering the variability of the steps in our fabrication process, our mean cassette cycle time is most sensitive to the variance of the Developer step, and therefore our improvement effort should be put on stabilizing the Developer time.

However, we want to do more than just ranking the input-distribution sensitivities; we also want to quantify how much reduction in, say, mean cycle time we should expect

CONTACT Barry L. Nelson nelsonb@northwestern.edu
Supplemental data for this article can be accessed online at https://doi.org/10.1080/24725854.2021.1931571.

for a one unit decrease in the Developer variance without running separate experiments to assess it. More generally, we want to estimate *all* input sensitivities from the same nominal simulation experiment.

Surprisingly, defining the seemingly intuitive sensitivity described above is not easy, and in fact there are an infinite number of correct answers. *In this article we define a new family of output-property-with-respect-to-input-property sensitivity measures whose members are interpretable, can be estimated using well-established technology for stochastic gradient estimation, and any member of the family can be obtained from the nominal simulation experiment.*

This article is organized as follows: The next section contains a broad overview of "sensitivity analysis" in computer experiments and describes where our contribution fits. We define our new family of sensitivity measures in Section 3, followed by two representative examples of stochastic simulations to which they apply in Section 4. The sensitivity estimators are established in Section 5. Section 6 summarizes results from an empirical study employing the two examples, followed by conclusions in Section 7. The online supplemental material describes supporting gradient-estimation methods, some additional empirical evaluation, variance estimation methods and some technical results.

2. Background

Computer models, including stochastic simulations, can be regarded as functions mapping inputs, denoted generically by $\mathbf{X} = (X^{(1)}, X^{(2)}, ..., X^{(K)})^\top$, into one or more outputs, denoted generically by Y, via a collection of rules and algorithms that mimic the features of interest: $Y = g(\mathbf{X})$. Sensitivity analysis investigates how the output of a computer model responds to variations in its inputs. Sensitivity analysis is of critical importance for identifying the relative contributions of the inputs to output uncertainty, assessing model risk, designing robust systems, calibrating the model, and quantifying the interactions among inputs (Saltelli *et al.*, 2000).

Based on the type of uncertainty inherent in the inputs, and the purpose of the model-based analysis, there are two broad categories of sensitivity analysis: global and local. Global sensitivity analysis is applicable when the input is a random variable naturally varying within its range, $X^{(i)} \sim F_i$, such as the daily temperature or the wind speed at a site during a specific season, and also when the input is believed to be a constant but we have less-than-complete knowledge of its value, such as the failure rate of an electronic component or the stress tolerance of a material. The goal in global sensitivity analysis is to apportion the overall output uncertainty to each of the inputs as a measure of their contribution.

Whereas measures of global sensitivity attempt to discern the inputs that drive the output uncertainty across each input's overall range, measures of local sensitivity focus on the influence of the inputs near a nominal setting. Local sensitivity analysis makes sense when the input is some parameter or property of a random variable, such as its mean, and we have some confidence in its nominal value. The goal of local sensitivity analysis is to measure the impact on the output of small perturbations of an input around this value. A local sensitivity measure is conceptually (and in our new family, precisely) a partial derivative of the output property with respect to the input property. *The "sensitivity analysis" we consider in this article is local.*

In the context of stochastic simulation when the simulation is driven by parametric input probability distributions — denoted by $X^{(i)} \sim F_i(\cdot|\boldsymbol{\theta}^{(i)})$ — then the parameters of each distribution are one type of input, denoted here by $\boldsymbol{\theta}^{(i)}$. Output variability depends on both the input probability distributions themselves (i.e., the inherent randomness of the system), and possibly uncertainty about the parameter values (e.g., if $\boldsymbol{\theta}^{(i)}$ is estimated by $\widehat{\boldsymbol{\theta}}^{(i)}$). When the distributions' parameters are estimated from historical data, then this additional output variability is referred to as "input uncertainty" in the simulation literature; see for instance, Barton *et al.* (2002), Barton *et al.* (2014), Song *et al.* (2014) and Lam *et al.* (2016). Thus, there is both sensitivity of the performance measures to the nominal values of these input-distribution parameters, and also statistical uncertainty as to their nominal values. *In this article we focus on the former: the local sensitivity of simulation output properties to input-distribution properties, and not input uncertainty.* Thus, our measures are useful even if distribution parameters are obtained from experience, subjective judgment, process physics, or guesses, as well as from data.

The reason we emphasize "input-distribution *properties*" is that sensitivity of the simulation output to the natural input-distribution *parameters* themselves is often difficult to interpret; this can be true even when the mean or variance of the distribution is one of the parameters. For example, a common sensitivity measure implemented in commercial software (e.g., Simio®) is simply the slope coefficient of a linear regression relating simulation output Y to the sample mean of the input variates. This measure quantifies how much Y would change per unit change in the sample mean of the input random variable, say $\bar{X}^{(i)}$, but cannot necessarily be interpreted as the partial derivative of $\mathrm{E}(Y)$ with respect to $\mathrm{E}(X^{(i)})$. Of course, the mean and variance are not the natural parameters of many distributions, such as the Weibull which is usually parameterized by shape and scale. Local sensitivities to such parameters are rarely meaningful to the simulation user; however, there are good methods for estimating them that we exploit.

In this article we reach beyond the partial derivative of the output mean with respect to the natural input-distribution parameters, to the partial derivative of an output *property* with respect to an input *property*. This can be represented conceptually as $\partial H_O(Y)/\partial H_I(X^{(i)})$, where $H(\cdot)$ is an operator yielding a property of a random variable, and the subscripts O and I are for the "output" and "input," respectively. Here we consider input distributions that are parametric, having parameters such as mean, variance, shape, scale, rate, etc. Thus, their properties can be represented as functions of their distribution parameters: $H_I(X^{(i)}) = r(\boldsymbol{\theta}^{(i)})$. We focus on $\mathrm{E}(\cdot)$ and $\mathrm{Var}(\cdot)$ here due to

their practical usefulness, but our family is more general, a point we return to the Section 7. Stated directly, we estimate the sensitivity of the *mean* or *variance* of the simulation output to the *mean* or *variance* of each input distribution around a nominal value of its parameters, $\boldsymbol{\theta}_0^{(i)}$. To achieve our goal, we propose a new family of local sensitivity measures that enable us to quantify $\partial H_O(Y)/\partial r(\boldsymbol{\theta}_0^{(i)})$ along a *meaningful direction* in the input-parameter space.

Our new sensitivity measures require the estimation of a *stochastic gradient* of the output property with respect to the natural input-distribution parameters, denoted by $\nabla_{\boldsymbol{\theta}_0^{(i)}} H_O(Y)$. This is a well-studied problem, and our methods apply to any output property for which they exist. Simulation-based gradient estimators can be categorized into two groups: indirect and direct methods. Indirect methods estimate an approximation of the true gradient by running additional simulations beyond the nominal setting; they require no knowledge of the underlying mechanics of the simulation model (Fu, 2008, 2015). The direct methods, which do require additional knowledge, lead to estimators that are often unbiased. We also employ the less-well-known method of Wieland and Schmeiser (2006) that is particularly appropriate for estimating output gradients with respect to input-distribution parameters. However, an appropriate stochastic gradient estimator depends on characteristics of the specific problem. Therefore, we describe three methods that apply to distinct situations that we expect to encounter in practice and provide practical advice as to how to choose and use them to obtain point and error estimators of our sensitivity measures in the online supplemental material.

Jiang *et al.* (2019) first introduced the idea of using directional derivatives in conjunction with standard gradient estimators. The focus of that paper is a tractable example — specifically the $M/G/\infty$ queue — to illustrate the impact and intuition behind choosing different directions. They also present an illustration using a semiconductor wafter fab simulation, as described earlier. However, Jiang *et al.* (2019) contains no theory, no recommendation on the choice of gradient estimator, and no standard error estimators for the sensitivity point estimators. Jiang *et al.* (2021) applies our local sensitivity estimators to clinical trial planning. An interesting feature of Jiang *et al.* (2021) is that for some outputs a direction that is *not* the steepest-ascent or minimum-mean-change directions emphasized in this article makes sense.

3. A new family of sensitivity measures

In this article we address the problem of local sensitivity of the mean or variance of the simulation output with respect to the mean or variance of its stochastic inputs. Some of the background material in this section is based on Section 2 of Jiang *et al.* (2019).

Consider a simulation model with K independent, scalar, parametric input distributions denoted $F^{(1)}(\cdot|\boldsymbol{\theta}^{(1)})$, $F^{(2)}(\cdot|\boldsymbol{\theta}^{(2)}), ..., F^{(K)}(\cdot|\boldsymbol{\theta}^{(K)})$, having in total $q \geq K$ input parameters (for some distributions $\boldsymbol{\theta}$ is a vector). Let $\boldsymbol{\Theta} = (\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, ..., \boldsymbol{\theta}^{(K)})$ be the vector of all input parameters, where $\boldsymbol{\theta}^{(i)} \in \Re^{p_i}$, with $p_i \geq 1$ the dimension of the parameter vector for input distribution i. The simulation output of interest can be represented as $Y(\boldsymbol{\Theta}) = \eta(\boldsymbol{\Theta}) + \varepsilon(\boldsymbol{\Theta})$ where $\eta(\boldsymbol{\Theta})$ is the expected value of the simulation output given the input parameters, and $\varepsilon(\boldsymbol{\Theta})$ is the corresponding stochastic noise with mean 0 and finite variance. In this article we consider the parameters $\boldsymbol{\Theta}$ to be fixed at $\boldsymbol{\Theta}_0$, so where no confusion is possible we will simply write Y. We also let $X^{(i)}$ represent a random variable with distribution $F^{(i)}$, whose mean μ_i and variance σ_i^2 are differentiable with respect to $\boldsymbol{\theta}^{(i)}$ at the nominal setting $\boldsymbol{\theta}_0^{(i)}$; this is true for the continuous-valued parameters of most standard distributions, provided that $\boldsymbol{\theta}^{(i)}$ is in the interior of the feasible parameter domain. Our local sensitivity is with respect to each input distribution separately, so for ease of exposition we focus first on a single input $X \sim F(\cdot|\boldsymbol{\theta})$ with parameter $\boldsymbol{\theta} \in \Re^p$, having mean $\mu = \mu(\boldsymbol{\theta})$, variance $\sigma^2 = \sigma^2(\boldsymbol{\theta})$ and nominal parameter value $\boldsymbol{\theta}_0$.

Suppose that we are interested in the effect of a unit change in the variance of an input random variable X on the variance of the output Y, which conceptually is $\partial \mathrm{Var}(Y)/\partial\sigma^2$. However, this partial derivative is not well defined when there are multiple ways to achieve a change in σ^2. That is, *different* changes in the distribution parameters that lead to the *same* change in the variance of the input might result in a *different* change in the variance of the output. *This fact is obvious, once stated, but is not well known or appreciated.* Therefore, the meaning of $\partial \mathrm{Var}(Y)/\partial\sigma^2$ is not clear, except for the special case of an input distribution that belongs to the location-scale family $X = \mu + \sigma W$, where W has mean 0 and variance 1. Similar issues arise if we want to estimate the impact on the variance of Y of changing the mean of X, or the impact on the mean of Y of changing the mean or variance of X. The key insight is that the mean and variance of both the output and the input are completely determined by $\boldsymbol{\theta}$; therefore, by *fixing* the direction of change in the input-parameter space we obtain a unique value for the desired sensitivity. We now formally introduce our new family of sensitivity measures. Given an output property H_O, an input property H_I, and a normed direction $\vec{\mathbf{d}}$ from the nominal parameter setting $\boldsymbol{\theta}_0$, we define the sensitivity of $H_O(Y)$ with respect to $H_I(X)$ as

$$\frac{\vec{\mathbf{d}}^T \nabla_{\boldsymbol{\theta}_0} H_O(Y)}{\vec{\mathbf{d}}^T \nabla_{\boldsymbol{\theta}_0} H_I(X)} \tag{1}$$

where ∇ is the gradient operator. This is simply an application of the chain rule for directional derivatives. The only requirements are that $\nabla_{\boldsymbol{\theta}_0} H_O(Y)$ exists and can be estimated, and that $\nabla_{\boldsymbol{\theta}_0} H_I(X)$ exists and can be computed. These are mild conditions, the obvious exception being when the input distribution has a discrete parameter, such as the binomial distribution. In the special case of a location-scale family $X = \mu + \sigma W$ our sensitivity measures with respect to the input mean and variance reduce to the $\partial H_O(Y)/\partial\mu$ and $\partial H_O(Y)/\partial\sigma^2$, respectively.

Remark. There are many possible ways to express "sensitivity," therefore, some sensible choices must be made to create a well-defined measure. A key choice that we have made is that the parametric family of the input distribution does not change as it is perturbed. Given this restriction, our definition is very flexible, as we illustrate later.

For practical reasons we focus on the four sensitivity measures relating means (M) and variances (V); we discuss percentile sensitivities in Section 7. We call the sensitivity of the mean of the output, $\mathrm{E}(Y) = \eta(\boldsymbol{\theta})$, with respect to the mean of the input $\mathrm{E}(X) = \mu(\boldsymbol{\theta})$, the *mean sensitivity to the mean (MSM)*. In the definitions below the first letter indicates the property of the output Y of interest and the final letter indicates sensitivity with respect to what property of the input X:

$$\mathrm{MSM}_{\vec{\mathbf{d}}} = \frac{\partial \mathrm{E}(Y)}{\partial \mu_{\vec{\mathbf{d}}}} = \frac{\vec{\mathbf{d}}^T \nabla_{\boldsymbol{\theta}_0} \mathrm{E}(Y)}{\vec{\mathbf{d}}^T \nabla_{\boldsymbol{\theta}_0} \mu} \tag{2}$$

$$\mathrm{MSV}_{\vec{\mathbf{d}}} = \frac{\partial \mathrm{E}(Y)}{\partial \sigma^2_{\vec{\mathbf{d}}}} = \frac{\vec{\mathbf{d}}^T \nabla_{\boldsymbol{\theta}_0} \mathrm{E}(Y)}{\vec{\mathbf{d}}^T \nabla_{\boldsymbol{\theta}_0} \sigma^2} \tag{3}$$

$$\mathrm{VSM}_{\vec{\mathbf{d}}} = \frac{\partial \mathrm{Var}(Y)}{\partial \mu_{\vec{\mathbf{d}}}} = \frac{\vec{\mathbf{d}}^T \nabla_{\boldsymbol{\theta}_0} \mathrm{Var}(Y)}{\vec{\mathbf{d}}^T \nabla_{\boldsymbol{\theta}_0} \mu} \tag{4}$$

$$\mathrm{VSV}_{\vec{\mathbf{d}}} = \frac{\partial \mathrm{Var}(Y)}{\partial \sigma^2_{\vec{\mathbf{d}}}} = \frac{\vec{\mathbf{d}}^T \nabla_{\boldsymbol{\theta}_0} \mathrm{Var}(Y)}{\vec{\mathbf{d}}^T \nabla_{\boldsymbol{\theta}_0} \sigma^2}. \tag{5}$$

For many input distributions the gradient of the mean or variance of X with respect its parameter $\boldsymbol{\theta}$ at $\boldsymbol{\theta}_0$, $\nabla_{\boldsymbol{\theta}_0}\mu$ or $\nabla_{\boldsymbol{\theta}_0}\sigma^2$, is available in closed form or easily computed numerically. The unknowns in (2)–(5) are $\nabla_{\boldsymbol{\theta}_0}\mathrm{E}(Y)$ and $\nabla_{\boldsymbol{\theta}_0}\mathrm{Var}(Y)$.

Remark. For clarity, we will use the terms "gradient" and "gradient estimators" when referring to $\nabla_{\boldsymbol{\theta}_0}\mathrm{E}(Y)$ and $\nabla_{\boldsymbol{\theta}_0}\mathrm{Var}(Y)$ and estimators of them, and we use the terms "sensitivity" and "sensitivity estimators" to refer to the right-hand side of (2)–(5) and estimators of them. Obviously, $\nabla_{\boldsymbol{\theta}_0}\mathrm{E}(Y)$ and $\nabla_{\boldsymbol{\theta}_0}\mathrm{Var}(Y)$ themselves could also be considered sensitivities.

Estimating $\nabla_{\boldsymbol{\theta}_0}\mathrm{E}(Y)$ has been studied extensively (L'Ecuyer, 1990; Fu, 2008, 2015). There exist many simulation-based techniques to estimate this gradient and we extend some of them to estimate $\nabla_{\boldsymbol{\theta}_0}\mathrm{Var}(Y)$. However, gradient estimation for the mean is *not* our contribution, and our extension to the variance is straightforward. The interested reader should see the cited references for conditions under which the various estimators exist, as well as their properties. What we do instead is to identify gradient estimators that fit our needs based on different practical situations described in Section 4; see the online supplemental material. Then in Section 5 we obtain point and error estimators of our proposed sensitivity measures; these are new. Although we focus on the sensitivity of the mean and variance, other properties such as quantiles also fit into this framework (see Section 7).

3.1. Choosing a meaningful direction

The proposed sensitivity measures can be computed along any direction $\vec{\mathbf{d}}$ chosen by the analyst, but our definition will only be valuable if there are practically meaningful directions. We address the choice of direction here. Before doing so, we point out that no additional simulation effort is required to compute the output sensitivity in one, two or more meaningful directions, and therefore every direction of interest can be evaluated.

The key to choosing a direction is how the analyst wants the *input property* to change. To be conservative, it will often make sense to choose the *steepest-ascent direction*, which is also the gradient direction, because in this direction the input property will change the fastest. Steepest ascent will be the best case or worst case, depending on whether change is good or bad. Any input property that has a gradient has a steepest-ascent direction. For instance, for sensitivity with respect to the variance of the input, the *steepest-ascent direction* along which σ^2 increases the fastest is $\vec{\mathbf{d}} = \nabla_{\boldsymbol{\theta}_0}\sigma^2 / ||\nabla_{\boldsymbol{\theta}_0}\sigma^2||$. In industrial engineering applications, variance is often a problem, so rapidly increasing variance is a pessimistic direction.

Another class of directions is to allow the input property of interest to change subject to constraints on other properties of the input. For instance, the *minimum-mean-change direction* of the input variance tries to maintain the mean of the input while increasing its variance:

$$\begin{aligned} \underset{\vec{\mathbf{d}} \in \Re^p}{\text{Minimize}} : & \left| \vec{\mathbf{d}}^\top \nabla_{\boldsymbol{\theta}_0} \mu(\boldsymbol{\theta}) \right| \\ \text{subject to} : & \ \vec{\mathbf{d}}^\top \nabla_{\boldsymbol{\theta}_0} \sigma^2(\boldsymbol{\theta}) > 0 \\ & ||\vec{\mathbf{d}}|| = 1. \end{aligned}$$

For many distributions the mean can be held constant (no change). This direction makes sense in applications when we expect to be able to reduce variability without compromising the central value.

We have illustrated steepest-ascent and constrained directions for output variance with respect to input variance, but the mean can play the role of output or input property as well. It is also possible that the analyst has a direction that makes particular sense for their application; perhaps the basic physics of the process suggests that when this input changes it changes in a particular way. In addition, a problem-specific choice is needed when there are shifted distributions or alternative parameterizations; we address those scenarios in the next two subsections.

3.2. Shifted distribution

Many useful distributions have their support on $[0, \infty)$, including the exponential, gamma, log-logistic, lognormal, Rayleigh and Weibull. Thus, a "shift" parameter is often needed when, say, the lower bound of the physical process is greater than zero. Notice that the minimum-mean change direction of the input variance may not be unique for input distributions with $p > 2$ parameters. Here we address the

special case of a three-parameter distribution obtained by shifting the lower bound of a two-parameter distribution.

Consider the shifted gamma distribution as an example, $X' = X + \xi$ where $X \sim \text{gamma}(\alpha, \beta)$, α is the shape parameter, β is the rate parameter, and ξ is the shift parameter (i.e., $\boldsymbol{\theta} = (\alpha, \beta, \xi)$). Notice that ξ does not affect the variance. Thus, the steepest-ascent direction for sensitivity with respect to the variance of X' is

$$\vec{\mathbf{d}} = \frac{\nabla_{\boldsymbol{\theta}_0}\sigma^2}{||\nabla_{\boldsymbol{\theta}_0}\sigma^2||} = \left(\frac{\beta}{\sqrt{4\alpha^2+\beta^2}}, -\frac{2\alpha}{\sqrt{4\alpha^2+\beta^2}}, 0\right) \tag{6}$$

where $(\beta/\sqrt{(4\alpha^2+\beta^2)}, -2\alpha/\sqrt{4\alpha^2+\beta^2})$ is the direction that most rapidly increases the variance of the X.

As ξ can compensate any change in the mean, there are multiple ways to do a minimum-mean-change direction unless we fix ξ. We argue that fixing ξ is typically the most relevant case in practice because it defines the support of the distribution; if sensitivity with respect to the support is the goal then it should be assessed directly, rather than indirectly through a change in the mean or variance. With the lower bound ξ fixed, the minimum-mean-change direction for sensitivity with respect to the variance of X' is given by

$$\vec{\mathbf{d}} = \left(-\frac{\alpha}{\sqrt{\alpha^2+\beta^2}}, -\frac{\beta}{\sqrt{\alpha^2+\beta^2}}, 0\right)$$

where $\left(-\alpha/\sqrt{\alpha^2+\beta^2}, -\beta/\sqrt{\alpha^2+\beta^2}\right)$ is the minimum-mean-change direction for X.

3.3. Alternative parameterizations

Another issue of note is that even for sensitivity measures from the same family along conceptually the same direction, a different parametrization of the input distribution might result in a different sensitivity value. Consider again the gamma distribution that has two parameterizations in common use: gamma(α, β) for which $\mu = \alpha/\beta, \sigma^2 = \alpha/\beta^2$, and gamma$(k, \theta)$ for which $\mu = k\theta$ and $\sigma^2 = k\theta^2$. Thus, $\alpha = k$ and $\beta = 1/\theta$. The corresponding unit-norm steepest-ascent directions of the variance of the gamma distribution under these two parameterizations are

$$\vec{\mathbf{d}}_1 = \left(\frac{\beta}{\sqrt{4\alpha^2+\beta^2}}, -\frac{2\alpha}{\sqrt{4\alpha^2+\beta^2}}\right), \text{and}$$
$$\vec{\mathbf{d}}_2 = \left(\frac{\theta}{\sqrt{\theta^2+4k^2}}, \frac{2k}{\sqrt{\theta^2+4k^2}}\right) \tag{7}$$

respectively, and the minimum-mean-change directions are

$$\vec{\mathbf{d}}_1 = \left(-\frac{\alpha}{\sqrt{\alpha^2+\beta^2}}, -\frac{\beta}{\sqrt{\alpha^2+\beta^2}}\right), \text{and}$$
$$\vec{\mathbf{d}}_2 = \left(-\frac{k}{\sqrt{\theta^2+k^2}}, \frac{\theta}{\sqrt{\theta^2+k^2}}\right) \tag{8}$$

respectively.

Does it matter? Suppose that the service-time distribution of an $M/G/\infty$ queue is gamma, and the output performance of interest, Y, is the number of customers in the system in steady state. Let λ be the rate parameter for the interarrival-time distribution. Then $\text{E}(Y) = \lambda\alpha/\beta = \lambda k\theta$. Thus, along the steepest ascent directions in (7), the corresponding $\text{MSV}_{\vec{\mathbf{d}}}$'s are given by

$$\text{MSV}_{\vec{\mathbf{d}}_1} = \frac{\lambda\beta^3 + 2\lambda\alpha^2\beta}{\beta^2+4\alpha^2}$$
$$\text{MSV}_{\vec{\mathbf{d}}_2} = \frac{\lambda\theta^2 + 2\lambda k^2}{\theta^3+4k^2\theta} = \frac{\lambda\beta + 2\lambda\alpha\beta^3}{1+4\alpha^2\beta^2}.$$

Apparently, $\text{MSV}_{\vec{\mathbf{d}}_1} \neq \text{MSV}_{\vec{\mathbf{d}}_2}$, which can be explained by the different rates of change of the output mean and the input variance while increasing β vs. θ. The $\text{MSV}_{\vec{\mathbf{d}}}$ along the two minimum-mean-change directions in (8), on the other hand, are both equal to zero, which makes sense because $\text{E}(Y)$ does not depend on the variance of the service-time distribution, only the mean.

What should be done in practice? We suggest employing the parameterization that was originally chosen for the input distribution. However, within our family the user can pick *any*, or *multiple*, directions $\vec{\mathbf{d}}$ that they find meaningful without affecting our definition, or the point and error estimators presented below.

4. Two examples

In Section 3 we defined four members of our family of sensitivity measures and noted that the key to applying them is estimating $\nabla_{\boldsymbol{\theta}_0}\text{E}(Y)$ and $\nabla_{\boldsymbol{\theta}_0}\text{Var}(Y)$; in the online supplemental material we review the Finite-Difference (FD), Likelihood-Ratio (LR), and Wieland-and-Schmeiser (WS) gradient estimators. In brief, FD provides a biased gradient estimator by executing simulations at the nominal and at perturbed parameter settings and taking differences; LR provides an unbiased estimator of the entire gradient from the nominal experiment by reweighting the output using the input distributions' score functions; and WS provides an estimator of the entire gradient from the slope coefficients of a least-squares regression of the simuation outputs on the inputs, and is unbiased under a multivariate normal assumption.

An appropriate gradient estimator depends on characteristics of the input and the output; all gradient-estimation methods use observed outputs Y, and possibly observed inputs X, but in different ways. We employ the following two examples to illustrate three distinct contexts that arise frequently in practice: An $M/G/1$ queue with gamma-distributed service time illustrates the situation when there are within-replication estimators of both the input-distribution parameter and the output property. A Stochastic Activity Network (SAN) illustrates two further cases: (i) when neither the input parameter nor the output property can be estimated within each replication (so multiple replications are essential); and (ii) when only an estimator of the output property, but not of the input-distribution parameter, is observed within each replication. As we will show, FD is

always applicable but expensive; WS is natural for the $M/G/1$ queue; whereas LR is ideal for situations like the SAN example.

4.1. $M/G/1$ queue

An $M/G/1$ queue with gamma-distributed service time has $K=2$ input distributions and $q=3$ parameters: the interarrival time following an exponential distribution with $\boldsymbol{\theta}^{(1)} = \lambda$, and the service time following a gamma distribution with $\boldsymbol{\theta}^{(2)} = (\alpha, \beta)$. To execute the simulation we set the value of these parameters to $\boldsymbol{\theta}_0^{(1)}$ and $\boldsymbol{\theta}_0^{(2)}$, respectively. Among a total of n replications, the jth replication generates m independent and identically distributed (i.i.d.) interarrival times, $X_{ij}^{(1)}, i = 1, 2, \ldots, m,$ and m i.i.d. service times, $X_{ij}^{(2)}, i = 1, 2, \ldots, m,$ where $m > 1$.

Since multiple input variates are observed within each replication, the input parameter $\boldsymbol{\Theta}_0$ can be estimated, for instance via maximum likelihood. Denote the estimators of the input parameters from within the jth replication as $\widehat{\boldsymbol{\Theta}}_j = (\widehat{\boldsymbol{\theta}}_j^{(1)}, \widehat{\boldsymbol{\theta}}_j^{(2)})$. We define this estimator even though $\boldsymbol{\Theta}_0$ is *known*, because one of the gradient estimators exploits an internally generated estimate of this known parameter.

Replication j also generates m outputs, $W_{\ell j}, \ell = 1, 2, \ldots, m$. Suppose $W_{\ell j}$ is the waiting time of the ℓth of a total of m customers arriving to the system after a sufficient warm-up period and before the stopping time within the jth replication. Then one key output from the jth replication is $Y_j = \sum_{\ell=1}^{m} W_{\ell j}/m$, an estimator of the steady-state mean waiting time of customers in the system. If the performance measure of interest is the steady-state variance of the waiting time of customers in the system, then the key output is $Y_j = \sum_{\ell=1}^{m} (W_{\ell j} - \bar{W}_j)^2/(m-1)$ where $\bar{W}_j = \sum_{\ell=1}^{m} W_{\ell j}/m$. Thus, the $M/G/1$ queue illustrates a setting in which we observe i.i.d. pairs $(Y_j, \widehat{\boldsymbol{\Theta}}_j), j = 1, 2, \ldots, n$.

4.2. A SAN

This example is based on a problem created by Burt and Garman (1971). A small instance of a project planning problem is modeled as a SAN. The network is shown in Figure 1 where the nodes (circles) represent project milestones and the arcs (arrows) are activities to be completed. The project starts from the source node a and is completed when the sink node d is reached, with the rule that all outgoing activities from a node begin when all of the incoming activities to that node are completed. The duration of the ith activity is a random variable $X^{(i)}$. Thus, the time to complete the project, Y, will be the longest path through the network: $Y = \max\{X^{(1)} + X^{(4)}, X^{(1)} + X^{(3)} + X^{(5)}, X^{(2)} + X^{(5)}\}$.

In this example, there are $K=5$ inputs whose distributions and parameters in our numerical illustration are given in Table 5 (in a later section). To execute the simulation we set the values of these parameters to their nominal values and run a total of n replications. Notice that for this simulation each replication generates exactly one sample from each input variate and one output value. Let $X_j^{(i)}$ be the sample generated from the distribution of the ith activity and Y_j be the output, both from the jth replication. Due to the single input variate from each input distribution within each replication, there is no natural within-replication estimator of $\boldsymbol{\theta}^{(3)}, \boldsymbol{\theta}^{(4)}$ and $\boldsymbol{\theta}^{(5)}$.

If the output property of interest is the mean time to complete the project, then Y_j returned from replication j is the corresponding estimator. However, if the property of interest is the variance of the time to complete the project, then no estimator of this output is observed within each replication. In this case we need a method to obtain the gradient of the variance of Y with respect to input parameters; we provide such a method in the online supplemental material, which is a small extension to the existing literature on gradient estimation. Thus, the SAN example illustrates a setting in which we observe i.i.d. pairs $(Y_j, \mathbf{X}_j)$.

5. Sensitivity measures and their variances

From here on $\boldsymbol{\theta}$ and $\widehat{\boldsymbol{\theta}}$ are $p \times 1$, denoting the parameter and its estimator of a single input distribution with nominal value $\boldsymbol{\theta}_0$; and $\boldsymbol{\Theta}$ and $\widehat{\boldsymbol{\Theta}}$ are $q \times 1$, containing the parameters across all K input distributions with nominal value $\boldsymbol{\Theta}_0$.

For the four families of sensitivity measures introduced in Section 3, the corresponding point estimator is obtained by plugging the appropriate gradient estimator into Definitions (2)–(5), i.e.,

$$\begin{aligned}
\widehat{\mathrm{MSM}}_{\vec{\mathbf{d}}} &= \vec{\mathbf{d}}^{\top} \widehat{\nabla}_{\boldsymbol{\theta}_0} \mathrm{E}(Y) \left(\vec{\mathbf{d}}^{\top} \nabla_{\boldsymbol{\theta}_0} \mu\right)^{-1} \\
\widehat{\mathrm{MSV}}_{\vec{\mathbf{d}}} &= \vec{\mathbf{d}}^{\top} \widehat{\nabla}_{\boldsymbol{\theta}_0} \mathrm{E}(Y) \left(\vec{\mathbf{d}}^{\top} \nabla_{\boldsymbol{\theta}_0} \sigma^2\right)^{-1} \\
\widehat{\mathrm{VSM}}_{\vec{\mathbf{d}}} &= \vec{\mathbf{d}}^{\top} \widehat{\nabla}_{\boldsymbol{\theta}_0} \mathrm{Var}(Y) \left(\vec{\mathbf{d}}^{\top} \nabla_{\boldsymbol{\theta}_0} \mu\right)^{-1} \\
\widehat{\mathrm{VSV}}_{\vec{\mathbf{d}}} &= \vec{\mathbf{d}}^{\top} \widehat{\nabla}_{\boldsymbol{\theta}_0} \mathrm{Var}(Y) \left(\vec{\mathbf{d}}^{\top} \nabla_{\boldsymbol{\theta}_0} \sigma^2\right)^{-1}.
\end{aligned} \tag{9}$$

All of these are linear functions of a gradient estimator $\widehat{\nabla}_{\boldsymbol{\theta}_0}$. Thus, if $\widehat{\nabla}_{\boldsymbol{\theta}_0}$ is unbiased or consistent, then so is the corresponding sensitivity estimator.

Notice that the only uncertain quantities in these expressions are the gradient estimators; therefore, their variances are

$$\begin{aligned}
\mathrm{Var}\left(\widehat{\mathrm{MSM}}_{\vec{\mathbf{d}}}\right) &= \vec{\mathbf{d}}^{\top} \mathrm{Var}\left(\widehat{\nabla}_{\boldsymbol{\theta}_0} \mathrm{E}(Y)\right) \vec{\mathbf{d}} \left(\vec{\mathbf{d}}^{\top} \nabla_{\boldsymbol{\theta}_0} \mu\right)^{-2} \\
\mathrm{Var}\left(\widehat{\mathrm{MSV}}_{\vec{\mathbf{d}}}\right) &= \vec{\mathbf{d}}^{\top} \mathrm{Var}\left(\widehat{\nabla}_{\boldsymbol{\theta}_0} \mathrm{E}(Y)\right) \vec{\mathbf{d}} \left(\vec{\mathbf{d}}^{\top} \nabla_{\boldsymbol{\theta}_0} \sigma^2\right)^{-2} \\
\mathrm{Var}\left(\widehat{\mathrm{VSM}}_{\vec{\mathbf{d}}}\right) &= \vec{\mathbf{d}}^{\top} \mathrm{Var}\left(\widehat{\nabla}_{\boldsymbol{\theta}_0} \mathrm{Var}(Y)\right) \vec{\mathbf{d}} \left(\vec{\mathbf{d}}^{\top} \nabla_{\boldsymbol{\theta}_0} \mu\right)^{-2} \\
\mathrm{Var}\left(\widehat{\mathrm{VSV}}_{\vec{\mathbf{d}}}\right) &= \vec{\mathbf{d}}^{\top} \mathrm{Var}\left(\widehat{\nabla}_{\boldsymbol{\theta}_0} \mathrm{Var}(Y)\right) \vec{\mathbf{d}} \left(\vec{\mathbf{d}}^{\top} \nabla_{\boldsymbol{\theta}_0} \sigma^2\right)^{-2}.
\end{aligned} \tag{10}$$

The key to estimating the variance of our sensitivity measure is estimating the variance of the corresponding gradient estimator $\widehat{\nabla}_{\boldsymbol{\theta}_0}$, where the situations we consider can be categorized into the following three settings:

- *Setting 1 (FD, LR)*: The gradient estimator with respect to the parameters of a single input distribution, $\widehat{\nabla}_{\boldsymbol{\theta}_0}$, is

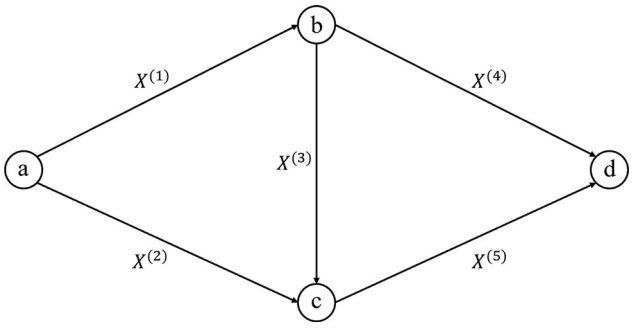

Figure 1. A small SAN.

the average of i.i.d. observations of the basic gradient estimator, $\widehat{\nabla}_1, \widehat{\nabla}_2, \ldots, \widehat{\nabla}_n$. Thus, the variance-covariance matrix of the gradient estimator can be estimated by $\widehat{\mathbf{V}} = \widehat{\Sigma}/n$, where $\widehat{\Sigma} = (n-1)^{-1}\sum_{j=1}^{n}(\widehat{\nabla}_j - \bar{\nabla})(\widehat{\nabla}_j - \bar{\nabla})^\top$ and $\bar{\nabla} = \sum_{j=1}^{n}\widehat{\nabla}_j/n = \widehat{\nabla}_{\boldsymbol{\theta}_0}$.

- *Setting 2 (WS)*:
The gradient estimator across all K distributions, $\widehat{\nabla}_{\boldsymbol{\Theta}_0} = (\widehat{\nabla}_{\boldsymbol{\theta}_0^{(1)}}^\top, \widehat{\nabla}_{\boldsymbol{\theta}_0^{(2)}}^\top, \ldots, \widehat{\nabla}_{\boldsymbol{\theta}_0^{(K)}}^\top)^\top$ is the ordinary least squares estimator of the slope coefficient $\widehat{\nabla}_{\boldsymbol{\Theta}_0} = \widehat{\boldsymbol{\beta}}_{1,\mathrm{OLS}}$, where

$$\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y} = \begin{bmatrix} \widehat{\beta}_{0,\mathrm{OLS}} \\ \widehat{\boldsymbol{\beta}}_{1,\mathrm{OLS}} \end{bmatrix}$$

with $\mathbf{Y} = [Y_1, Y_2, \ldots, Y_n]^\top$ and

$$\mathbf{X} = \begin{bmatrix} 1 & \mathbf{x}_1^\top \\ 1 & \mathbf{x}_2^\top \\ \vdots & \vdots \\ 1 & \mathbf{x}_n^\top \end{bmatrix}$$

with $\mathbf{x}_j$ the predictor variables from the jth replication. Assuming the joint distribution of $(Y_j, \mathbf{x}_j)$ is multivariate normal, the variance-covariance matrix of the slope coefficients is

$$\mathbf{V} = \frac{\sigma_\varepsilon^2}{n-q-2}\Sigma_{\mathbf{x},\mathbf{x}}^{-1} \tag{11}$$

where σ_ε^2 is the conditional variance of Y given $\mathbf{x}$. Therefore, we can estimate it by

$$\widehat{\mathbf{V}} = \frac{s_\varepsilon^2}{n-q-2}\left(\widehat{\Sigma}_{\mathbf{x},\mathbf{x}}\right)^{-1} \tag{12}$$

where $s_\varepsilon^2 = \mathrm{SSE}/(n-q-1)$, SSE is the sum of squared errors of the multiple linear regression of Y on $\mathbf{x}$, and $\widehat{\Sigma}_{\mathbf{x},\mathbf{x}}$ is the sample variance-covariance matrix of $\mathbf{x}$. The estimator of the variance-covariance matrix of $\widehat{\nabla}_{\boldsymbol{\theta}_0^{(i)}}$ is the ith $p_i \times p_i$ submatrix on the diagonal of $\widehat{\mathbf{V}}$. The complete derivation of this variance-covariance matrix and its estimator are found in Jiang *et al.* (2019).

- *Setting 3 (LR)*: The gradient estimator with respect to the parameters of a single input distribution, $\widehat{\nabla}_{\boldsymbol{\theta}_0}$, can be expressed as the average of $W_j + \widehat{\mu}U_j, j = 1, 2, \ldots, n$, where (W_j, U_j) are i.i.d. observations of the basic gradient estimator pairs, and

$$\widehat{\mu} \underset{n\to\infty}{\overset{\text{a.s.}}{\to}} \mu.$$

As n gets larger Setting 3 behaves like Setting 1, with $\widehat{\nabla}_j = W_j + \widehat{\mu}U_j$. Thus, we treat Setting 3 as Setting 1 with a plug-in estimator for μ; however, the small-sample properties of this variance estimator cannot be obtained.

In the online supplemental material we provide variance estimators for the FD, LR, and WS methods separately by categorizing each situation into one of the three settings above.

6. Empirical analysis

In this section we illustrate the estimation and interpretation of the proposed sensitivity measures, and evaluate our point and variance estimators, using the two examples introduced in Section 4; a less-controlled illustration for a wafer fabrication simulation may be found in Jiang *et al.* (2019).

Since the true gradients (and therefore sensitivity measures) are not known for either example, but the systems are computationally inexpensive to simulate, we employ intensive simulation to precisely estimate the true gradients for each output property with respect to each input parameter using the FD method; this in turn yields a "true" value of the corresponding sensitivity measures. To evaluate our sensitivity estimators using LR and WS, which is what we would do in practice, we compare to these "true" values. We also evaluate our variance estimators by observing whether the estimated standard error captures the difference between our sensitivity point estimate and the true value of the sensitivity measure.

Notice that we are *not* presenting an evaluation or a comprehensive study of the LR and WS gradient estimators themselves, as these are studied in the gradient-estimation literature. Rather, we demonstrate how these gradient estimators can be combined with our new family of sensitivity measures to yield useful and interpretable point and error estimators in typical settings. If and when better gradient estimators are invented, our sensitivity measures will benefit from them.

There is no obvious competitor for our sensitivity measures unless the gradient with respect to the input parameters itself is meaningful to the analyst. However, an alternative way to estimate our measures is by running the nominal experiment and then running a second experiment at a finite step along the desired parameter direction. This approach inherits all of the problems of FD — deciding how large a step to take, and the need for an additional simulation for each input and each direction of interest — so we focus on exploiting gradient estimators like LR and WS that obtain the entire gradient at once.

Recall that the proposed sensitivity measures reveal the change in the output mean or variance per unit change in the mean or variance of an input distribution along a meaningful direction. When we refer to "per unit change" for the mean it is in the natural units, whereas for the variance it in the natural units squared. Stating sensitivities as standard

Table 1. Experiment setup of $M/G/1$ queue example.

Input	*Distribution*	*Parameter*	*Nominal Value*
Interarrival time (ARR)	exponential	mean	$\theta_0^{(1)} = 1$
Service time (SER)	gamma	(shape, scale)	$\theta_0^{(2)} = (4, 5)$

Table 2. Regression results for $M/G/1$ queue example ($^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$; $^{****}p < 2e-16$).

Parameter	*Coefficient*	*Significance*	*StdErr (SE)*
ARR_{mean}	−10.6315	***	(0.3976)
$\text{SER}_{\text{shape}}$	3.0695	***	(0.2145)
$\text{SER}_{\text{scale}}$	−2.5316	***	(0.1596)
Intercept	13.0264	***	(0.5100)
Observations	900		
R^2	0.514		
Adjusted R^2	0.5124		
Residual Std. Error	0.1967 (df = 896)		
F Statistic	316**** (df = 3; 896)		

deviation rather than variance is possible, and probably more useful in practice.

6.1. $M/G/1$ queue

The output property of interest is the steady-state mean waiting time of customers in an $M/G/1$ queue. We illustrate estimating its sensitivity with respect to the mean of each input distribution when the mean changes along the steepest-ascent direction, and with respect to the variance of each input distribution when the variance changes along the steepest-ascent and minimum-mean-change directions.

The waiting time is simulated via Lindley's equation and, to speed up the convergence to steady state, the system is preloaded with one waiting time at the steady-state expected value obtained from the Pollaczek–Khinchine formula. The warm-up period is the first 200 customers; after that, the waiting times of 4000 customers arriving to the system are averaged to estimate the steady-state mean. The nominal experiment ran 900 replications with the input distributions specified in Table 1. For simplicity of notation, we use "ARR" for the interarrival-time input and "SER" for the service-time input. The intensive simulation to estimate the true sensitivities ran 64,000 replications, ensuring the relative error of the gradient estimator to be less than 0.001.

Since multiple variates are observed from both the interarrival-time and the service-time distributions within each replication of the nominal experiment the WS gradient estimator is particularly appropriate. Furthermore, because the distributions of the MLEs of the distribution parameters, $\widehat{\boldsymbol{\Theta}}$, are asymptotically normal and the mean waiting time is the average of the waiting times of a large number of customers arriving to system within each replication, it is plausible to approximate the distribution of $(Y, \widehat{\boldsymbol{\Theta}})$ as multivariate normal and thus the relationship between $\mathrm{E}(Y|\widehat{\boldsymbol{\Theta}})$ and $\widehat{\boldsymbol{\Theta}}$ as approximately linear. Accordingly, we have the sufficient conditions to apply the WS method — linear regression of Y on $\widehat{\boldsymbol{\Theta}}$ — to obtain the gradient estimator, $\widehat{\nabla}_{\boldsymbol{\Theta}_0}\mathrm{E}(Y)$, and its variance-covariance matrix.

A summary of the fitted model is shown in Table 2, where we see that although the adjusted R^2 of 0.51 is low, all predictors are significant. We also applied model diagnostics to validate assumptions including normality, homoscedasticity, and linearity. In summary, we conclude that the linear model fits the data well. Thus, we can draw important conclusions about how changes in the input distribution parameters affect the mean waiting time from the fitted model. For example, the coefficient associated with the mean of the interarrival time is negative, which makes sense because longer interarrival times will help mitigate the congestion and reduce the expected waiting time. A similar explanation applies to the positive (negative) sign of the shape (scale) parameter of the service-time input distribution because increasing (decreasing) shape (scale) increases the mean of the service time which is the main driver of congestion in the queue.

Table 3. MSM estimates for $M/G/1$ queue example.

$\text{MSM}_{\text{Input, Direction}}$	*Estimator (WS)*	*SE*	*"True" value (FD)*
$\text{MSM}_{\text{ARR, SA}}$	−10.6315	0.3976	−9.9594
$\text{MSM}_{\text{SER, SA}}$	15.5142	1.0243	14.6577

Table 4. MSV estimates for $M/G/1$ queue example.

$\text{MSV}_{\text{Input, Direction}}$	*Estimator (WS)*	*SE*	*"True" value (FD)*	$\Delta\mu$
$\text{MSV}_{\text{ARR, SA}}$	−5.3158	0.1988	−4.9797	0.5
$\text{MSV}_{\text{SER, SA}}$	49.9413	3.2365	47.2310	3.2
$\text{MSV}_{\text{SER, MM}}$	2.3815	1.9415	2.6760	0

After plugging the gradient estimates into (9) and their variances into (10), we report the MSM and MSV estimates and their Standard Errors (SEs) along with their "true" values in Tables 3 and 4. The two subscripts specifying the direction of sensitivity measures are "SA," denoting the steepest-ascent direction, and "MM," denoting the minimum-mean-change direction. Notice for all MSM and MSV estimates, the "true" value is included in the $\pm 2 \times$ SE interval, indicating that the MSM and MSV are pretty well estimated using the WS method with 900 simulation replications.

In Table 3, the $\text{MSM}_{\text{SER, SA}}$ estimator suggests that the steady-state mean waiting time is expected to increase by about 16 time units per unit increase in the mean of the service time at the fastest rate. The $\text{MSM}_{\text{ARR, SA}}$ estimator, on the other hand, implies that the steady-state mean waiting time is expected to decrease by around 11 time units per unit increase in the mean of the interarrival time. Thus, this table suggests that the steady-state mean waiting time is more sensitive to the mean service time at this nominal setting.

In Table 4, the $\text{MSV}_{\text{SER, SA}}$ estimator implies that the steady-state mean waiting time would increase by around 50 time units when the variance of the service time increases by one unit at the fastest rate, which is about three times the $\text{MSM}_{\text{SER, SA}}$ estimator. This can be explained by the fact in the SA direction for the variance both the the mean and the variance of the service time increase. The $\Delta\mu$ column in Table 4, where $\Delta\mu(\boldsymbol{\theta}_0) = \vec{\mathbf{d}}^T \nabla_{\theta_0}\mu / \vec{\mathbf{d}}^T \nabla_{\theta_0}\sigma^2$, tells us approximately how much the mean of each input distribution, $\mu(\theta)$, would change if the variance of the distribution $\sigma^2(\boldsymbol{\theta})$ changes one unit. Notice that the $\text{MSV}_{\text{SER, MM}}$ estimator

Table 5. Experiment setup of SAN example.

Input	Distribution	Parameter	Nominal value
$X^{(1)}$	exponential	mean	$\theta_0^{(1)} = 5$
$X^{(2)}$	exponential	mean	$\theta_0^{(2)} = 15$
$X^{(3)}$	Weibull	(shape, scale)	$\boldsymbol{\theta}_0^{(3)} \equiv (\vartheta_1^{(3)}, \vartheta_2^{(3)}) = (5, 11)$
$X^{(4)}$	gamma	(shape, rate)	$\boldsymbol{\theta}_0^{(4)} \equiv (\vartheta_1^{(4)}, \vartheta_2^{(4)}) = (30, 2)$
$X^{(5)}$	gamma	(shape, rate)	$\boldsymbol{\theta}_0^{(5)} \equiv (\vartheta_1^{(5)}, \vartheta_2^{(5)}) = (20, 4)$

indicates that per unit increase in the variance of the service time in the minimum-mean-change direction would lead to only a 2 time unit increase in the mean waiting time. Moreover, the $\pm 2 \times$ SE interval includes zero, implying this sensitivity might not be statistically significant. This substantial difference in $\text{MSV}_{\text{SER, SA}}$ and $\text{MSV}_{\text{SER, MM}}$ emphasizes the critical importance of specifying a direction of change to be able to interpret results.

6.2. SAN

In this example we measure the sensitivity of two output performance measures of the SAN — the mean and the variance of the time to complete the project — to the mean and variance of each of the five input distributions along meaningful directions. Specifically, for sensitivities with respect to the input mean (i.e., MSM and VSM measures), we consider the steepest-ascent direction of the mean of the input, and for sensitivities with respect to the input variance (i.e., MSV and VSV measures), the directions are the steepest-ascent and the minimum-mean-change directions.

The nominal setup of the experiment is specified in Table 5. The three paths connecting the source node and the sink node, $X^{(1)} + X^{(4)}, X^{(1)} + X^{(3)} + X^{(5)}$, and $X^{(2)} + X^{(5)}$, have balanced means so that each path is approximately equally likely to be the longest path. The two output properties of interest represent two different situations: whether there is, or is not, a within-replication estimator of the property of interest. We illustrate the estimation and interpretation of the sensitivity measures to the mean time to complete the project in the next subsections. Note that results for sensitivity of the variance of the project completion time can be found in the online supplemental material.

Despite the simplicity of this problem, gradients with respect to the activity time parameters are notoriously hard to estimate with generic methods; this fact has nothing to do with our sensitivity measures, it is simply a property of this noisy problem. To guarantee the relative error of the gradient estimator is less than 0.001, we ran 200,000 replications in the intensive simulation. The nominal experiment was also run with 200,000 replications, which is larger than we would expect to make in practice, but we wanted to have a precise comparison of sensitivity measures obtained using different gradient estimation methods. We also ran nominal experiments with a more reasonable number of replications (10,000) and report those results at the end of this subsection. A gradient estimator tailored specifically for this problem, perhaps employing variance-reduction techniques, would also help our sensitivity measures.

Table 6. LR Gradient estimates of SAN example with output E(Y).

Parameter	Gradient Estimator (LR)	SE	"True" value (FD)
$X^{(1)}_{\text{mean}}$	0.7284	0.0166	0.7385
$X^{(2)}_{\text{mean}}$	0.6886	0.0089	0.7015
$X^{(3)}_{\text{shape}}$	−0.0465	0.0183	−0.0332
$X^{(3)}_{\text{scale}}$	0.3521	0.0310	0.3468
$X^{(4)}_{\text{shape}}$	0.1379	0.0123	0.1667
$X^{(4)}_{\text{rate}}$	−2.2401	0.1844	−2.6337
$X^{(5)}_{\text{shape}}$	0.1745	0.0150	0.1752
$X^{(5)}_{\text{rate}}$	−0.9087	0.0748	−0.8820

The LR method is a good fit for the case when only one input variate is generated within each replication. The gradient estimator with respect to each input parameter is then an average of 200,000 corresponding LR gradient estimates. The LR gradient estimator of $\text{E}(Y)$ with respect to the mean of an exponential distribution (e.g., $X^{(1)}, X^{(2)}$), the shape and scale of a Weibull distribution (e.g, $X^{(3)}$), and the shape and rate of a gamma distribution (e.g.,$X^{(4)}, X^{(5)}$) are given in the online supplemental material. The estimated values of the gradients are shown in Table 6, along with the "true value" estimated using FD.

We also applied the WS method because there are sufficient replications to batch with a large enough batch size to obtain precise MLEs of each input parameter, and at the same time with enough batches for the subsequent regression. Specifically, we batched the observed input variates from each input distribution with batch size $b = 100$ to estimate the MLEs of each input distribution parameter and, to be consistent, the observed output with the same batch size to estimate its mean. Therefore, for the same reason as stated for the $M/G/1$ queue example, it is reasonable to approximate the joint distribution of the batch means of Y and the MLEs of all input parameters, $\widehat{\boldsymbol{\Theta}}$, as multivariate normal and we can use the WS method to estimate the gradient, $\widehat{\nabla}_{\boldsymbol{\Theta}_0}\text{E}(Y)$, and its variance-covariance matrix through regressing $\bar{Y}$ on $\widehat{\boldsymbol{\Theta}}$. The summary of the fitted model is shown in Table 7 where the coefficient column is the WS gradient estimates. As can be seen from Tables 6 and 7, both the LR and WS gradient estimates are consistent with the "true" values and their SEs are small.

In Table 7 all predictors are significant except the shape parameter of the distribution of $X^{(3)}$, which might be because the rate of change in the mean of $X^{(3)}$ with respect to its shape is the smallest compared with that of the other parameters at the nominal setting. The positive signs associated with the means of $X^{(1)}$ and $X^{(2)}$ are not surprising because increasing the mean should increase the length of the corresponding path, and accordingly, the probability of being the longest. A similar explanation applies to the signs associated with other predictors. The adjusted R^2 is 0.88. We also performed regression diagnostics to test the standard multiple linear regression assumptions and checked multicollinearity and outliers. In summary, we conclude that the linear model fits the data well.

After plugging the gradient estimates into (9) and their variances into (10), the MSM and MSV estimators using the LR and WS methods, their SEs, and their true values are reported in Tables 8 and 9. For both MSM and MSV

Table 7. Regression results for SAN example with output E(Y) (" "$p < 1$; "$*$" $p < 0.05$; "$**$" $p < 0.01$; "$***$" $p < 0.001$).

Parameter	*Coefficient*	*Significance*	*SE*
$X^{(1)}_{\text{mean}}$	0.7620	***	(0.0182)
$X^{(2)}_{\text{mean}}$	0.7047	***	(0.0064)
$X^{(3)}_{\text{shape}}$	−0.0231		(0.0246)
$X^{(3)}_{\text{scale}}$	0.3130	***	(0.0421)
$X^{(4)}_{\text{shape}}$	0.1381	***	(0.0163)
$X^{(4)}_{\text{rate}}$	−2.2031	***	(0.2430)
$X^{(5)}_{\text{shape}}$	0.17801	***	(0.0200)
$X^{(5)}_{\text{rate}}$	−0.8942	***	(0.0985)
Intercept	9.4142	***	(0.4690)
Observations	2000		
R^2	0.8765		
Adjusted R^2	0.876		
Residual Std. Error	0.4144 (df = 1991)		
F Statistic	1767*** (df = 8; 1991)		

Table 8. MSM estimates of SAN example.

$MSM_{Input,\,Direction}$	*Estimator (LR)*	*SE*	*Estimator (WS)*	*SE*	*"True" value (FD)*
$\text{MSM}_{X^{(1)},\text{SA}}$	0.7284	0.0166	0.7620	0.0182	0.7385
$\text{MSM}_{X^{(2)},\text{SA}}$	0.6886	0.0089	0.7047	0.0065	0.7015
$\text{MSM}_{X^{(3)},\text{SA}}$	0.3694	0.0322	0.3311	0.0440	0.3657
$\text{MSM}_{X^{(4)},\text{SA}}$	0.2986	0.0246	0.2937	0.0324	0.3511
$\text{MSM}_{X^{(5)},\text{SA}}$	0.7258	0.0598	0.7152	0.0788	0.7055

Table 9. MSV estimates of SAN example.

$MSV_{Input,\,Direction}$	*Estimator (LR)*	*SE*	*Estimator (WS)*	*SE*	*"True" value (FD)*	$\Delta\mu$
$\text{MSV}_{X^{(1)},\text{SA}}$	0.0728	0.0017	0.0762	0.0018	0.0738	0.1
$\text{MSV}_{X^{(2)},\text{SA}}$	0.0230	0.0003	0.023	0.0002	0.0234	0.0333
$\text{MSV}_{X^{(3)},\text{SA}}$	0.0896	0.0114	0.0722	0.0150	0.0832	0.1325
$\text{MSV}_{X^{(3)},\text{MM}}$	0.0452	0.0096	0.0315	0.0128	0.0385	0
$\text{MSV}_{X^{(4)},\text{SA}}$	0.2990	0.0246	0.2940	0.0324	0.3515	1.0011
$\text{MSV}_{X^{(4)},\text{MM}}$	0.0459	0.0064	0.0350	0.0084	0.0357	0
$\text{MSV}_{X^{(5)},\text{SA}}$	1.4671	0.1208	1.4447	0.1591	1.4250	2.0198
$\text{MSV}_{X^{(5)},\text{MM}}$	0.1163	0.0379	0.0132	0.0507	0.0184	0

measures estimated using either the LR or WS method, the value of almost every estimator is close to the true value obtained using the FD method and the SE is always smaller than the estimate itself by at least one order of magnitude. The only estimator that appears to have relatively large error is the LR $\text{MSV}_{X^{(5)},\text{MM}}$ estimator. This might be because the estimation error of the LR gradient estimator with respect to $X^{(5)}_{\text{rate}}$ is magnified in the minimum-mean-change direction and that the mean project completion time is relatively insensitive to the change in the variance of $X^{(5)}$ when holding its mean constant (as indicated by the "true value"). Since the WS method slightly outperforms the LR method in this setting, we use the corresponding estimates for illustrating the interpretation of MSM and MSV sensitivities.

In Table 8, the $\text{MSM}_{X^{(1)},\text{SA}}$ estimator is the largest, indicating that a unit increase in the mean of $X^{(1)}$ would lead to an increase in the mean project completion time by about 0.76 units, which is larger than the case when the mean duration of any other activity increases at the fastest rate. However, since the differences between $\text{MSM}_{X^{(1)},\text{SA}}$, $\text{MSM}_{X^{(2)},\text{SA}}$, and $\text{MSM}_{X^{(5)},\text{SA}}$ are not significant, the mean duration of all three activities should receive attention when managing the mean project completion time.

Table 10. Gradient estimates of SAN example with output E(Y) and 10,000 observations.

Parameter	*Gradient estimator (LR)*	*SE*	*Gradient estimator (WS)*	*SE*	*"True" value (FD)*
$X^{(1)}_{\text{mean}}$	0.7373	0.0736	0.7559	0.0366	0.7385
$X^{(2)}_{\text{mean}}$	0.6688	0.0384	0.7011	0.0121	0.7015
$X^{(3)}_{\text{shape}}$	0.0271	0.0781	−0.0423	0.0376	−0.0332
$X^{(3)}_{\text{scale}}$	0.2089	0.1364	0.6107	0.0842	0.3468
$X^{(4)}_{\text{shape}}$	0.1327	0.0545	0.1387	0.0274	0.1667
$X^{(4)}_{\text{rate}}$	−2.0549	0.8202	−2.1011	0.4089	−2.6337
$X^{(5)}_{\text{shape}}$	0.1575	0.0658	0.1093	0.0345	0.1752
$X^{(5)}_{\text{rate}}$	−0.7754	0.3301	−0.4936	0.1687	−0.8820

In Table 9, the $\text{MSV}_{X^{(1)},\text{SA}}$ implies that the mean project completion time is likely to increase around 0.076 time units, i.e., one-tenth of the $\text{MSM}_{X^{(1)},\text{SA}}$ estimator, per unit increase in the variance of $X^{(1)}$. This can be explained by $\Delta\mu$, which suggests that every unit increase in the variance along the steepest-ascent direction comes with 0.1 unit increase in the mean, and that the mean is more influential on the length of the longest path of the SAN. A similar explanation applies to the difference between the $\text{MSM}_{X^{(i)},\text{SA}}$ and the $\text{MSV}_{X^{(i)},\text{SA}}$ estimates for all of the other activities. Additionally, throughout we see the sensitivity to the variance in the steepest-ascent direction is consistently larger than in the minimum-mean-change direction, and in some cases when the mean is held constant the sensitivity may not be statistically significant, e.g., $\text{MSV}_{X^{(5)},\text{MM}}$. This is because the mean duration of activities is the primary determinant of the longest path, and the steepest-ascent direction of the variance also changes the mean, but the minimum-mean-change direction does not, as shown in the $\Delta\mu$ column. Comparing all the MSM and MSV estimates, the MSV with respect to $X^{(5)}$ along the steepest-ascent direction is the largest, suggesting that the variance of $X^{(5)}$ should receive attention if we want to control the length of the longest path.

Applying the same estimation process to the nominal experiment with 10,000 replications, we report the LR gradient estimates and the WS gradient estimates obtained with batch size $b = 20$ in Table 10. The resulting MSM and MSV estimates are reported in Tables 11 and 12. To assist with comparison, we also include the true values of the gradients and the corresponding sensitivity measures in these tables.

Comparing Tables 6 and 7 with Table 10, the WS gradient estimator obviously has the advantage because its SE does not suffer as seriously as the LR gradient estimator when the number of replications is smaller, even though most of the LR estimates themselves are still relatively close to true values. The big increase in the SE also explains the discrepancy in the sign of the LR gradient estimator with respect to $X^{(3)}_{\text{shape}}$.

In Table 11, the $\pm 2 \times$ SE interval for each LR MSM estimator includes the true value, but in some cases wrongly

Table 11. MSM estimates of SAN example with 10,000 observations.

$MSM_{Input,\,Direction}$	*Estimator (LR)*	*SE*	*Estimator (WS)*	*SE*	*"True" value (FD)*
$MSM_{X^{(1)},\,SA}$	0.7373	0.0736	0.7559	0.0366	0.7385
$MSM_{X^{(2)},\,SA}$	0.6688	0.0384	0.7011	0.0121	0.7015
$MSM_{X^{(3)},\,SA}$	0.2273	0.1419	0.6464	0.0889	0.3657
$MSM_{X^{(4)},\,SA}$	0.2739	0.1094	0.2801	0.0545	0.3511
$MSM_{X^{(5)},\,SA}$	0.6207	0.2639	0.3965	0.1350	0.7055

Table 12. MSV estimates of SAN example with 10,000 observations.

$MSV_{Input,\,Direction}$	*Estimator (LR)*	*SE*	*Estimator (WS)*	*SE*	*"True" value (FD)*	$\Delta\mu$
$MSV_{X^{(1)},\,SA}$	0.0737	0.0074	0.0755	0.0037	0.0738	0.1
$MSV_{X^{(2)},\,SA}$	0.0223	0.0013	0.0234	0.0004	0.0234	0.0333
$MSV_{X^{(3)},\,SA}$	0.0309	0.0490	0.1399	0.0252	0.0832	0.1325
$MSV_{X^{(3)},\,MM}$	0.0009	0.0412	0.0602	0.0198	0.0385	0
$MSV_{X^{(4)},\,SA}$	0.2743	0.1095	0.2805	0.0546	0.3515	1.0011
$MSV_{X^{(4)},\,MM}$	0.0174	0.0285	0.0054	0.0119	0.0357	0
$MSV_{X^{(5)},\,SA}$	1.2533	0.5332	0.7992	0.2726	1.4250	2.0198
$MSV_{X^{(5)},\,MM}$	−0.0389	0.1644	−0.1689	0.0668	0.0184	0

includes zero, e.g., $MSM_{X^{(3)},\,SA}$. On the other hand, for the WS MSM estimates, their $\pm 2 \times SE$ interval might fail to include the true value because of larger bias of the estimator itself, e.g., $MSM_{X^{(3)},\,SA}$ and $MSM_{X^{(5)},\,SA}$. Thus, when the number of observations is reasonable but still large, it is hard to tell which method has an absolute advantage over the other based on this experiment. Similar observations can be drawn from those MSV estimates in Table 12. Notice the wrong signs of the LR and WS $MSV_{X^{(5)},\,MM}$ estimates, which might be due to the minimum-mean-change direction of $X^{(5)}$ magnifying the moderate estimation error of the gradient estimator.

In the online supplemental material we show that estimating sensitivities for the project completion time variance requires a much higher simulation budget no matter what gradient estimation method is adopted. Nevertheless, the results in the online supplemental material show that the LR gradient estimators always outperform the WS gradient estimators with the same simulation budget, because of the smaller SE. Additionally, VSV is harder to estimate than VSM, especially if using the WS method. However, neither of the sensitivity estimates using LR or WS is uniformly accurate and precise.

In summary, if we obtain a large enough number of replications, then both WS and LR can work for this example; at smaller (but still large) sample sizes there are issues, especially when the output is $\mathrm{Var}(Y)$. The WS method is better for estimating the sensitivity of the $\mathrm{E}(Y)$, even with a moderate number of observations, and the LR method has obvious benefits for estimating the sensitivity of the $\mathrm{Var}(Y)$. As noted earlier, gradient estimation is difficult for the SAN, even with FD.

7. Conclusions

In this article we defined a new family of sensitivity measures for a simulation output property with respect to some input property based on directional derivatives. Unlike gradients with respect to the input-distribution parameters, our sensitivity measures are easy to interpret and allow for the selection of a direction that is meaningful for the problem at hand.

Although we focused on sensitivities of the output mean or variance with respect to the input mean or variance, the only restriction is that the input and output properties must be differentiable with respect to the input-distribution parameters. For instance, we might be interested in the sensitivity of an output quantile, or sensitivity with respect to an input quantile, and this is possible in our framework. As an example, if the input of interest has a Weibull distribution with parameters $\boldsymbol{\theta} = (\alpha, \beta)$, then for any $0 < p < 1$ the pth quantile of this input is $\beta[-\ln(1-p)]^{\frac{1}{\alpha}}$ whose gradient with respect to $\boldsymbol{\theta}$ is easily computed. Furthermore, both Hong (2009) and Jiang and Fu (2015) provide methods for estimating the gradient of output quantiles with respect to input-distribution parameters. Therefore, our family of sensitivity measures also applies to quantiles.

Nevertheless, identifying the inputs whose mean or variance has the greatest impact on output performance is often of interest for system design and control (e.g., Schoemig (1999) and Hopp and Spearman (2011)). For this case we identified two specific directions that seem appropriate for many applications, and by using existing gradient estimators, both point and error estimators for any member of the family were obtained solely using output data from the nominal experiment.

Our *definition* of the family of sensitivity measures does not depend on the gradient estimator used, but the statistical properties of our *estimators* do. We illustrated estimation of sensitivity in different contexts in Section 6. Although we considered generic gradient estimation methods, specifically FD, LR and WS, problem-specific approaches may also be employed.

An open issue is that our family of sensitivity measure exploit specifying a direction, but alternative parameterizations of an input distribution might lead to different values of the sensitivity measure along conceptually the same direction. Although we suggested adopting whatever parameterization was used in the simulation model, it makes sense to search for a parameterization-free definition of "direction."

Finally, in this article we only considered univariate input distributions. Our framework extends naturally to the steepest-ascent direction of parameter change for multivariate input distributions when gradients are available with respect to the natural parameters. However, other meaningful directions are harder to specify and remain a topic for future work.

Acknowledgments

We thank the Department Editor, Associate Editor and two referees for helpful improvements to the paper. Some background material in this paper was previously published in Jiang et al. (2019).

Funding

This research was partially supported by National Science Foundation Grant No. CMMI-1634982.

Notes on contributors

Xi Jiang is an Operations Research Specialist at SAS Institute. She received her Ph.D. in Industrial Engineering and Management Sciences at Northwestern University, where she majored in applied statistics and statistical learning, with minors in analytics and optimization. Her research focus is on stochastic simulation methodology and uncertainty and sensitivity analysis.

Barry L. Nelson is the Walter P. Murphy Professor of the Department of Industrial Engineering and Management Sciences at Northwestern and a Visiting Scholar at Lancaster University. His research is on the design and analysis of computer simulation experiments on models of discrete-event, stochastic systems, including methodology for simulation optimization, quantifying and reducing model risk, variance reduction, output analysis, metamodeling and multivariate input modeling. He has published numerous papers and three books, including *Foundations and Methods of Stochastic Simulation: A First Course* (Springer, 2013). Nelson is a Fellow of INFORMS and IISE.

Jeff Hong is Fudan Distinguished Professor, Hongyi Chair Professor, Head of Department of Management Science and Associate Dean of School of Data Science at Fudan University. His research interests include stochastic simulation, stochastic optimization, financial risk management and supply chain management. He is currently the Associate Editor-in-Chief of the *Journal of Operations Research Society of China*, the Simulation Area Editor of *Operations Research, an Associate Editor of Management Science*, and the President of INFORMS Simulation Society.

ORCID

Xi Jiang http://orcid.org/0000-0002-0083-5345
Barry L. Nelson http://orcid.org/0000-0002-1325-2624
L. Jeff Hong http://orcid.org/0000-0001-7011-4001

References

Barton, R.R., Chick, S.E., Cheng, R.C.H., Henderson, S.G., Law, A.M., Schmeiser, B.W., Leemis, L.M., Schruben, L.W. and Wilson, J.R. (2000) Panel discussion on current issues in input modeling, in *Proceedings of the 2002 Winter Simulation Conference*., IEEE Press, Piscataway, NJ, pp. 353–369.

Barton, R.R., Nelson, B.L. and Xie, W. (2014) Quantifying input uncertainty via simulation confidence intervals. *INFORMS Journal on Computing*, **26**(1), 74–87.

Burt, J.M. and Garman, M.B. (1971) Conditional Monte Carlo: A simulation technique for stochastic network analysis. *Management Science*, **18**(3), 207–217.

Fu, M.C. (2008) What you should know about simulation and derivatives. *Naval Research Logistics (NRL)*, **55**(8), 723–736.

Fu, M.C. (2015) Stochastic gradient estimation, in *Handbook of Simulation Optimization, Volume 216 of International Series in Operations Research & Management Science*. Springer, New York, pp. 105–147.

Hong, L.J. (2009) Estimating quantile sensitivities. *Operations Research*, **57**(1), 118–130.

Hopp, W.J. and Spearman, M.L. (2011) *Factory Physics*, Waveland Press, Long Grove, IL.

Jiang, G. and Fu, M.C. (2015) On estimating quantile sensitivities via infinitesimal perturbation analysis. *Operations Research*, **63**(2), 435–441.

Jiang, W.X., Nelson, B.L. and Hong, L.J. (2019) Estimating sensitivity to input model variance, in *Proceedings of the 2019 Winter Simulation Conference*, IEEE Press, Piscataway, NJ, pp. 3705–3716.

Jiang, X., Biller, B., Box, J. and Nelson, B.L. (2021) Simulation sensitivity analysis for clinical trial enrollment planning, in *Proceedings of the 2021 Winter Simulation. Conference*. IEEE Press, Piscataway, NJ.

Lam, H. (2016) Advanced tutorial: Input uncertainty and robust analysis in stochastic simulation, in *Proceedings of the 2016 Winter Simulation Conference*., IEEE Press, Piscataway, NJ, pp. 178–192.

L'Ecuyer, P. (1990) A unified view of the IPA, SF, and LR gradient estimation techniques. *Management Science*, **36**(11), 1364–1383.

Robinson, S. (2014) *Simulation: The Practice of Model Development and Use*. Palgrave Macmillan, London, UK.

Saltelli, A., Chan, K. and Scott, E.M. (2000) *Sensitivity Analysis*. John Wiley & Sons, New York, NY.

Schoemig, A.K. (1999) On the corrupting influence of variability in semiconductor manufacturing, in *Proceedings of the 1999 Winter Simulation Conference*, IEEE Press, Piscataway, NJ, pp. 837–842.

Song, E., Nelson, B.L. and Pegden, C.D. (2014) Advanced tutorial: Input uncertainty quantification, in *Proceedings of the 2014 Winter Simulation Conference*, IEEE Press, Piscataway, NJ, pp. 162–176.

Wieland, J.R. and Schmeiser, B.W. (2006) Stochastic gradient estimation using a single design point, in *Proceedings of the 2006 Winter Simulation Conference*, IEEE Press, Piscataway, NJ, pp. 390–397.