Proceedings of the 2016 Winter Simulation Conference
T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S. E. Chick, eds.

ADVANCED TUTORIAL: INPUT UNCERTAINTY AND ROBUST ANALYSIS IN STOCHASTIC SIMULATION

Henry Lam

Department of Industrial and Operations Engineering
University of Michigan
1205 Beal Ave.
Ann Arbor, MI 48109, USA

ABSTRACT

Input uncertainty refers to errors caused by a lack of complete knowledge about the probability distributions used to generate input variates in stochastic simulation. The quantification of input uncertainty is one of the central topics of interest and has been studied over the years among the simulation community. This tutorial overviews some methodological developments in two parts. The first part discusses major established statistical methods, while the second part discusses some recent results from a robust-optimization-based viewpoint and their comparisons to the established methods.

1 INTRODUCTION

Stochastic simulation takes in probabilistic assumptions that govern the generation of input variates, through system logic, to produce random outputs. When these probabilistic assumptions are misspecified or uncertain, the simulation outputs and hence the performance analysis of interest are subject to *input error*, *variability* or *uncertainty*. This uncertainty is distinct from errors due to finite simulation runs, known as *simulation* or *stochastic uncertainty*. In the simulation literature, these two sources of uncertainty are also called *extrinsic* and *intrinsic uncertainty*, respectively.

Input uncertainty could lead the modeler to underestimate risk or make suboptimal decisions. In the conventional practice of simulation modeling, the modeler conducts simulation runs under the assumed input model (obtained from, say, expert opinion or statistical fit). The modeler assesses the stochastic error by, for instance, constructing a confidence interval (CI). Typically, as the replication size increases, this CI shrinks to length zero, showing that the stochastic error becomes negligible. A modeler ignorant of the input uncertainty could get a false confidence of the simulation report, since the experimental results concluded under the presumption that the model is correct could be very different from the actual performance measure. This issue has been highlighted in several previous Winter Simulation Conference (WSC) Advanced Tutorials (Barton et al. 2002, Henderson 2003, Chick 2006a, Barton 2012, Song et al. 2014) and Nelson (2013), Section 7.2.

Input uncertainty can be categorized at a few levels. *Parametric uncertainty* refers to the error in misspecifying a parameter value, but otherwise under a known parametric family. *Model uncertainty* refers to the error in selecting the parametric family from among a finite or potentially infinite number of choices. More broadly, the term *model* here is viewed as any probabilistic specifications related to the input variates, including not only the marginal distributions but also cross-sectional or serial dependency structures. For instance, in a generic discrete-event system, the service times constitute one type of input variates. While physical knowledge could assert the use of an exponential service time distribution, uncertainty about its rate parameter comprises a source of parametric uncertainty. When there is no knowledge about whether the service time distribution belongs to, say, two classes of distributions like exponential or lognormal, or

worse still, when there are no specific choices at all, model uncertainty arises. Furthermore, uncertainty about the validity of the i.i.d. structure of the service times is also a type of model uncertainty.

This tutorial overviews the methodological developments in input uncertainty in two aspects. After Section 2 reviews some background concepts, Section 3 surveys the statistical methods in handling parametric and model uncertainty. Section 4 discusses some recent works from a robust optimization viewpoint that targets model uncertainty, and compares these works to the established methods. We will focus on performance analyses involving the estimation of expectation-type performance measures. Because of space limitation, our references are in no way exhaustive. Lastly, we also mention that there are other salient sources of modeling errors besides input uncertainty, including notably the misspecification of system logic, that are out of the scope of this tutorial.

2 WHAT TO DO IN INPUT UNCERTAINTY

Put in simple terms, the goals of input uncertainty are to:

- *Goal 1:* Quantify the sensitivity or the amount of contribution to output variability from the uncertainty of (each of) the input distributions.
- *Goal 2:* Generate an interval that covers the true performance measure with high confidence or credibility that takes into account both stochastic and input uncertainty.

Sensitivity refers to the extent of changes in the performance measure when the input parameter varies. Output variability measures the scope of errors in the simulation output propagated from the errors in misspecifying the input models. This depends on both the sensitivity of the performance measure with respect to the input parameter (from the system structure) and the statistical uncertainty of the parameter (from the input data). Goal 1 is important as it evaluates the risk of misspecifying each input parameter and also guides the allocation of data collection efforts across different input models. Goal 2 gives an overall assessment on the reliability of the output estimates; A wide interval signifies high uncertainty on the true value of the performance measure. Goal 1 is related to Goal 2 in that output variance could often be translated into the standard error needed in a CI.

Though beyond the scope of this tutorial, it is important to note that the third, and arguably the most important, goal is to incorporate input uncertainty quantification to make better *decision*. Nelson (2016) listed optimization and feasibility tests among the common decision-making tasks that motivate simulation modeling. Integrating input uncertainty in these tasks has gained growing interest and is an extensive topic by itself (e.g., Chick 2006b, Fan et al. 2013, Corlu and Biller 2013).

What is the best method for input uncertainty depends on the particular context. But at the very least, a good method to handle input uncertainty should be:

1. Computationally tractable, i.e., capable to accurately assess the input uncertainty with affordable computational overhead beyond the nominal simulation (i.e., the simulation under the best model assumption from the modeler's belief). The overhead here could refer to the additional simulation runs, resampling of data, and possibly other types of computation such as optimization (which is particularly relevant in Section 4). Availability of good experimental design for computational budget allocation is also important in this regard.
2. Flexible to incorporate different types of data (input, output) and expert beliefs.
3. Transparent, including the use of few and easily tuned method parameters.

3 STATISTICAL METHODS FOR INPUT UNCERTAINTY

The majority of the input uncertainty literature focuses on the use of input data. Some methods could also naturally integrate with expert opinion (e.g., Bayesian). This section first lays out the basic statistical

principles for input uncertainty. Then we will overview some frequentist and Bayesian methods in handling parametric and model uncertainty.

To facilitate discussion, let us denote the input variates as $\mathbf{X}$, and the random output of a simulation as $y(\mathbf{X})$. The variates $\mathbf{X}$ are generated from a distributional assumption, which we call ν generically. ν could be a (vector-valued) parameter $\theta \in \mathbb{R}^l$ in a chosen parametric family, an index of parametric family m together with the parameter, i.e., (m,θ), or a whole distribution function F. We denote $\mathbf{X}_\nu$ as the input variates under assumption ν. The estimation target is $\psi(\nu_0) = E[y(\mathbf{X}_\nu)|\nu = \nu_0]$, where ν_0 is what we assume to be the existing "true" or "correct" probabilistic description for $\mathbf{X}_\nu$.

The nominal simulation typically outputs the simulation average

$$\bar{\psi}(\hat{\nu}) = \frac{1}{r}\sum_{j=1}^{r} \hat{\psi}_j(\hat{\nu})$$

where $\hat{\psi}_j(\hat{\nu}) := y_j(\mathbf{X}_{\hat{\nu}})$ denotes the j-th simulation replication under the fitted model $\hat{\nu}$. As summarized neatly in Song et al. (2014), each simulation replication $\hat{\psi}_j(\hat{\nu})$ can be decomposed as a random effect model

$$\hat{\psi}_j(\hat{\nu}) = \psi(\nu_0) + (\psi(\hat{\nu}) - \psi(\nu_0)) + (\hat{\psi}_j(\hat{\nu}) - \psi(\hat{\nu})). \tag{1}$$

Thus, $Var(\bar{\psi}(\hat{\nu}))$ can be decomposed as

$$Var(\bar{\psi}(\hat{\nu})) = Var(\psi(\hat{\nu})) + \frac{E[Var(\hat{\psi}_j(\hat{\nu})|\hat{\nu})]}{r}. \tag{2}$$

The term $Var(\psi(\hat{\nu}))$ is the contribution of input uncertainty to the output variance due to the fact that $\hat{\nu}$ is only an estimate of ν_0. The term $E[Var(\hat{\psi}_j(\hat{\nu})|\hat{\nu})]/r$ is the stochastic uncertainty that goes away as simulation size $r \to \infty$. Note that the first variance in the RHS of (2) is with respect to the data noise, which, depending on the school of thoughts, could be summarized as the sampling distribution or the posterior distribution. The second variance in the RHS of (2) is with respect to the simulation noise.

Besides variance, bias from input errors could also play a role in the output variability (e.g., Henderson 2003, Song et al. 2014). In this regard, a more comprehensive measurement of the variability due to input uncertainty is the mean square error defined as $E(\psi(\hat{\nu}) - \psi(\nu_0))^2 = Var(\psi(\hat{\nu})) + (E[\psi(\hat{\nu})] - \psi(\nu_0))^2 = Var(\psi(\hat{\nu})) + Bias(\psi(\nu))^2$. Having said that, however, the effect of bias typically goes away at a faster rate than the variance as input data size increases, and this effect is not well-defined in the Bayesian setting. Perhaps due to these reasons, the literature focus has been on the variance, though it should be cautioned that bias can be significant in small samples.

Goal 1 in Section 2 is essentially about estimating $Var(\psi(\hat{\nu}))$, and better yet, estimating the amount of variability contributed by each model in $Var(\psi(\hat{\nu}))$. This can be viewed as a statistical estimation problem, but the non-standard challenges are: 1) ψ is a nonlinear function of $\hat{\nu}$. Thus, even if the sampling distribution (or posterior distribution) of $\hat{\nu}$ is known, effort is needed to convert it to the distribution of $\psi(\hat{\nu})$. 2) ψ can only be approximated by simulation in general. This leads to the question of how to separate the variance contributions from the input and the simulation (i.e., the two terms in (2)). This also raises the issue of computational effort and experimental design.

To connect Goal 1 with Goal 2, an application of the central limit theorem (CLT) gives, typically,

$$\psi(\hat{\nu}) - \psi(\nu_0) \approx N\left(0, \frac{\sigma^2}{n}\right)$$

where n is the input sample size, or in the case of q independent input models each with sample size n_i,

$$\psi(\hat{\nu}) - \psi(\nu_0) \approx N\left(0, \sum_{i=1}^{q} \frac{\sigma_i^2}{n_i}\right)$$

where σ^2 and σ_i^2 are some asymptotic variances of the normalization. On the other hand,

$$\bar{\psi}(\hat{\nu}) - \psi(\hat{\nu}) \approx N\left(0, \frac{Var(\hat{\psi}_j(\nu_0))}{r}\right).$$

Therefore, an asymptotically valid $(1-\alpha)$ level CI that takes into account both the input and the stochastic uncertainties is typically

$$\bar{\psi}(\hat{\nu}) \pm z_{1-\alpha/2}\sqrt{\sum_{i=1}^{q}\frac{\sigma_i^2}{n_i} + \frac{Var(\hat{\psi}_j(\nu_0))}{r}} \tag{3}$$

where, of course, σ_i^2 and $Var(\hat{\psi}_j(\nu_0))$ need to be plugged in by some empirical estimates. Even though (3) may be challenging to compute directly, it forms the basis for many methods in the literature.

3.1 Parametric Uncertainty: The Frequentist View

This subsection discusses parametric uncertainty from the frequentist view. In other words, we assume known parametric families for the input models, with an unknown (vector-valued) parameter $\theta \in \mathbb{R}^l$. The nominal simulation estimates $\bar{\psi}(\hat{\theta})$ using a point estimate $\hat{\theta}$ for the true input parameter θ_0.

3.1.1 Delta Method, Two-Point Method

Cheng and Holland (1995, 1997) show that, with large input data size n and simulation replication size r, the variance decomposition (2) can be approximated by the delta method

$$Var(\bar{\psi}(\hat{\theta})) = \mathbf{g}(\theta_0)'V(\theta_0)\mathbf{g}(\theta_0) + \frac{Var(\hat{\psi}_j(\theta_0))}{r} + \text{higher order term.} \tag{4}$$

Here $V(\theta_0)$ is the sampling variance of $\hat{\theta}$, typically of order $1/n$. For example, using the maximum likelihood estimator (MLE), $V(\theta_0) \approx I(\theta_0)^{-1}/n$ where $I(\theta_0)$ is the Fisher information matrix of the input parametric model. The gradient vector $\mathbf{g}(\theta_0) = \nabla_\theta \psi(\theta_0)$ is the *sensitivity coefficient* with respect to θ. Thus, in view of Goal 1, the input uncertainty quantification amounts to finding $\mathbf{g}(\theta_0)'V(\theta_0)\mathbf{g}(\theta_0)$.

Furthermore, as discussed in Lin et al. (2015), when $\theta_0 = (\theta_0^1, \ldots, \theta_0^q)$ where θ_0^i is the parameter of each of q independent input models (think about, for instance, the interarrival time distribution at each station in a queueing network), the input-contributed variance $\mathbf{g}(\theta_0)'V(\theta_0)\mathbf{g}(\theta_0)$ can be decomposed into $\sum_{i=1}^q \mathbf{g}^i(\theta_0^i)'V^i(\theta_0^i)\mathbf{g}^i(\theta_0^i)$, where $\mathbf{g}^i$ is the sensitivity coefficient associated with each model. Moreover, this is asymptotically equivalent to saying that $Var(\psi(\hat{\theta})) \approx \sum_{i=1}^q Var(E[\psi(\hat{\theta})|\hat{\theta}^i])$. In other words, the input-contributed variance is approximately the sum of the variances of conditional expectations on the individual models.

The matrix $V(\theta_0)$ relies only on the input model, but not the simulation model, and can be substituted approximately by a point estimate $V(\hat{\theta})$. The challenge in using the delta method is to approximate $\mathbf{g}(\theta_0)$. In principle, this can be estimated by a finite-difference (e.g., Fu 2006), but with a fixed simulation budget, the variance of the resulting estimator for $\mathbf{g}(\theta_0)'V(\theta_0)\mathbf{g}(\theta_0)$ is typically of order p, where p is the total number of parameters, since each component of $\mathbf{g}(\theta_0)$ needs to be estimated via a separate set of simulation runs. This thus becomes less appealing when there are many models and parameters. To remedy this issue, Cheng and Holland (1998) suggests a delta/two-point method to get an upper estimate of $(\mathbf{g}(\theta_0)'V(\theta_0)\mathbf{g}(\theta_0))^{1/2}$ by splitting the simulation budget between the first stage whereby a finite-difference estimation is carried out for each component with a smaller number of runs, and the second stage where all the runs are devoted to only two carefully chosen points of θ. Cheng and Holland (1998) shows that a judicious choice of the budget split can reduce the estimation variance to have a leading term independent of p. The first stage can be further eliminated if the sign of each component of $\mathbf{g}$ is known, leading to what they call the simplified two point method. Recently, Lin et al. (2015) developed a method to estimate

$\mathbf{g}(\theta_0)$ that requires only simulation runs at a single point $\hat{\theta}$. This method utilizes the idea in Wieland and Schmeiser (2006) to regress the output responses $y_j(\mathbf{X}_{\hat{\theta}})$ on $\hat{\theta}_j$, the point estimates of θ using only the input variates in the same run. This scheme is justified if the conditional mean of $(y_j(\mathbf{X}_{\hat{\theta}}), \hat{\theta}_j)$ is locally linear in $\hat{\theta}_j$, which holds if the distribution is multivariate normal or when batching is used. Finally, it is evident that all these methods can be applied for estimating the variance contribution of each individual input model.

Estimates of $Var(\psi(\hat{\theta}))$ could be readily converted into a CI by using (3). Moreover, Cheng and Holland (2004) studied a direct two-point method in constructing CIs without explicitly approximating the variance, under normality assumption of y_j.

3.1.2 Bootstrap and Metamodel-assisted Bootstrap

An alternative to directly estimating sensitivity coefficients is to use the bootstrap (Efron 1992). This is a widely used technique that resamples data to measure the variability of a given statistic. The bootstrap principle, roughly speaking, is to approximate the distribution of a statistic $t(\hat{\theta};\theta_0)$ with that of $t(\hat{\theta}^*;\hat{\theta})$, where $\hat{\theta}^*$ is the *bootstrapped* parameter that plays the role of a point estimate pretending that the true parameter is $\hat{\theta}$ (known) instead of θ_0 (unknown).

With regard to Goal 1, we take $t(\hat{\theta};\theta_0)$ as $Var(\psi(\hat{\theta}))$, where Var is evaluated under θ_0. Cheng and Holland (1997) suggested drawing n simple random samples with replacement from the data, and evaluating the bootstrapped parameter estimates B times, thus leading to θ^b, $b = 1,\ldots,B$. r simulation runs are conducted on each θ^b to get $\bar{\psi}(\theta^b)$. Cheng and Holland (1997) showed that the variance $Var(\bar{\psi}(\theta_0))$ approximates (4). Moreover, with a modified scheme, r can be replaced by any arbitrary replication size r' in the bootstrap computation, and separate estimates on the input-contributed variance and the stochastic variance can be obtained.

The bootstrap principle can be applied to approximate the CI in Goal 2. This entails constructing a $(1-\alpha)$ CI for $\psi(\hat{\theta}) - \psi(\theta)$ by approximating it with the $\alpha/2$ and $(1-\alpha/2)$ empirical quantiles of $\psi(\hat{\theta}^*) - \psi(\hat{\theta})$ (the so-called basic bootstrap). Barton et al. (2013) utilizes this idea by first building a metamodel to approximate $\psi(\cdot)$, followed by the bootstrap computation. Alternately, Cheng and Holland (2004) uses a similar principle, but directly evaluates the quantiles of $\bar{\psi}(\hat{\theta}^b)$ instead of $\psi(\hat{\theta}^*) - \psi(\hat{\theta})$ (the percentile bootstrap). As Barton (2012) pointed out, techniques to evaluate the distribution of conditional expectations, such as those obtained in Lee and Glynn (2003) and Steckley et al. (2016), can be used to approximate the quantiles of $\psi(\hat{\theta}) - \psi(\theta_0)$.

3.1.3 Some Comments on Computational Effort

It is a general phenomenon in input uncertainty quantification that it often requires additional computational effort beyond the nominal simulation. Direct implementations of the delta method or the bootstrap are typically computationally intensive (order $p-$ and B-fold of the nominal simulation size respectively). Techniques such as two-point method, regression-based method and metamodeling curb the computational burden. Other distinct viewpoints related to parametric input uncertainty include the experimental design for input data collection (e.g., Freimer and Schruben 2002, Kleijnen 1998).

3.2 Nonparametric Uncertainty

Nonparametric uncertainty refers to situations where there are no a priori choices for the parametric families of the input models. Though being a different scope of uncertainty, the concepts for handling nonparametric uncertainty can be borrowed from the parametric counterpart. The "parameter" ν in (1) can be viewed as the distribution function of the input variate F, and the variance decomposition (2) holds with $\hat{\nu}$ replaced by the empirical distribution $\hat{F}$.

The nonparametric delta method (Serfling 2009) approximates the input-contributed variance $Var(\psi(\hat{F}))$ as

$$Var(\psi(\hat{F})) \approx \frac{Var_{F_0}(IF(X))}{n} \tag{5}$$

when n is large, where $IF(\cdot)$ is called the influence function (Hampel 1974), defined on the domain of the input variate X and Var_{F_0} is taken under the true distribution F_0. $IF(\cdot)$ is a suitably defined Gateaux derivative of ψ taken with respect to F. An analogous approximation holds in the case of q independent input models

$$Var(\psi(\hat{F}^1,\ldots,\hat{F}^q)) \approx \sum_{i=1}^{q} \frac{Var(IF^i(X^i))}{n_i} \tag{6}$$

when the individual input data size n_i are all large enough. CIs can be constructed as in (3) as well.

Unlike in the parametric case, directly using the influence function in estimating $Var(\psi(\hat{F}))$ and $Var(\psi(\hat{F}^1,\ldots,\hat{F}^q))$ has not been considered in the simulation literature. Though the expression of IF could be explicitly derivable, direct estimation typically requires a computational budget that scales at least quadratically with the length of the input variates needed in each simulation run. Techniques such as bootstrapping is thus a natural alternative approach.

3.2.1 Resampling Methods

Barton and Schruben (1993, 2001) considered three resampling methods in constructing CIs: direct resampling, bootstrap resampling, and uniformly randomized empirical distribution function (EDF). Direct resampling splits the input data into several sections, and uses the data in each section to construct an empirical distribution and drive r simulation runs. Barton and Schruben (1993, 2001) used $r = 1$ in their examples and computed the empirical quantiles of the sections' outputs, since the input uncertainty overshadowed the stochastic uncertainty in their considered systems. In general, r could be chosen larger than 1, and the distribution of conditional expectations should be used to generate the quantiles. This scheme is closely related to the sectioning technique described in Asmussen and Glynn (2007) Chapter III, which used an estimate of the variance $Var(\psi(\hat{F}))$ and normality approximation instead of empirical quantiles.

When input data are less abundant, direct resampling is not appropriate since each section can have few data. Bootstrap resampling can be used instead to generate a collection of B new empirical distributions by drawing with replacement from the data. Alternately, uniformly randomized EDF generates copies of empirical distributions by first generating an ordered sequence of uniform random variates, and translating them into the weights on the ordered data. As noted by Barton and Schruben (2001), this method is equivalent to the so-called Bayesian bootstrap (Rubin 1981), and has similar computational load and numerical performance as bootstrap resampling. Much like in the parametric case, the simulation burden of bootstrap resampling and uniformly randomized EDF is of order Br.

Ankenman and Nelson (2012) and Song and Nelson (2015) studied the estimation of the relative variance contribution among several input models. Ankenman and Nelson (2012) used $Var(E[\psi(F^1,\ldots,F^{i-1},\hat{F}^i,F^{i+1},\ldots,F^q)|\hat{F}^i])$ as a measurement of the variance contribution from the i-th input model. Acknowledging that $Var(\psi(\hat{F}^1,\ldots,\hat{F}^q)) \neq \sum_{i=1}^{q} Var(E[\psi(F^1,\ldots,F^{i-1},\hat{F}^i,F^{i+1},\ldots,F^q)|\hat{F}^i])$ in general (though asymptotic equality holds thanks to (6)), Song and Nelson (2015) refined their measurement by including the variance of conditional variance, in addition to condition mean, in the calculation. They derived CIs for approximating the relative contribution according to this conditional-mean-and-variance approximation.

3.2.2 Comparison with Parametric Methods

The main advantage of nonparametric modeling is the alleviation of the issue of model misspecification. Direct application of bootstrap resampling to nonparametric models does not appear to bear much overhead compared to the parametric counterpart. However, nonparametric models offer less flexibility than parametric

since the former is intrinsically high-dimensional. For example, it is possible to build metamodels to improve the efficiency in quantifying parametric uncertainty. Serial dependency structure can be modeled conveniently using parametric models (e.g., Biller and Nelson 2005, Melamed et al. 1992), and so is the reflection of expert knowledge.

3.3 Bayesian Methods

The Bayesian approach provides a philosophically different viewpoint than in Sections 3.1 and 3.2, by modeling the unknown parameter ν as a random object. The prior distribution on ν captures the initial belief of the modeler (from expert opinion or physical implications, etc.). After collecting data, the belief is updated to a posterior distribution, which is the conditional distribution of ν given the data. This quantifies the uncertainty of the input parameters, subsequently propagated through the stochastic simulation runs. The posterior distribution thus plays an analogous role as the sampling distribution in the frequentist framework.

The Bayesian approach can in principle be applied to parametric uncertainty, uncertainty among several choices of models, and nonparametric uncertainty (using, e.g., nonparametric Bayes). The simulation literature has studied the first two types of uncertainty. Glynn (1986) first discussed the role of Bayesian method in simulation. Chick (2001) studied the posterior inference of input parameters and the computation of posterior mean output under both parametric uncertainty and the choice of several models. Zouaoui and Wilson (2004) refined the efficiency of the procedure in Chick (2001) and studied the decomposition of the posterior output variance, i.e., (2), into the variances due to model selection, parametric uncertainty, and simulation noise, extending the work of Zouaoui and Wilson (2003) which considers only parametric uncertainty. The approaches by Chick (2001) and Zouaoui and Wilson (2004) come down to Bayesian model average (BMA), which means taking a posterior average over the model choices. In contrast to the frequentist approach that identifies a single best model using information criterion or goodness-of-fit, this approach keeps all the candidate models in consideration and weights their importance through the posterior probabilities.

Chick and Ng (2002) and Ng and Chick (2006) studied the quantification and reduction of parameter uncertainty under polynomial-type assumptions on ψ. Xie et al. (2014) further investigated the construction of Gaussian-process-based metamodels for ψ, removing the functional assumption on ψ and the use of asymptotics in Ng and Chick (2006). More discussion on Bayesian methods in other simulation-related problems can be found in Chick (2006a). Additionally, it is worth pointing out that Bayesian methodology has been widely used in handling model errors in deterministic computer experiments (e.g., Kennedy and O'Hagan 2001).

Major advantages of Bayesian methods include: 1) It is a principled framework in uncertainty quantification in which the posterior distribution is all that is needed. In particular, this quantification holds exactly for any data size, in contrast to frequentist techniques that primarily rely on asymptotics and need to be tailor-made case-by-case; 2) It is convenient for handling multiple choices of parametric models, for instance through BMA; 3) It naturally incorporates expert opinions through the prior; 4) It is computationally advantageous for complex problems in dynamic optimization, ranking & selection, sampling allocation etc., by using the guiding posterior.

One challenge of Bayesian methods is the computational effort in estimating the posterior distribution. In general, this needs to be handled through Markov chain Monte Carlo (MCMC) or importance sampling. This issue can become less pressing with advances in sampling methods (Brooks et al. 2011) and the use of conjugate prior. Another common criticism is the specification of prior, and its philosophical controversy. However, the bottom line is that under common scenarios the Bayesian posterior distribution behaves asymptotically the same as the frequentist sampling distribution as data size increases (the Bernstein-von Mises Theorem; e.g., Van der Vaart 2000). In this sense, both criticisms above only apply to small samples in which the frequentists do not seem to have perfect solutions either.

4 ROBUST SIMULATION ANALYSIS

Robust simulation analysis takes a distinct viewpoint from the methods discussed in Section 3, using the premise of worst-case optimization. A generic formulation to handle input uncertainty in this framework is to solve the optimization

$$\max_{v \in \mathscr{U}} \psi(v) \quad \text{and} \quad \min_{v \in \mathscr{U}} \psi(v) \tag{7}$$

where $\mathscr{U}$ is the *uncertainty set*, borrowed from the terminology in robust optimization (Ben-Tal et al. 2009). $\mathscr{U}$ captures the belief, information, or uncertainty of the input model. This uncertainty is propagated to the mean output response through the optimizations. Since ψ typically needs to be evaluated by simulation, (7) often becomes a simulation optimization problem. In summary, the pair of optimizations in (7) gives the worst-case values of the performance measure subject to the uncertainty imposed by $\mathscr{U}$.

Formulation (7) is related to Goals 1 and 2 in the following way. If $\mathscr{U}$ specifies a high confidence region of the input model parameter, the optimal values of (7) will give upper and lower bounds on the performance measure at the same confidence level. This is Goal 2. When $\mathscr{U}$ denotes a direction or a small neighborhood of the input parameter, (7) quantifies the worst-case variability and can be interpreted as a local sensitivity analysis, thus related to Goal 1.

Several immediate questions arise: 1) How to construct $\mathscr{U}$? 2) How to solve (7)? 3) What are the pros/cons in using (7) compared to the conventional methods for simulation input uncertainty?

Before answering these questions, we should first mention that the robust approach could be applied both parametrically and nonparametrically, depending on how we specific v. In the parametric case, Hu et al. (2012) studied the computation of worst-case performance measure of a Gaussian model subject to mean and covariance uncertainty represented by linear matrix inequalities. Ryzhov et al. (2012) and Fan et al. (2013) studied worst-case frameworks for ranking & selection that blend with Bayesian and frequentist methods, either under parametric uncertainty or several candidate parametric models, respectively. In the rest of this section, we will focus on using (7) in the nonparametric regime, as this is where its strength appears the most. To answer Question 1, common types of $\mathscr{U}$ include:

- constraints based on a nonparametric distance, e.g., Kullback-Leibler (KL), χ^2, etc., from a nominal distribution.
- moment constraints, e.g., mean, second moment, covariance matrix etc., and support.
- shape constraints on the distributions or densities, e.g., convexity, monotonicity.

The next few subsections will discuss Questions 2 and 3 in detail. Some motivations of using the robust approach are: 1) It provides a natural framework for handling unexpected *model risk* beyond past data or expert opinion. 2) The uncertainty set $\mathscr{U}$ can represent prior belief on the input variates in some unconventional and nonparametric way, offering more flexibility than Bayesian methods (in which prior knowledge typically needs to be expressed parametrically) and nonparametric methods (which do not focus on prior knowledge). $\mathscr{U}$ can also be applied to conduct sensitivity analysis with respect to a nonparametric aspect of the inputs. 3) It provides new computational machinery using optimization, rather than resampling, to quantify input uncertainty.

Much of the discussion coming next is very recent and is still in active research. At present, this approach faces a few challenges, including most prominently computational tractability and the incapability to separate input uncertainty from stochastic uncertainty (both due to the simulation optimization nature). On the other hand, given its potential to handle forms of uncertainty beyond the scope of the conventional methods, and its successes witnessed in other areas (see some literature review next), it could be valuable to the stochastic simulation community.

4.1 Related Literature

The notion of robustness and worst-case optimization originates from two main, overlapping, literatures. The first is robust dynamic control in which the transition probability is ambiguous and is modeled as an adversary against which the control is optimized (e.g., Petersen et al. 2000). This approach has been applied across various areas like economics (Hansen and Sargent 2008) and finance (Glasserman and Xu 2013). The second line of literature is what is known as distributionally robust optimization (DRO). Instead of handling uncertain deterministic parameters as in the conventional robust optimization framework (Ben-Tal et al. 2009, Bertsimas et al. 2011), DRO finds robust solutions under ambiguous probability distributions in stochastic optimizations. Scarf et al. (1958) first proposed a minimax formulation for newsvendor problems under moment information. Recent works studied other types of constraints such as covariance (Delage and Ye 2010) and divergence balls (Ben-Tal et al. 2013).

4.2 A Formulation to Handle Unexpected Model Risk

Glasserman and Xu (2014) considered a formulation to quantify model risk in financial applications, through the worst-case optimizations

$$\max / \min_P E_P[h(X)] \quad \text{subject to} \quad D(P\|P_0) \le \eta \tag{8}$$

where D is the KL divergence and P_0 is a known nominal distribution (fitted from historical data or from expert knowledge). The decision variable in (8) is the unknown true distribution P, which is not assumed to belong to any parametric family except that it is absolutely continuous with respect to P_0. Glasserman and Xu (2014) interprets η as the amount of model misspecification due to lack of knowledge or inadequacy in data modeling. Obviously, the analysis coming out from the optimizations is more conservative as η grows, and reduces to the nominal simulation as η shrinks to zero.

Glasserman and Xu (2014) models this ambiguity on the output level, i.e., P is the distribution of the output X that can be approximated with a complex simulation. Typical in the robust framework, (8) is an infinite-dimensional optimization program if P is a continuous distribution. However, it turns out that by using Lagrangian duality, (8) admits a finite-dimensional representation of the solution given by an exponential change of measure on P_0

$$P^*(x) \propto P_0(x) e^{h(x)/\alpha^*}$$

where α^* is the Lagrange multiplier. The solution thus resembles an importance sampler, but, instead of being a variance reduction technique, it is used to evaluate the worst-case distribution. Glasserman and Xu (2014) formulates a sample average approximation (SAA) to solve for α^*, and call their scheme *robust Monte Carlo*. Hu and Hong (2013), Vidyashankar and Xu (2015) and Love and Bayraksan (2015), etc. studied related approaches for solving stochastic and chance-constrained optimizations.

How about applying this calculation on the input level? In other words, we want to solve

$$\max / \min_P \psi(P) \quad \text{subject to} \quad D(P\|P_0) \le \eta \tag{9}$$

where P now represents an input distribution. In general, ψ is nonlinear and nonconvex in P, thus (9) is not a convex program, not to mention that it can be infinite-dimensional. In this sense (9) is an intractable problem. However, Lam (2016a) shows that, if the amount of model misspecification is small, the optimal values of (9) can be Taylor-approximated as

$$\max / \min_P \psi(P) = \psi(P_0) \pm \xi_1(P_0)\sqrt{\eta} + \xi_2(P_0)\eta + \cdots \tag{10}$$

under the assumption that $\psi(P_0)$ is an expectation performance measure with i.i.d. input variates. The expansion coefficients in (10) can be expressed in terms of certain moments of $y_j(\mathbf{X})$ under the known nominal distribution P_0. Similar types of linearization as (10) have also been studied for simulation models in physical and biological systems (Dupuis et al. 2015).

There is a close connection of (10) with the nonparametric delta method in (5). The first order coefficient ξ_1 is $\sqrt{2Var_0(IF(X))}$, where Var_0 denotes the variance under P_0, and $IF(\cdot)$ is the influence function of ψ. The factor 2 comes from the local smoothness property of the KL divergence, i.e., this number changes if KL is replaced by other ϕ-divergences (Pardo 2005). Thus, the standard deviation of the influence function quantifies the local worst-case sensitivity of the performance measure as the model P deviates from P_0 in the nonparametric space. Note that "nonparametric" here refers to the assessment of model misspecification, where P_0 can as well be a parametric model. As shown in Lam (2016a), this sensitivity dominates any parametric sensitivities in some well-defined sense.

The choice of η in quantifying model risk in practice needs to be calibrated case-by-case. For instance, Glasserman and Xu (2014) suggested to calibrate η by matching the simulation output with real-world data. Further research is needed to identify principled techniques to calibrate η.

4.3 Modeling Nonparametric Beliefs

As discussed before, one strength of robust analysis is the capability to represent nonparametric knowledge as constraints, and likewise, to conduct sensitivity analysis with respect to a nonparametric aspect of the input. Below we will give some examples.

4.3.1 Nonparametric Assessment of the Effect of Dependency

Contrary to the established parametric framework in quantifying dependency (e.g, Biller and Nelson 2005, Xie et al. 2016), the robust approach could in a way quantify the nonparametric uncertainty of dependency. In the bivariate case, Glasserman and Yang (2015) studied the formulation

$$\max / \min_P E_P[h(X,Y)] \quad \text{subject to} \quad P \in \Pi(p,q),\ D(P\|P_0) \le \eta \tag{11}$$

where P is the joint distribution between X and Y, and $\Pi(p,q)$ denotes the set of all distributions with known marginal distributions p and q respectively. $D(P\|P_0)$ is the KL divergence between the bivariate distribution P and the known P_0, an independent model between X and Y. In other words, $D(P\|P_0)$ is the mutual information of P. Glasserman and Yang (2015) investigated the so-called iterative proportional fitting process to solve (11) in the discrete case, and applied it to quantify the impact of dependency between market and credit risks in derivative valuation.

As in Section 4.2, a parallel formulation can be used to quantify the effect of serial dependency in the input uncertainty context. Lam (2016b) elaborated the scheme

$$\max / \min_{\mathbf{P}} \psi(\mathbf{P}) \quad \text{subject to} \quad \mathbf{P} \in \Pi(p),\ D(\mathbf{P}\|\mathbf{P}_0) \le \eta \tag{12}$$

where $\mathbf{P}$ denotes the joint probability measure of a time series of input variates, $\Pi(p)$ denotes the class of all stationary time series that have one-lag dependence (i.e., Markovian) with known marginal distribution p, and $D(\mathbf{P}\|\mathbf{P}_0)$ is a nonparametric distance between $\mathbf{P}$ and the nominal i.i.d. probability measure $\mathbf{P}_0$ that is expressed in terms of Pearson's ϕ-coefficient. Note that only the lag information is assumed, and otherwise the true transition structure is assumed nonparametric. (12) thus quantifies the worst-case impact in the performance measure if a modeler ignores the input serial dependency in the nominal simulation, with the amount of ignored dependency nonparametrically denominated by the ϕ-coefficient.

Lam (2016b) analyzed the solution of (12) when η is small, and showed an expansion similar to (10) with the first-order coefficient $\xi_1 = \sqrt{Var_0(R(X_1,X_2))}$, where

$$R(X_1,X_2) = IF_2(X_1,X_2) - E_0[IF_2(X_1,X_2)|X_1] - E_0[IF_2(X_1,X_2)|X_2] + E_0[IF_2(X_1,X_2)]$$

for a suitably defined bivariate influence function IF_2 and two arbitrary consecutive states X_1 and X_2 of the i.i.d. nominal input process. Here $R(x_1,x_2)$ can be interpreted as the nonlinear interaction between X_1 and X_2 in $IF_2(X_1,X_2)$, and the interaction variance signifies the local impact of one-lag dependence. Lam (2016b) proposed a scheme using two-way analysis-of-variance to estimate $\sqrt{Var_0(R(X,Y))}$.

4.3.2 Nonparametric Assessment on Rare Event and Extremal Uncertainty

Rare-event simulation concerns the estimation of small probabilities (e.g., system overload), which depend substantially on the tail behavior of the input variates. A nonparametric assessment of the impact of input tail misspecification is often difficult because of a lack of data in the tail region. One could in principle fit extreme value distributions such as generalized Pareto (Embrechts et al. 2013), but there is a documented bias-variance tradeoff in conducting such fits, and it does not always give satisfactory accuracy.

Atar et al. (2015), Dey and Juneja (2012), Blanchet et al. (2014) and Blanchet and Murthy (2016) investigated the impacts of input misspecification in rare-event probabilities measured by KL or Renyi divergences, using formulations similar to (8). Alternately, Lam and Mottet (2015b) proposed using tail convexity as a prior belief, and solved for the worst-case probabilities built on knowledge of the non-tail region of the distribution. Lam and Mottet (2015a) combined this analysis with a heuristic iterative algorithm to compute the worst-case bound on the tail probability involving a random walk.

4.3.3 Robust Simulation under Expert Opinion

As a standard practice, when there are little or no input data, a modeler could adopt a simple distribution such as the triangular distribution, based on intuitive opinion like the support and the mean (Biller and Nelson 2002). However, since the performance measure is often a complex nonlinear function of the input distribution, it is unclear how influential the triangular distribution assumption is, even if the support and the mean are correctly specified. A modeler who is concerned about this uncertainty could instead compute the worst-case upper and lower estimates of the performance measure subject to support and mean constraints. Ghosh and Lam (2015a, 2015b) studied stochastic approximation (SA) algorithms and related gradient estimators for such type of computation. Their approach, however, requires the ad hoc tuning typically needed in SA and a discretization of the input probability distribution.

4.4 Optimization as a Substitute for Resampling: The Empirical Likelihood Method

There is a close statistical link between robust analysis and the statistical methods for nonparametric uncertainty discussed in Section 3.2. In (10), observe that by substituting $\eta = \chi^2_{1,1-\alpha}/(2n)$ where $\chi^2_{1,1-\alpha}$ is the $(1-\alpha)$-quantile of χ^2-distribution with degree of freedom 1, and setting the nominal input distribution as the empirical distribution, the first order term in the RHS of the expansion recovers the confidence bounds implied by the nonparametric delta method. In other words, the divergence-constrained optimization (9), when setting $\eta = \chi^2_{1,1-\alpha}/(2n)$ and P_0 empirically defined, is a statistically accurate upper and lower confidence bounds for the performance measure.

This connection turns out to be a consequence of the empirical likelihood (EL) method (Owen 2001), a nonparametric analog of maximum likelihood theory. Lam and Qian (2016) studied this connection and the resulting statistical guarantees in the generality of multiple independent input models. Relatedly, Lam and Zhou (2015) applied EL to quantify the uncertainty of SAA. This framework essentially gives a simulation-optimization-based approach as an alternate to the bootstrap in quantifying nonparametric input uncertainty. Using optimization instead of resampling potentially reduces the overall computational efforts and variability in the CI construction. The drawback, however, is that there is not yet a way to separate the input uncertainty from the stochastic uncertainty, and a premature halt of the optimization algorithm before convergence could lead to uninterpretable results.

5 DISCUSSION AND FUTURE DIRECTIONS

We close this tutorial by asking the following fundamental questions:

1. *Parametric vs nonparametric, or does it matter?* While obviously context-dependent, most academic works set out to *assume* a specific statistical framework. We now have an abundant literature on dealing with these assumed parametric, model and nonparametric uncertainties. We know that, in general,

nonparametric approaches alleviate the risk of model misspecification, but they are less efficient (in the statistical sense) and flexible (in the simulation modeling sense) than parametric methods. But, in a given class of problems, would we lose too much due to model bias with a parametric approach, or is it worth the effort to go with less assumptions with a nonparametric approach? How to assess these tradeoffs? The second part of this tutorial has suggested robust analysis as a potential route, but there could certainly be other approaches.

2. *Which asymptotic regime?* Though there are some exceptions, we often justify our uncertainty quantification methods based on large sample asymptotics. This forms the fundamentals in the delta method and the construction of CIs. For large-scale modeling, the input data size may not be that large relative to the system scale, and our proposed methods may not be justified either theoretically or practically. The situation could be similar to modern high-dimensional statistics, with a large model dimension p and a not-so-large sample size n.
3. *Is input uncertainty all?* Clearly not. The errors in specifying system logic can overwhelm input uncertainty and, in practice, simulation model as a whole is at best an approximation to reality. A systematic approach to calibrate and validate both input and system models is quite open.

ACKNOWLEDGMENTS

The author gratefully acknowledges support from the National Science Foundation under grants CMMI-1400391/1542020 and CMMI-1436247/1523453.

REFERENCES

Ankenman, B. E., and B. L. Nelson. 2012. "A quick assessment of input uncertainty". In *Proceedings of the 2012 Winter Simulation Conference*, edited by C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A. Uhrmacher, 1–10. Piscataway, New Jersey: IEEE.

Asmussen, S., and P. W. Glynn. 2007. *Stochastic Simulation: Algorithms and Analysis*, Volume 57. Springer Science & Business Media.

Atar, R., K. Chowdhary, and P. Dupuis. 2015. "Robust bounds on risk-sensitive functionals via Renyi divergence". *SIAM/ASA Journal on Uncertainty Quantification* 3 (1): 18–33.

Barton, R. R. 2012. "Tutorial: Input uncertainty in output analysis". In *Proceedings of the 2012 Winter Simulation Conference*, edited by R. P. O. R. C. Laroque, J. Himmelspach and A. Uhrmacher, 1–12. Piscataway, New Jersey: IEEE.

Barton, R. R., S. E. Chick, R. C. Cheng, S. G. Henderson, A. M. Law, B. W. Schmeiser, L. M. Leemis, L. W. Schruben, and J. R. Wilson. 2002. "Panel discussion on current issues in input modeling". In *Proceedings of the 2002 Winter Simulation Conference*, edited by E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes, 353–369. Piscataway, New Jersey: IEEE.

Barton, R. R., B. L. Nelson, and W. Xie. 2013. "Quantifying input uncertainty via simulation confidence intervals". *INFORMS Journal on Computing* 26 (1): 74–87.

Barton, R. R., and L. W. Schruben. 1993. "Uniform and bootstrap resampling of empirical distributions". In *Proceedings of the 1993 Winter Simulation Conference*, edited by G. W. Evans, M. Mollaghasemi, E. Russell, and W. Biles, 503–508. ACM.

Barton, R. R., and L. W. Schruben. 2001. "Resampling methods for input modeling". In *Proceedings of the 2001 Winter Simulation Conference*, edited by B. A. Peters, J. S. Smith, D. J. Medeiros, and M. W. Rohrer, Volume 1, 372–378. Piscataway, New Jersey: IEEE.

Ben-Tal, A., D. Den Hertog, A. De Waegenaere, B. Melenberg, and G. Rennen. 2013. "Robust solutions of optimization problems affected by uncertain probabilities". *Management Science* 59 (2): 341–357.

Ben-Tal, A., L. El Ghaoui, and A. Nemirovski. 2009. *Robust Optimization*. Princeton University Press.

Bertsimas, D., D. B. Brown, and C. Caramanis. 2011. "Theory and applications of robust optimization". *SIAM Review* 53 (3): 464–501.

Biller, B., and B. L. Nelson. 2002. "Answers to the top ten input modeling questions". In *Proceedings of the 2002 Winter Simulation Conference*, edited by E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes, Volume 1, 35–40. Piscataway, New Jersey: IEEE.

Biller, B., and B. L. Nelson. 2005. "Fitting time-series input processes for simulation". *Operations Research* 53 (3): 549–559.

Blanchet, J., C. Dolan, and H. Lam. 2014. "Robust rare-event performance analysis with natural non-convex constraints". In *Proceedings of the 2014 Winter Simulation Conference*, edited by A. Tolk, S. Diallo, I. Ryzhov, L. Yilmaz, S. Buckley, and J. Miller, 595–603. Piscataway, New Jersey: IEEE.

Blanchet, J., and K. R. Murthy. 2016. "On distributionally robust extreme value analysis". *arXiv preprint arXiv:1601.06858*.

Brooks, S., A. Gelman, G. Jones, and X.-L. Meng. 2011. *Handbook of Markov Chain Monte Carlo*. CRC press.

Cheng, R., and W. Holland. 1995. "The effect of input parameters on the variability of simulation output". In *Proceedings of the Second United Kingdom Simulation Society Conference*, edited by R. Cheng and R. Pooley, 29–36. EUCS Reprographics, Edinburgh University.

Cheng, R. C., and W. Holland. 1997. "Sensitivity of computer simulation experiments to errors in input data". *Journal of Statistical Computation and Simulation* 57 (1-4): 219–241.

Cheng, R. C., and W. Holland. 1998. "Two-point methods for assessing variability in simulation output". *Journal of Statistical Computation Simulation* 60 (3): 183–205.

Cheng, R. C., and W. Holland. 2004. "Calculation of confidence intervals for simulation output". *ACM Transactions on Modeling and Computer Simulation* 14 (4): 344–362.

Chick, S. E. 2001. "Input distribution selection for simulation experiments: Accounting for input uncertainty". *Operations Research* 49 (5): 744–758.

Chick, S. E. 2006a. "Bayesian ideas and discrete event simulation: Why, what and how". In *Proceedings of the 2006 Winter Simulation Conference*, edited by L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto, 96–106. Piscataway, New Jersey: IEEE.

Chick, S. E. 2006b. "Subjective probability and Bayesian methodology". *Handbooks in Operations Research and Management Science* 13:225–257.

Chick, S. E., and S. H. Ng. 2002. "Simulation input analysis: Joint criterion for factor identification and parameter estimation". In *Proceedings of the 2002 conference on Winter simulation*, edited by E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes, 400–406. Piscataway, New Jersey: IEEE.

Corlu, C. G., and B. Biller. 2013. "A subset selection procedure under input parameter uncertainty". In *Proceedings of the 2013 Winter Simulation Conference*, edited by R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, 463–473. Piscataway, New Jersey: IEEE.

Delage, E., and Y. Ye. 2010. "Distributionally robust optimization under moment uncertainty with application to data-driven problems". *Operations Research* 58 (3): 595–612.

Dey, S., and S. Juneja. 2012. "Incorporating fat tails in financial models using entropic divergence measures". *arXiv preprint arXiv:1203.0643*.

Dupuis, P., M. A. Katsoulakis, Y. Pantazis, and P. Plechac. 2015. "Path-space information bounds for uncertainty quantification and sensitivity analysis of stochastic dynamics". *arXiv preprint arXiv:1503.05136*.

Efron, B. 1992. *Bootstrap Methods: Another Look at the Jackknife*. Springer.

Embrechts, P., C. Klüppelberg, and T. Mikosch. 2013. *Modelling Extremal Events for Insurance and Finance*, Volume 33. Springer Science & Business Media.

Fan, W., L. J. Hong, and X. Zhang. 2013. "Robust selection of the best". In *Proceedings of the 2013 Winter Simulation Conference*, edited by R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, 868–876. Piscataway, New Jersey: IEEE.

Freimer, M., and L. Schruben. 2002. "Simulation input analysis: Collecting data and estimating parameters for input distributions". In *Proceedings of the 2002 Winter Simulation Conference*, edited by E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes, 393–399. Piscataway, New Jersey: IEEE.

Fu, M. C. 2006. "Gradient estimation". *Handbooks in Operations Research and Management Science* 13:575–616.

Ghosh, S., and H. Lam. 2015a. "Computing worst-case input models in stochastic simulation". *arXiv preprint arXiv:1507.05609*.

Ghosh, S., and H. Lam. 2015b. "Mirror descent stochastic approximation for computing worst-case stochastic input models". In *Proceedings of the 2015 Winter Simulation Conference*, edited by L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, 425–436. Piscataway, New Jersey: IEEE.

Glasserman, P., and X. Xu. 2013. "Robust portfolio control with stochastic factor dynamics". *Operations Research* 61 (4): 874–893.

Glasserman, P., and X. Xu. 2014. "Robust risk measurement and model risk". *Quantitative Finance* 14 (1): 29–58.

Glasserman, P., and L. Yang. 2015. "Bounding wrong-way risk in CVA calculation". *Available at SSRN 2607649*.

Glynn, P. W. 1986. "Problems in Bayesian analysis of stochastic simulation". In *Proceedings of the 1986 Winter Simulation Conference*, edited by J. Wilson, J. Henrikserl, and S. Roberts, 376–379. ACM.

Hampel, F. R. 1974. "The influence curve and its role in robust estimation". *Journal of the American Statistical Association* 69 (346): 383–393.

Hansen, L. P., and T. J. Sargent. 2008. *Robustness*. Princeton University Press.

Henderson, S. G. 2003. "Input modeling: Input model uncertainty: Why do we care and what should we do about it?". In *Proceedings of the 2003 Winter Simulation Conference*, edited by S. Chick, P. J. Sánchez, D. Ferrin, and D. J. Morrice, 90–100. Piscataway, New Jersey: IEEE.

Hu, Z., J. Cao, and L. J. Hong. 2012. "Robust simulation of global warming policies using the DICE model". *Management science* 58 (12): 2190–2206.

Hu, Z., and L. J. Hong. 2013. "Kullback-Leibler divergence constrained distributionally robust optimization". *Available at Optimization Online*.

Kennedy, M. C., and A. O'Hagan. 2001. "Bayesian calibration of computer models". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63 (3): 425–464.

Kleijnen, J. P. 1998. "Experimental design for sensitivity analysis, optimization, and validation of simulation models". *Handbook of Simulation*:173–223.

Lam, H. 2016a. "Robust sensitivity analysis for stochastic systems". *to appear in Mathematics of Operations Research*.

Lam, H. 2016b. "Sensitivity to serial dependency of input processes: A robust approach". *under minor revision in Management Science. arXiv preprint arXiv:1606.06645*.

Lam, H., and C. Mottet. 2015a. "Simulating tail events with unspecified tail models". In *Proceedings of the 2015 Winter Simulation Conference*, edited by L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, 392–402. Piscataway, New Jersey: IEEE.

Lam, H., and C. Mottet. 2015b. "Tail analysis without tail information: A worst-case perspective". *arXiv preprint arXiv:1507.03293*.

Lam, H., and H. Qian. 2016. "The empirical likelihood approach to simulation input uncertainty". *to appear in Proceedings of the 2016 Winter Simulation Conference*.

Lam, H., and E. Zhou. 2015. "Quantifying uncertainty in sample average approximation". In *Proceedings of the 2015 Winter Simulation Conference*, edited by L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, 3846–3857. Piscataway, New Jersey: IEEE.

Lee, S.-H., and P. W. Glynn. 2003. "Computing the distribution function of a conditional expectation via Monte Carlo: Discrete conditioning spaces". *ACM Transactions on Modeling and Computer Simulation* 13 (3): 238–258.

Lin, Y., E. Song, and B. Nelson. 2015. "Single-experiment input uncertainty". *Journal of Simulation* 9 (3): 249–259.

Love, D., and G. Bayraksan. 2015. "Phi-divergence constrained ambiguous stochastic programs for data-driven optimization". Technical report, The Ohio State University, Columbus.

Melamed, B., J. R. Hill, and D. Goldsman. 1992. "The TES methodology: Modeling empirical stationary time series". In *Proceedings of the 1992 Winter Simulation Conference*, edited by J. J. Swain, D. Goldsman, R. C. Crain, and J. R. Wilson, 135–144. ACM.

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

Nelson, B. 2016. "Some tactical problems in digital simulationfor the next 10 years". *Journal of Simulation* 10 (1): 2–11.

Ng, S. H., and S. E. Chick. 2006. "Reducing parameter uncertainty for stochastic systems". *ACM Transactions on Modeling and Computer Simulation* 16 (1): 26–51.

Owen, A. B. 2001. *Empirical Likelihood*. CRC press.

Pardo, L. 2005. *Statistical Inference Based on Divergence Measures*. CRC Press.

Petersen, I. R., M. R. James, and P. Dupuis. 2000. "Minimax optimal control of stochastic uncertain systems with relative entropy constraints". *IEEE Transactions on Automatic Control* 45 (3): 398–412.

Rubin, D. B. 1981. "The Bayesian bootstrap". *The Annals of Statistics* 9 (1): 130–134.

Ryzhov, I. O., B. Defourny, and W. B. Powell. 2012. "Ranking and selection meets robust optimization". In *Proceedings of the 2012 Winter Simulation Conference*, edited by C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A. Uhrmacher, 48. Piscataway, New Jersey: IEEE.

Scarf, H., K. Arrow, and S. Karlin. 1958. "A min-max solution of an inventory problem". *Studies in the Mathematical Theory of Inventory and Production* 10:201–209.

Serfling, R. J. 2009. *Approximation Theorems of Mathematical Statistics*, Volume 162. John Wiley & Sons.

Song, E., and B. L. Nelson. 2015. "Quickly assessing contributions to input uncertainty". *IIE Transactions* 47 (9): 893–909.

Song, E., B. L. Nelson, and C. D. Pegden. 2014. "Advanced tutorial: Input uncertainty quantification". In *Proceedings of the 2014 Winter Simulation Conference*, edited by A. Tolk, S. Diallo, I. Ryzhov, L. Yilmaz, S. Buckley, and J. Miller, 162–176. Piscataway, New Jersey: IEEE.

Steckley, S. G., S. G. Henderson, D. Ruppert, R. Yang, D. W. Apley, J. Staum et al. 2016. "Estimating the density of a conditional expectation". *Electronic Journal of Statistics* 10 (1): 736–760.

Van der Vaart, A. W. 2000. *Asymptotic Statistics*, Volume 3. Cambridge University Press.

Vidyashankar, A. N., and J. Xu. 2015. "Stochastic optimization using Hellinger distance". In *Proceedings of the 2015 Winter Simulation Conference*, edited by L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, 3702–3713. Piscataway, New Jersey: IEEE.

Wieland, J. R., and B. W. Schmeiser. 2006. "Stochastic gradient estimation using a single design point". In *Proceedings of the 2006 Winter Simulation Conference*, edited by L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto, 390–397. Piscataway, New Jersey: IEEE.

Xie, W., B. L. Nelson, and R. R. Barton. 2014. "A Bayesian framework for quantifying uncertainty in stochastic simulation". *Operations Research* 62 (6): 1439–1452.

Xie, W., B. L. Nelson, and R. R. Barton. 2016. "Multivariate input uncertainty in output analysis for stochastic simulation". *working paper*.

Zouaoui, F., and J. R. Wilson. 2003. "Accounting for parameter uncertainty in simulation input modeling". *IIE Transactions* 35 (9): 781–792.

Zouaoui, F., and J. R. Wilson. 2004. "Accounting for input-model and input-parameter uncertainties in simulation". *IIE Transactions* 36 (11): 1135–1151.

AUTHOR BIOGRAPHY

HENRY LAM is an Assistant Professor in the Department of Industrial and Operations Engineering at the University of Michigan, Ann Arbor. His email address is khlam@umich.edu.