Proceedings of the 2001 Winter Simulation Conference
B. A. Peters, J. S. Smith, D. J. Medeiros, and M. W. Rohrer, eds.

REDUCING INPUT PARAMETER UNCERTAINTY FOR SIMULATIONS

Szu Hui Ng
Stephen E. Chick

Department of Industrial and Operations Engineering
The University of Michigan
1205 Beal Avenue
Ann Arbor, Michigan 48109-2117, U.S.A.

ABSTRACT

Parameters of statistical distributions that are input to simulations are typically not known with certainty. For existing systems, or variations on existing systems, they are often estimated from field data. Even if the mean of simulation output were estimable exactly as a function of input parameters, there may still be uncertainty about the output mean because inputs are not known precisely. This paper considers the problem of deciding how to allocate resources for additional data collection so that input uncertainty is reduced in a way that effectively reduces uncertainty about the output mean. The optimal solution to the problem in full generality appears to be quite challenging. Here, we simplify the problem with asymptotic approximations in order provide closed-form sampling plans for additional data collection activities. The ideas are illustrated with a simulation of a critical care facility.

1 INTRODUCTION

Simulations of design proposals for nonexistent systems can help a modeler evaluate the performance of the system as a function of a range of design and statistical input parameters. Simulations of existing systems, or variations of existing systems, can similarly help modelers evaluate design proposals and inform decisions to improve system performance.

An advantage of a system that already exists is that field data may be available to estimate statistical input parameters. Statistical software is widely used to identify an input distribution and parameter to describe demand; processing times; failure and repair distributions; routing probabilities; and other distributions that determine system performance (Law and Kelton 2000).

However, input parameter estimates based on field data are subject to random variation because the field data are observations of random phenomena. Barton and Schruben (1993, 2001) provide compelling examples that show how ignoring input parameter uncertainty can significantly reduce the coverage of confidence intervals for the mean.

The problem is easy to identify.

Simulations that use a single parameter estimate can model the stochastic randomness for the given input parameters. But they do not model the uncertainty associated with not knowing the right parameter (Kleijnen 1994). Cheng and Holland (1997) use classical statistical techniques to approximate the output uncertainty of the mean simulation output as a function of the input uncertainty. An alternate method is to account for uncertainty in input parameters using a Bayesian model average, or BMA (Draper 1995). Chick (2001) addresses implementation issues for the BMA in simulation experiments of queuing systems. The idea is to run multiple replications, with input parameters sampled before replications to represent input uncertainty. Both bootstrap sampling (Cheng and Holland 1997; Barton and Schruben 2001) and Bayesian posterior distributions (Chick 2001) have been proposed for input sampling.

While the BMA and bootstrap sampling indicate how to quantify structural uncertainty, they do not show how to reduce input parameter uncertainty and its effect on estimates of the mean system performance.

This article illustrates one way to allocate resources to collect additional field data to reduce input uncertainty in a way that optimally reduces uncertainty about the mean system performance. We assume that there are several sources of structural uncertainty, such as imprecisely known demand and service rates. But uncertainty about parameters can be reduced by collecting additional field data (such as actual demand and service times).

We assume that additional field data can be collected to better infer the input parameters, and the goal is to plan data collection activities to reduce input uncertainty in a way that effectively reduces output uncertainty. Here, uncertainty is quantified as the variance of an unknown quantity. Our approach uses asymptotic normality approximations for the input parameter uncertainty, and the mean output performance is estimated by a linear approximation. This may

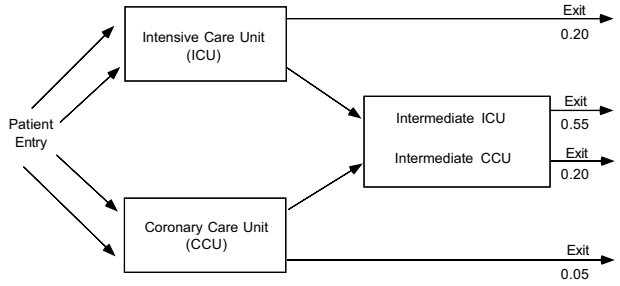

Figure 1: Fraction of patients routed through different units of a critical care facility.

cause suboptimal performance if the actual mean system response is highly nonlinear over the range of likely input parameter values. However, the approach is very general and can give guidance for data collection activities in a variety of situations.

A critical care facility simulation illustrates the ideas.

2 EXAMPLE: CRITICAL CARE FACILITY

The critical care facility illustrated in Figure 1 was originally studied by Schruben and Margolin (1978). Patients arrive according to a Poisson process and are routed through the system depending upon their specific health condition. Stays in the intensive care (ICU), coronary care (CCU), and intermediate care units are presumed to be lognormally distributed.

Schruben and Margolin (1978) studied how to allocate random number streams to reduce variability in response surface parameter estimates. Their response model predicts the expected number of patients per month $E[Y]$ that are denied entry to the facility as a function of the number of beds in the ICU, CCU, and intermediate care units. They presume fixed point estimates for $k = 6$ input parameters, one per source of randomness, to describe the patient arrival process (Poisson arrivals, mean $\hat{\lambda} = 3.3$/day), ICU stay duration (lognormal, mean 3.4 and standard deviation 3.5 days), intermediate ICU stay duration (lognormal, mean 15.0, standard deviation 7.0), intermediate CCU stay duration (lognormal, mean 17.0, standard deviation 3.0), CCU stay duration (lognormal, mean 3.8, standard deviation 1.6), and routing probabilities (multinomial, $\hat{p_1} = 0.2$, $\hat{p_3} = 0.2$, $\hat{p_4} = 0.05$). Some parameters are multivariate, and there are a total of $K = 1 + 4 * 2 + 3 = 12$ dimensions of parameters.

The actual system parameters are not known with certainty, and the estimated system performance will be in error if the actual parameter values differ from their point estimates. Here we fix the number of beds in each of the three units (14 in ICU, 5 in CCU, 16 in intermediate care), and study how the expected number of patients per month that are denied entry depends on the unknown parameters. Parameter uncertainty is modeled by taking the above point estimates, presuming that they are maximum likelihood estimates (MLEs) based on data from 100 patients, then using a Bayesian approach to model parameter uncertainty described in Section 3.

Section 4 describes an asymptotic relationship between the input parameter uncertainty and uncertainty in the mean system performance. If the mean performance is a known function of the input parameters, equations in Section 4.1 can be employed to suggest how many additional data points should be collected for each source of randomness in a way that efficiently reduces output uncertainty. In many cases, including the simulation model here, the response function is unknown and must be estimated. Section 4.2 uses Bayesian metamodeling techniques to select a simple linear response model for the system. The metamodel is then used as a surrogate for the actual response model. Section 5 presents numerical results for the critical care facility and specific suggestions for further data collection.

3 ASYMPTOTIC APPROXIMATIONS FOR INPUT UNCERTAINTY

There may be several sources of randomness in a system. The critical care facility has six sources of randomness modeled by the Poisson, lognormal, and multinomial distributions.

However, the specific values of the input parameters are unknown. For instance, the number of arrivals per day is Poisson, but λ is unknown. The value of the parameters are inferred from prior information and from field data. We presume that observations satisfy the standard independence assumptions. As the number of observations increases, uncertainty about the parameter's value decreases.

The distributions used in the critical care facility simulation satisfy asymptotic normality properties. The normality is with respect to the mode of the posterior distribution of the unknown parameter and the Bayesian observed information matrix. We exploit this property to approximate the posterior distribution of each parameter. We further use the expected information of additional observations to approximate the posterior distributions if a few more data points were collected.

A motivation for the asymptotic approximations is presented in Section 3.1. The formalism for the Poisson distribution is then presented in Section 3.2. The lognormal is treated in Section 3.3. The multinomial distribution is handled similarly, see (Chick and Ng 2001) for details.

3.1 Analogy to Output with Normal Distribution

One way to think about the approximations below for input parameter uncertainty is by analogy with the confidence interval (CI) size for the mean of output from terminating

simulations. The result is a well-known approximation for the number of additional replications required to obtain a CI of a given size.

Suppose that simulation output $y_1, \ldots, y_n$ is observed for a fixed set of inputs, with sample mean $\bar{y}_n$ and sample variance S_n^2. Then the $100(1-\alpha)\%$ CI for the mean is

$$\bar{y}_n \pm t_{n-1,1-\alpha/2} \left(\frac{S_n^2}{n} \right)^{1/2}.$$

Law and Kelton (2000, Sec 9.4.1) present an *approximate* expression for the number of additional replications m^* required to achieve an absolute error of δ is

$$m^* = \min \left\{ m \geq 0 : t_{n+m-1,1-\alpha/2} \left(\frac{S_n^2}{n+m} \right)^{1/2} \leq \delta \right\}.$$

For large n, $t_{n+m-1,1-\alpha/2}$ does not change much as a function of m. The size of the CI essentially scaled by presuming the variance shrinks from S_n^2/n to

$$S_n^2/(n+m). \tag{1}$$

A Bayesian analog for this approximation is given to describe input parameter uncertainty, where n represents the amount of information (prior and data points) that is available, and m is the number of additional observations to make for a particular source of uncertainty.

3.2 Poisson Arrivals

Suppose that number of arrivals per day $x_1, \ldots, x_n$ have a Poisson distribution with unknown mean λ, and that observations are conditionally independent given λ. The probability mass function (pmf) is

$$\Pr(x_1, \ldots, x_n \mid \lambda) = \prod_{i=1}^{n} e^{-\lambda} \frac{\lambda^{x_i}}{x_i!}.$$

A common choice of probability model for $\Lambda = \lambda$ is a gamma distribution Gamma (α, β) with probability density function (pdf)

$$f(\lambda) = c\lambda^{\alpha-1} e^{-\beta\lambda}.$$

The usefulness of the gamma distribution for λ in large part comes from the fact that a gamma prior distribution results in a gamma posterior distribution for λ.

Suppose now that prior information and field data combine to result in a Gamma (α, β) posterior distribution for λ, with $\alpha > 1$. The normal distribution Normal $\left(\tilde{\lambda}, \Sigma\right)$ approximation to a Gamma (α, β) distribution (Bernardo and Smith 1994) has a mean equal to the mode $\tilde{\lambda} = (\alpha - 1)/\beta$ of the distribution of λ, and a variance Σ that is the inverse of the Bayesian analog of the observed information,

$$\begin{aligned} \Sigma^{-1} &= -\frac{d^2 \log f(\lambda \mid x_1, \ldots, x_n)}{d\lambda^2}\bigg|_{\tilde{\lambda}} \\ &= \frac{\alpha - 1}{\tilde{\lambda}^2} = \frac{\beta^2}{\alpha - 1}. \end{aligned}$$

The variance, colloquially, is roughly 1/information. This normal distribution approximation is better for larger values of α. Here, α grows with the number of arrivals $\sum_{i=1}^{n} x_i$, and β grows with n.

How would the distribution of λ change if m more observation are to be made? The expected value of information of one more day of arrival information is

$$\begin{aligned} H(\tilde{\lambda}) &= E_X \left(-\frac{d^2 \log f(X \mid \lambda)}{d\lambda^2} \right)\bigg|_{\tilde{\lambda}} \\ &= \frac{1}{\tilde{\lambda}} = \frac{\beta}{\alpha - 1}. \end{aligned} \tag{2}$$

The analog to the approximation in Equation 1 to assess the effect of m additional observations is to change Σ to

$$(\Sigma^{-1} + mH(\tilde{\lambda}))^{-1} = \frac{(\alpha - 1)/\beta}{\beta + m}. \tag{3}$$

To see the analogy with Equation 1, recall that β grows at the same rate as n, and that the variance of observations given $\tilde{\lambda}$ is $(\alpha - 1)/\beta$.

Specific data for the arrival rate to the critical care facility was not published. We therefore presumed a standard noninformative prior distribution $f(\lambda) \propto \lambda^{-1/2}$ for the arrival rate λ, and presumed that the point estimate $\hat{\lambda} = 3.3$ is an MLE based on 100 patient arrivals. This results in $\alpha = 100.5$, $\beta = 30.3$, $\tilde{\lambda} = 3.28$, $E[\lambda] = 3.32$, $\Sigma^{-1} = 9.23$, and $H(\tilde{\lambda}) = 0.3045$.

The data collection activities suggested below in Section 5 are relatively insensitive to small perturbations in these assumptions. For example, if there were 99 arrivals in 30 days, then $\alpha = 99.5$, $\beta = 30$, and the resulting analysis and conclusions would be quite similar.

3.3 Lognormal Service Times

The process for Poisson arrivals can be adapted for lognormal service times. The logarithm of a random variable X with lognormal distribution has a normal distribution with mean μ and precision $\tau = \sigma^{-2}$. The likelihood function is

$$f(x_1, \ldots, x_n \mid \mu, \tau) = \prod_{i=1}^{n} \frac{\tau^{1/2}}{\sqrt{2\pi}} e^{(\log x_i - \mu)^2 \tau/2}.$$

A common choice of probability model for an unknown $\theta = (\mu, \tau)$ is the normal-gamma distribution (Bernardo and Smith 1994). The normal-gamma distribution has parameters $\mu_0, s_0, \alpha_0, \beta_0$,

$$f(\mu, \tau) = \frac{(s_0\tau)^{1/2}}{\sqrt{2\pi}} e^{(\mu-\mu_0)^2 s_0 \tau/2} \cdot \frac{{\beta_0}^{\alpha_0}}{\Gamma(\alpha_0)} \tau^{\alpha_0 - 1} e^{-\beta_0 \tau}.$$

If the prior for (μ, τ) is a normal-gamma distribution, then the posterior distribution is also a normal-gamma distribution, with parameters μ_n, $s_0 + n$, $\alpha_0 + n/2$, β_n, where

$$\mu_n = \left(s_0\mu_0 + \sum_{i=1}^{n} \log x_i \right) \Big/ (s_0 + n)$$

is a weighted mean of the logarithms of the observations,

$$ns^2 = \sum_{i=1}^{n} \left(\log x_i - \sum_{i=1}^{n} \log x_i / n \right)^2$$

and $\beta_n = \beta_0 + ns^2/2 + s_0 n(\mu_0 - \sum_{i=1}^{n} \log x_i/n)/(s_0 + n)$.

The noninformative prior distribution $f(\mu, \tau) \propto \tau^{-1}$ results in a normal-gamma posterior distribution with parameters $\mu_n = \sum_{i=1}^{n} \log x_i/n$, $s_n = n$, $\alpha_n = (n-1)/2$, and $\beta_n = ns^2/2$. The mode of the variance is then $\sigma_n^2 = ns^2/(n-1)$, the usual estimator for the variance.

Suppose that the prior distribution and data combine to give a normal-gamma distribution for $\theta = (\mu, \tau)$ with parameters $\mu_n, s_n, \alpha_n, \beta_n$. Then $\tilde{\mu} = \mu_n$ and $\tilde{\tau} = (\alpha_n - 1)/\beta_n$. The Bayesian observed information matrix is

$$\begin{aligned} \Sigma^{-1} &= -\frac{\partial^2 \log f(\theta \mid x_1, \ldots, x_n)}{\partial \theta^2}\bigg|_{\tilde{\theta}} \\ &= \begin{bmatrix} s_n \frac{\alpha_n - 1}{\beta_n} & 0 \\ 0 & \left(\alpha_n - \frac{1}{2}\right)\left(\frac{\beta_n}{\alpha_n - 1}\right)^2 \end{bmatrix}. \end{aligned}$$

And the expected information from a single observation is

$$\begin{aligned} H(\tilde{\theta}) &= E_X\left(-\frac{\partial^2 \log f(X \mid \theta)}{\partial \theta^2} \right)\bigg|_{\tilde{\theta}} \\ &= \begin{bmatrix} \frac{\alpha_n - 1}{\beta_n} & 0 \\ 0 & \frac{1}{2}\left(\frac{\beta_n}{\alpha_n - 1}\right)^2 \end{bmatrix}. \end{aligned}$$

In this case, Σ^{-1} and $H(\tilde{\theta})$ are not proportional, as $\Sigma_{11}^{-1} = nH(\tilde{\theta})_{11}$, but $\Sigma_{22}^{-1} = (n-2)H(\tilde{\theta})_{22}$.

Specific data for the four types of lognormal service times at the critical care facility was not published. We presumed the noninformative prior distribution, and that the point estimates for the mean and standard deviation published in Schruben and Margolin (1978) were MLEs based on a total of 100 patients through the system (e.g., 75 observations at ICU, 55 observations at intermediate ICU). We solved for the sufficient statistics that would lead to the published MLEs. At ICU, for example, the estimated mean of $3.4 = e^{\mu_n + \sigma_n^2/2}$ and variance $3.5 = e^{2\mu_n + \sigma_n^2}(e^{\sigma_n^2} - 1)$ result in $\mu_n = 0.8625$ and $\beta_n^2 = 0.7195$.

4 INPUT TO OUTPUT UNCERTAINTY

Chick (1997) proposed general ideas for a Bayesian framework to relate simulation input parameters and output parameters. Here, we discuss some specific asymptotic approximations that extend that framework.

The mean simulation response $E[Y \mid \boldsymbol{\theta}] = g(\boldsymbol{\theta})$ is a function of k input distribution parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$, where each θ_i is an input parameter (potentially multivariate). The output of the r-th replication is denoted

$$Y_r = g(\boldsymbol{\theta}) + \sigma Z_r,$$

where σ is a standard deviation and Z_r is sequence of i.i.d. zero mean, unit variance, random variables. In general, σ and the distribution of Z_r may depend on $\boldsymbol{\theta}$.

The parameters used for the critical care facility are the arrival rate $\theta_1 = \lambda$; the mean and precision of the logarithms of the ICU service times $\theta_2 = (\mu_2, \tau_2)$, with $\tau_2 = 1/\sigma_2^2$, similar parameters for the logarithms of the intermediate ICU service times $\theta_3 = (\mu_3, \tau_3)$, intermediate CCU service times $\theta_4 = (\mu_4, \tau_4)$, and CCU service times $\theta_5 = (\mu_5, \tau_5)$; and finally the multinomial routing probabilities $\theta_6 = (p_1, p_3, p_4)$.

4.1 Known Response Function

Suppose that the mean simulation response $E[Y] = g(\boldsymbol{\theta})$ were a known function of the parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$. Assume for the moment that n observations are available from each of the sources of randomness.

Informally, the posterior distribution of $\boldsymbol{\theta}$, given n observations, is asymptotically normal as the number of observations n increases, with mean $\tilde{\boldsymbol{\theta}}_n$ and variance/covariance matrix Σ_n, where Σ_n^{-1} is the Bayesian analog of the observed information matrix, as above in Section 3. Bernardo and Smith (1994, Sec 5.3.2) provide a more formal statement and technical conditions for the assertion to hold.

The distribution of $\boldsymbol{\theta}$ induces a distribution on $g(\boldsymbol{\theta})$. We approximate $g(\boldsymbol{\theta})$ in the neighborhood of $\tilde{\boldsymbol{\theta}}$ with a local linear approximation. A linear transformation of normally distributed random variables is also a normal random variable. Let $\partial g(\tilde{\boldsymbol{\theta}}_n)/\partial\theta_i$ be the gradient of the response with respect to the i-th input parameter, evaluated at $\tilde{\boldsymbol{\theta}}_n$. That gradient is a vector if θ_i is multivariate. Set

$$\nabla g(\boldsymbol{\theta}) = [\partial g(\boldsymbol{\theta})/\partial\theta_1 \ldots \partial g(\boldsymbol{\theta})/\partial\theta_k].$$

If the distribution of $\boldsymbol{\theta}$ is approximated as a Normal $\left(\tilde{\boldsymbol{\theta}}, \boldsymbol{\Sigma}\right)$ distribution, and $g(\boldsymbol{\theta})$ is linear, then $g(\boldsymbol{\theta})$ is distributed

$$\text{Normal}\left(g(\tilde{\boldsymbol{\theta}}), \nabla g(\tilde{\boldsymbol{\theta}})\boldsymbol{\Sigma}\nabla^{\mathrm{T}} g(\tilde{\boldsymbol{\theta}})\right).$$

The number of observations for each of the k source of randomness may differ, as may the number of additional data points to collect. Suppose that n_i is the equivalent number of data points for source of randomness i, resulting in information Σ_{i,n_i}. A simple variation on the proposition indicates that the (asymptotic) approximation to the output variance as a function of the input uncertainty is

$$\sum_{i=1}^{k} \frac{\partial g(\tilde{\boldsymbol{\theta}}_n)}{\partial \theta_i} \Sigma_{i,n_i} \frac{\partial g(\tilde{\boldsymbol{\theta}}_n)^{\mathrm{T}}}{\partial \theta_i}.$$

Further suppose that m_i additional points are to be collected for source of randomness i. Using the approximations of Section 3 for input parameter uncertainty, the output variance after collecting additional information is approximately

$$V_{in} = \sum_{i=1}^{k} \frac{\partial g(\tilde{\boldsymbol{\theta}}_n)}{\partial \theta_i} (\Sigma_{i,n_i}^{-1} + m_i H(\tilde{\theta}_n))^{-1} \frac{\partial g(\tilde{\boldsymbol{\theta}}_n)^{\mathrm{T}}}{\partial \theta_i}. \quad (4)$$

If each observed information matrix is proportional to the corresponding expected information matrix (say, $n_i H(\tilde{\theta}_{n_i}) = \Sigma_{i,n_i}^{-1}$), then Equation 4 simplifies to

$$V_{in} = \sum_{i=1}^{k} \frac{\partial g(\tilde{\boldsymbol{\theta}}_n)}{\partial \theta_i} \frac{H^{-1}(\tilde{\theta}_{n_i})}{n_i + m_i} \frac{\partial g(\tilde{\boldsymbol{\theta}}_n)^{\mathrm{T}}}{\partial \theta_i} \quad (5)$$

One formal optimization problem for reducing this asymptotic approximation to output variance can be formulated by assuming that sampling costs for source i is linear in the number of samples, $c_i m_i$, and solving:

$$\begin{array}{lll} \min & V_{in} & \\ \text{s.t.} & m_i \geq 0 & \text{for } i = 1, \ldots, k \\ & \sum c_i m_i = b, & \end{array} \quad (6)$$

where b is the sampling budget. Special features of the data collection process can be handled by adding constraints. For instance, if service times and routing decisions are collected and reported together, then a constraint can be added to require that the corresponding m_is be equal.

Proposition 1. *If V_{in} simplifies to Equation 5, and the integer restriction is relaxed (let the m_i be continuous), and b is large, then the solution to Equation 6 is:*

$$m_i^* = \frac{b + \sum_{\ell=1}^{k} n_\ell c_\ell}{\sum_{j=1}^{k} \left(\frac{\xi_j c_i c_j}{\xi_i}\right)^{1/2}} - n_i \quad (7)$$

where

$$\xi_i = \frac{\partial g(\tilde{\boldsymbol{\theta}}_n)}{\partial \theta_i} H^{-1}(\tilde{\theta}_{n_i}) \frac{\partial g(\tilde{\boldsymbol{\theta}}_n)^{\mathrm{T}}}{\partial \theta_i}.$$

Proof. See Chick and Ng (2001). □

For small b, the nonnegativity constraints for the m_i need consideration. If V_{in} does not simplify to Equation 5, then the solution to Equation 6 can be computed numerically. That optimization was implemented with a spreadsheet solver for this paper.

If $g(\boldsymbol{\theta})$ is not linear, then the naive estimator of $E_{\boldsymbol{\theta}}[g(\boldsymbol{\theta})]$ can be biased (further, the variance and expectation might not exist!). While a Taylor series expansion correction can correct for first order bias when the expectation exists, we do not do so here. The current proposal seeks to reduce output variance, but does not consider mean squared error. When the variance does not exist, we reduce an asymptotic approximation to the variance induced by a linear metamodel for $g(\cdot)$ in the neighborhood of $\tilde{\boldsymbol{\theta}}$.

4.2 Unknown Response Function

The response function $g(\boldsymbol{\theta})$ is not known for the critical care facility simulation. This is typical of many simulation projects. An approximation to the response function is therefore needed. Since the asymptotic approximations introduced above 'hit their limit' in some sense with linear approximations, we only seek a reasonable linear approximation to $g(\boldsymbol{\theta})$ in the region around the most likely values of the unknown parameters, $\tilde{\boldsymbol{\theta}}_n$.

There are several approaches to approximating the gradient near $\tilde{\boldsymbol{\theta}}_n$. Any might be appropriate.

Here we use a Bayesian model average to identify a reasonable response surface metamodel, using the general approach of Raftery, Madigan, and Hoeting (1997). Namely, we consider 2^K candidate linear metamodels, if there are K dimensions of parameters. Each metamodel contains an intercept term, but is distinguished by the presence or absence of a linear coefficient for each dimension of the parameter. For the critical care facility, the $k = 6$ parameters have a total of $K = 12$ dimensions. This results in $2^{12} = 4096$ linear response metamodels of the form:

$$\begin{aligned} Y &= \beta_0 + \sigma Z \\ Y &= \beta_0 + \beta_1 \lambda + \sigma Z \end{aligned}$$

Table 1: The five most likely response models. A 1 indicates the presence of a regressor in the linear model, a 0 indicates absence. 'Post' is the posterior probability of the model given the data. The most likely linear metamodel is $Y = \beta_0 + \beta_1\lambda + \beta_4\mu_3 + \beta_5\tau_3 + \beta_6\mu_4 + \beta_{10}p_1 + \beta_{11}p_3 + \beta_{12}p_4 + \sigma Z$.

λ	μ_2	τ_2	μ_3	τ_3	μ_4	τ_4	μ_5	τ_5	p_1	p_3	p_4	Post
1	0	0	1	1	1	0	0	0	1	1	1	0.692
1	1	0	1	1	1	0	0	0	1	1	1	0.069
1	0	0	1	1	1	0	1	0	1	1	1	0.059
1	0	0	1	1	0	0	0	0	1	1	1	0.037
1	0	0	1	1	1	0	0	1	1	1	1	0.027

$$\begin{aligned} Y &= \beta_0 + \beta_2\mu_2 + \sigma Z \\ Y &= \beta_0 + \beta_3\tau_2 + \sigma Z \\ &\vdots \\ Y &= \beta_0 + \beta_1\lambda + \beta_2\mu_2 + \ldots + \beta_{12}p_4 + \sigma Z \end{aligned}$$

The Z are presumed to be standard normal random variables.

Here we use the same formulation for prior probability distributions for the unknown metamodel form and vector of βs that were used by Raftery, Madigan, and Hoeting (1997). Each metamodel is given prior probability of 2^{-K}, and the probability for the vector of βs, given the metamodel, is chosen to be proper yet relatively noninformative. See Raftery, Madigan, and Hoeting (1997) for further details, an argument for the reasonableness of the approach, and numerical experiments that show good performance in some sense. Chick (2000) used the same approach for inferring the response surface of an M/M/1 queue, with some numerical evidence that shows that the correct response can be identified even though the homoscedasticity assumption is violated.

Simulation replications with input/output realizations $(\boldsymbol{\theta}_r, y_r)$ are used to infer the marginal probability that a given metamodel is best and to estimate the βs. We sample the $\boldsymbol{\theta}_r$ from the Bayesian posterior distributions (given field data) determined above in Section 3. This focuses attention to the likely values of $\boldsymbol{\theta}$.

The overall output variance consists of input uncertainty terms (like V_{in}) that decrease with the amount of field data, and uncertainty about the βs and σ decreases with the number of replications.

5 ANALYSIS: CRITICAL CARE UNIT

We use simulation to estimate how input parameter uncertainty results in uncertainty about the mean number of patients per month that are denied entry to the critical care unit described in Section 2. Input parameter uncertainty is approximated as in Section 3. We then used the Bayesian model average (Draper 1995; Chick 2001) to sample 32 independent sets of input parameters for the critical care simulation. For each input parameter, we ran 4 independent replications of the critical care unit, for a total of $r = 128$ replications. Each replication simulated 50 months of operation (after a 10 month warm-up period).

The input-output combinations were then used to select a linear metamodel to represent the local behavior of the mean response in the area of the most likely parameter values, as described in Section 4. The posterior probability for each of the 2^{12} metamodels was computed, and the most likely five response metamodels are listed in Table 1. The column 'Post' gives the posterior probability for the given metamodel. The top 5 metamodels accounted for 88.4% of the probability. The presence of a 1 under each input parameter indicates that a linear factor with that parameter is included in the model. A 0 indicates that the factor is less important than stochastic noise. The most likely linear model, based on the observed output, is $Y = \beta_0 + \beta_1\lambda + \beta_4\mu_3 + \beta_5\tau_3 + \beta_6\mu_4 + \beta_{10}p_1 + \beta_{11}p_3 + \beta_{12}p_4 + \sigma Z$, where σ is the standard deviation for the random noise in the output. Point estimates (posterior modes) of the βs are $\tilde{\beta}_1 = 28.97$, $\tilde{\beta}_4 = 29.5$, $\tilde{\beta}_5 = -0.63$, $\tilde{\beta}_6 = 11.8$, $\tilde{\beta}_{10} = -56.2$, $\tilde{\beta}_{11} = -10.5$, $\tilde{\beta}_{12} = -46.72$, and $\tilde{\sigma} = 1.6$. We set $\tilde{\beta}_i = 0$ for all 'unimportant' β_i. Estimates of V_{in} presumed that the response function is this most likely metamodel, and that gradients were given by the $\tilde{\beta}_i$.

The approximation to uncertainty about the mean performance that is due to input uncertainty is $\sqrt{V_{in}} = 10.0$. Uncertainty about the mean performance due to stochastic variation in simulations is about $\hat{\sigma}/\sqrt{r} = 0.14$, or approximately $\sqrt{V_{in}}/70$. For this study, then, input uncertainty results in a much greater uncertainty about the performance of the system than stochastic uncertainty associated with random outcomes in the simulation. We do not report uncertainty in the output mean due to uncertainty about the βs.

The most likely model indicates that the parameters for the arrival rate, the intermediate ICU, the intermediate CCU, and the routing probabilities are the most important, in terms of influence on the output uncertainty. We then used the asymptotic approximations in Section 3 to identify

Table 2: Asymptotic approximation for the reduction in output uncertainty V_{in} due to input uncertainty, as a function of the number of observations b that can be collected. Also shown for each b is the optimal number n_1 of days of arrival data, n_3 intermediate ICU times, n_4 intermediate CCU times, and n_6 routing decisions.

Budget, b	n_1	n_3	n_4	n_6	V_{in}
0	0	0	0	0	100.0
100	100	0	0	0	30.3
400	313	37	0	50	13.9
1000	661	131	9	199	7.0

how many additional data points should be collected in those areas, as a function of the sampling budget b.

Table 2 describes how many data points in each of those areas should be collected. The table assumes that a total of b data points can be collected, and that the collection cost for each area is the same ($c_i = 1$ for $i = 1, \ldots, k$). Clearly the most important area for additional data collection, based on this analysis, is the arrival rate. The routing probabilities also appear to be relatively important as well.

6 CONCLUSIONS

Input distribution selection and output analysis are often treated separately in discrete-event simulation education and practice. Sensitivity and uncertainty analysis are areas where the two topics meet. The premise of this paper is that input uncertainty results in uncertainty about the mean performance of a system, and that additional data collection activities can be a mechanism for reducing input uncertainty. This in turn reduces output uncertainty. Analytical solutions to the problem of identifying optimal data collection plans for arbitrary stochastic systems seems to be a difficult problem. The approach in this paper is to use asymptotic approximations for uncertainty, and to estimate gradient information, in order to provide allocations of resources for additional data collection.

An advantage of this approach is its apparent broad generality. This paper dealt with the Poisson, lognormal and multinomial distributions, but the same idea seems to be generalizable to other members of the regular exponential family of distributions. The mean response is estimated from simulation output when it is not known. We used a specific Bayesian technique to estimate the gradient, but in principal any technique for estimating the gradient can be used.

There are several considerations, however, that warrant further exploration. A more formal mathematical formulation for the regular exponential family should be given. The sensitivity of the data collection allocation to the gradient estimation scheme should be evaluated. Linear approximations to the mean of nonlinear functions of input parameters may result in bias, particularly when the number of observed data points is small.

We note that some expectations, much less the variance, might not exist when independent conjugate prior distributions are used (see the M/M/1 example in Chick 2001). However, the nonexistence of some of those moments may be avoided by modifying the model to better reflect reality (assume capacitated queues, run transient rather than steady state simulations) . The asymptotic approximation might not reflect the actual distribution well when there are few observations. And in some situations, Bayesian representations of input uncertainty are based on expert opinion and have no obvious sampling mechanism for further inference.

Nonetheless, the formulas in this paper present an implementable mechanism for guiding data collection plans to reduce input uncertainty in a way that in some sense effectively reduces output uncertainty.

REFERENCES

Barton, R. R., and L. W. Schruben. 1993. Uniform and bootstrap resampling of empirical distributions. In *Proceedings of the Winter Simulation Conference*, ed. G. W. Evans, M. Mollaghesemi, E. C. Russell, and W. E. Biles, 503–508. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Barton, R. R., and L. W. Schruben. 2001. Simulating real systems. *in submission*.

Bernardo, J. M., and A. F. M. Smith. 1994. *Bayesian theory*. Chichester, UK: Wiley.

Cheng, R. C. H., and W. Holland. 1997. Sensitivity of computer simulation experiments to errors in input data. *Journal on Statistical Compututing and Simulation* 57:219–241.

Chick, S. E., and S.-H. Ng. 2001. Input parameter uncertainty reduction for stochastic systems. Technical report, University of Michigan, Ann Arbor, Dept. of Industrial & Operations Engineering.

Chick, S. E. 1997. Bayesian analysis for simulation input and output. In *Proceedings of the Winter Simulation Conference*, ed. S. Andradóttir, K. Healy, D. Withers, and B. Nelson, 253–260. Piscataway, NJ: Institute of Electrical and Electronics Engineers, Inc.

Chick, S. E. 2000. Bayesian methods for simulation. In *Proceedings of the Winter Simulation Conference*, ed. J. Joines, R. Barton, P. Fishwick, and K. Kang, 109–118. Piscataway, NJ: Institute of Electrical and Electronics Engineers, Inc.

Chick, S. E. 2001. Input distribution selection for simulation experiments: Accounting for input uncertainty. *Operations Research*:to appear.

Draper, D. 1995. Assessment and propogation of model uncertainty (with discussion). *Journal of the Royal Statistical Society, Series B* 57 (1): 45–97.

Kleijnen, J. P. C. 1994. Sensitivity analysis versus uncertainty analysis: When to use what? In *Predictability and Nonlinear Modeling in Natural Sciences and Economics*, ed. J. Grasman and G. van Straten. Dordrecht: Kluwer Academic Publishers.

Law, A. M., and W. D. Kelton. 2000. *Simulation modeling & analysis*. 3d ed. New York: McGraw-Hill, Inc.

Raftery, A. E., D. Madigan, and J. A. Hoeting. 1997. Bayesian model averaging for linear regression models. *Journal of the American Statistical Association* 92 (437): 179–191.

Schruben, L. W., and B. H. Margolin. 1978. Pseudorandom number assignment in statistically designed simulation and distribution sampling experiments. *Journal of the American Statistical Association* 73 (363): 504–525.

AUTHOR BIOGRAPHIES

STEPHEN E. CHICK is an assistant professor of Industrial and Operations Engineering at the University of Michigan, Ann Arbor. In addition to stochastic simulation, his research focuses on Bayesian statistics, public health, decision analysis and computational methods in statistics. His research is motivated by projects in manufacturing, operations, and health care. His work experience includes several years of materials handling system design for the automotive industry using simulation analysis. His e-mail address is <sechick@engin.umich.edu>, and his web page is <http://www-personal.engin.umich.edu/~sechick/>.

SZU HUI NG is a Ph.D. candidate in the Department of Industrial and Operations Engineering at the University of Michigan, Ann Arbor. Her interests include stochastic simulation, design of experiments, and Bayesian computation. Her e-mail address is <nszuhui@engin.umich.edu>.