Proceedings of the 2000 Winter Simulation Conference
J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, eds.

BAYESIAN METHODS FOR SIMULATION

Stephen E. Chick

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

ABSTRACT

This tutorial describes some ways that Bayesian methods address problems that arise during simulation studies. This includes quantifying uncertainty about input distributions and parameters, sensitivity analysis, and the selection of the best of several simulated alternatives. Focus is on illustrating the main ideas and their relevance to practical problems. Numerous citations for both introductory and more advanced material provide a launching pad into the Bayesian literature.

1 INTRODUCTION

It is easy to argue that Bayesian methods don't apply at all to simulation, if simulation is taken to be a numerical analysis tool for studying complex systems. However, if simulation is considered to be an analytical tool that informs decisions, then Bayesian methods are relevant and appropriate. Bayesian methods for simulation input and output uncertainty have been around for some time, but have been increasingly applied and developed in recent years (a partial list is Glynn 1986; Cooke 1994; Chen and Schmeiser 1995; Chen 1996; Scott 1996; Andradóttir and Bier 1997; Chick 1997; Nelson, Schmeiser, Taaffe, and Wang 1997; Chen, Chen, Lin, and Yücesan 1999; Cheng 1999; Chick and Inoue 2000b).

This tutorial on Bayesian methods for simulation adopts the view of Glynn (1986) and Chick (1997), in that Bayesian statistics can help a simulation analyst to deal with issues that arise in the decision-making process. Here we discuss input distribution selection, sensitivity analysis, and the selection of the best of several alternative systems. The focus is on developing intuition for Bayesian methods, understanding how those techniques apply to practical problems, and indicate similarities to and differences from some standard frequentist methods used in practice. Thorough foundational developments are not included here, as these are already supplied in the many references given in the paper (please excuse omissions, which are unintentional). Nor do we attempt to review the rapidly expanding use of Markov Chain Monte Carlo (MCMC) simulation to implement Bayesian inference, which use applies simulation to address computational issues in Bayesian inference (Gilks, Richardson, and Spiegelhalter 1996; Carlin and Louis 1996).

Although the article deals with Bayesian methods for simulation, many of the ideas apply to other areas. Techniques to model input parameter uncertainty apply equally well to simulations of customer arrivals to a service facility and to an analytical study of the M/G/1/K queue. And statistical selection procedures apply both to studies to identify the best simulated system, and to studies to identify the best disease therapy, be it medical or agricultural. Excellent references to various aspects of Bayesian methods, subjective probability, and decision analysis in general, not just as they apply to simulation, include de Groot (1970), Lindley (1972), Savage (1972), Winkler (1972), Berger (1985), de Finetti (1990), and Bernardo and Smith (1994). References that address issues that seem to be of more interest at present to simulation than to other areas, such as shifted input distributions and common random numbers, are indicated below as they are discussed.

2 INPUT UNCERTAINTY

When simulating existing systems, or modifications to existing systems, we often have data regarding how the system functions. For example, we may have collected interarrival and service times and rework probabilities. Sometimes a lot of data is available, in which case the system might be known fairly well, but other times there might not be much data, so that the parameters of the system are not known precisely.

Suppose we are simulating a simple facility where customers arrive to a waiting area with finite capacity, wait until a single server can handle their service request, then are processed before leaving the system (an M/G/1/K queue). Several interarrival and service times have been collected

to help identify an appropriate arrival rate λ and a service time distribution.

Some commonly-used software packages take that data, fit the parameters of several candidate distributions (say, using maximum likelihood or moment methods), then select a specific distribution using a goodness-of-fit test. This widely-used technique enjoys face validity by a large number of simulation modelers.

However, using a single input distribution and parameter to describe the system can strongly underestimate uncertainty about the performance of a system. 'The arrival rate is $\lambda = 5$ customers per minute' means very different things if that rate is a point estimate from 10 interarrival times, rather than 1000. Barton and Schruben (2000) provide a compelling example of how confidence intervals for the mean output can give only a small fraction of the desired coverage probability if input uncertainty is ignored. And Chick (2000) illustrates the potential ambiguity of goodness-of-fit tests, which don't specify how to select from among nonrejected distributions. Two simulation runs based on two nonrejected service distributions can result in two nonoverlapping confidence intervals for the mean output.

One way to account for uncertainty about input distributions is *subjective* probability assessment. Subjective probability assessments quantify the uncertainty of experts about simulation inputs, and do not necessarily reflect statements about repeated independent trials of a random variable. They have been used in simulation practice (Cooke 1994; Helton 1996; Scott 1996). Some software tools exist that can help specify such distributions (Wagner and Wilson 1995). Data can be incorporated with subjective distributions for unknown inputs (a.k.a. prior distributions) using Bayes rule. Sections 2.1 and 2.2 describe a parametric Bayesian approach to select a probability distribution(e.g., exponential versus gamma) and a parameter for that distribution. Section 2.3 summarizes a nonparametric Bayes alternative.

2.1 Parameter Uncertainty

Here we suppose that a continuous-valued input parameter to a simulation is unknown, such as an arrival rate λ to an M/G/1/K queue. Suppose that interarrival times $\mathbf{x} = (x_1, \ldots, x_n)$ are available to estimate λ. The likelihood function for λ is

$$p(\mathbf{x} \mid \lambda) = \lambda^n e^{-\lambda \sum_{i=1}^n x_i} \tag{1}$$

The standard maximum likelihood estimator (MLE) $\hat{\lambda} = \sum_{i=1}^n x_i / n$ is the arrival rate that maximizes the 'likelihood' of seeing the data $\mathbf{x}$. Common practice would be to use this estimator to choose an arrival rate parameter for input to a computer simulation.

On the other hand, a range of parameters may support the data nearly as well as the MLE, as shown by the likelihood function in Figure 1 (obtained for a particular $\mathbf{x}$). One might provide a confidence interval for λ, and run simulations that use the upper and lower confidence bounds to test the sensitivity of the output to input parameter uncertainty. Edwards (1984) proposes a 'method of support' to identify a range of values of λ that are reasonably well supported by the data. He suggests looking at the log-likelihood function $\ell(\lambda; \mathbf{x}) = \ln p(\mathbf{x} \mid \lambda)$ (the 'support'), and chooses the of λ such that $\ell(\hat{\lambda}; \mathbf{x}) - \ell(\lambda; \mathbf{x})$ is small to be an alternative to a confidence interval. For example, the values above the horizontal line in Figure 1 are within 1-log of the maximum, and are therefore indicate reasonable values of λ. The method of support is graphically intuitive even for multidimensional parameters, and uses more information about the likelihood than the maximum alone.

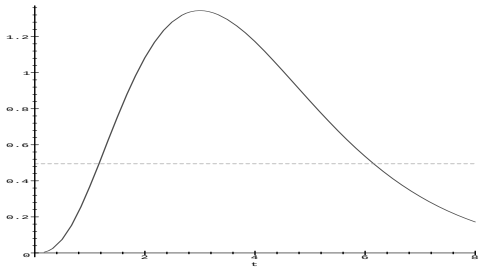

Figure 1: The Likelihood Function $p(\mathbf{x} \mid \lambda)$ for the Exponential Distribution (Here with $n = 3$ Points) Is Shaped like the Pdf of a Gamma Distribution

However, the MLE and method of support do not quantify uncertainty about λ with probability, so they are unable to describe the probability that λ exceeds a given level, or to discuss expectations of figures of merit of a queueing system that sees an arrival stream with rate λ. This is required if one is to estimate the probability that the steady-state fraction of M/G/1/K queue customers that balk exceeds a given threshold, for example. If λ were known with certainty, we could determine the fraction of balkers with analytical methods or long simulation runs. Since λ is not known, we can only answer the question 'What is the probability that more than 1% of customers balk?' if a probability model for λ is specified.

2.1.1 Accounting for Input Uncertainty

Bayesian methods do quantify uncertainty about λ with probability, using the data $\mathbf{x}$ and Bayes rule:

$$p_{\Lambda|\mathbf{x}}(\lambda) = \frac{p(\mathbf{x} \mid \lambda) f(\lambda)}{c(\mathbf{x})} \tag{2}$$

where $c(\mathbf{x}) = \int p(\mathbf{x} \mid \lambda) f(\lambda)$ is a normalizing constant. The 'catch' is that a prior distribution $f(\lambda)$ needs to be specified to describe the uncertainty a modeler has about λ before observing $\mathbf{x}$. The prior distribution reflects a *belief* about the value of λ, and does *not* necessarily provide a sampling distribution for λ over repeated experiments.

Sample averages $\hat{o} = \sum_{i=1}^{n} o_i/n$ of n simulation replications can be used to estimate the mean output $E_O[O \mid \lambda_0]$. To estimate the mean output $E_\Lambda[E_O[O \mid \Lambda]]$ in a way that accounts for input parameter uncertainty, input parameters are sampled according to their posterior distributions before each replication, and held constant during the replication. The procedure in Figure 2 gives the idea in general, where there may be several unknown inputs, and the data $\mathcal{E}$ is available to describe those inputs.

for $r = 1, \ldots, n$ replications

1. sample input distributions and parameters for the r-th replication (structural uncertainty), conditional on the information $\mathcal{E}$.
2. for the r-th replication:
   - a sample random variates given the r-th input distributions and parameters
   - b generate the simulation output o_r that is determined by the random variates.

end loop
generate the estimate $\hat{o} = \sum_{r=1}^{n} o_r/n$ of $E_O[O \mid \mathcal{E}]$.

Figure 2: Simulation algorithm that accounts for structural and stochastic uncertainty, given the information $\mathcal{E}$, using a Bayesian model average (BMA).

Glynn (1986) and Chick (1997) discuss this sampling mechanism, a *Bayesian model average* (BMA), in the simulation context. Draper (1995) and Raftery (1995), among others, use the BMA in other contexts and indicate how the BMA addresses known concerns with point estimators and goodness-of-fit tests. See also Berger and Pericchi (1996).

2.1.2 Prior Distributions.

Although any choice of prior distribution $f(\lambda)$ for λ on the positive real line that a decision maker feels comfortable with is appropriate, there are some prior probability distributions of particular interest. The *conjugate* prior distribution has the property that the posterior distribution has the same functional form as the prior distribution, a property that simplifies Bayesian inference. More general prior distributions may require Markov Chain Monte Carlo techniques to explore the posterior distribution (Gilks, Richardson, and Spiegelhalter 1996). For conditionally independent samples from an exponential distribution, the conjugate prior distribution for λ is the gamma distribution, because the posterior distribution given $\mathbf{x}$ is also a gamma distribution. It is therefore no coincidence that the likelihood function for the exponential in Figure 1 is shaped like the pdf of a gamma distribution.

Sometimes, a *noninformative*, or *reference* prior distribution is desired, so that the prior distribution doesn't overly influence the posterior distribution, or to explore the relationship of Bayesian inference with classical statistical results (Kass and Wasserman 1996; Berger 1985). For a given sampling distribution, the noninformative distribution may or may not be *proper* (i.e., integrate to 1). For the exponential model, the noninformative prior distribution (de Groot 1970) is $f(\lambda) = \lambda^{-1} d\lambda$, an improper distribution. The posterior distribution, given that noninformative prior pdf and the data $\mathbf{x}$, is a proper gamma distribution with shape parameter $n-1$, mean $n/\sum_{i=1}^{n} x_i$, and mode (a.k.a. maximum *a posteriori* or MAP estimator) is $(n-1)/\sum_{i=1}^{n} x_i$.

It turns out that all members of the regular exponential family of distributions have conjugate prior distributions. In particular, if the pdf can be written

$$f_{X|\lambda}(x) = a(x) g(\lambda) \exp\left[\sum_{j=1}^{d} c_j \phi_j(\lambda) h_j(x)\right] \tag{3}$$

for some $a(\cdot), g(\cdot), c_j, \phi_j(\cdot), h_j(\cdot)$, then the conjugate prior distribution is

$$f_{\mathbf{\Lambda}|\boldsymbol{t}}(\lambda) = [K(\boldsymbol{t})]^{-1} [g(\lambda)]^{t_0} \exp\left[\sum_{j=1}^{d} c_j \phi_j(\lambda) t_j\right] \tag{4}$$

where the parameter $\boldsymbol{t} = (t_0, \ldots, t_d)$ is chosen so that $f_{\mathbf{\Lambda}|\boldsymbol{t}}(\lambda)$ is proper (i.e., so that $K(\boldsymbol{t}) = \int [g(\lambda)]^{t_0} \exp\left[\sum_{j=1}^{d} c_j \phi_j(\lambda) t_j\right] d\lambda < \infty$). This family includes the exponential, gamma, Bernoulli, Poisson, normal, lognormal, and many other commonly used distributions (Bernardo and Smith 1994; de Groot 1970).

The choice of an appropriate prior distribution is a source of controversy. Lindley (1972), Savage (1972, Chapters 1-6) and de Finetti (1990) give foundational developments for subjective probability. Kahneman, Slovic, and Tversky (1982) study behavioral aspects of subjective probability assessments, see also Winkler (1972). Kass and Wasserman (1996) present mechanisms for developing 'default' noninformative priors, and others have suggested using subsets of the data to automate the selection of a prior (Berger and Pericchi 1996; O'Hagan 1995). In some situations the precise method for assessing a prior distribution may be somewhat irrelevant, in that the posterior distribution may be relatively insensitive to changes in the prior distribution (Berger 1994).

Chick (2000) gives several examples of prior distributions for simulation inputs, and implements the BMA algorithm in Figure 2 for a queueing simulation.

2.2 Input Distribution Uncertainty

Turn now to the selection of an input distribution and parameter for the service times of the M/G/1/K queue. In simulation practice, a set of q candidate distributions are proposed, then point estimation and goodness-of-fit tests are used to identify an appropriate distribution and parameter. For example, let λ_i be the input parameter associated with candidate distribution i for the service distribution, say with $q = 2$, using $m = 1$ for the exponential and $m = 2$ for the shifted gamma.

1. Exponential. $\lambda_1 = (\theta)$, with rate $\theta > 0$,

$$f_{Y|m=1,\lambda_1}(y) = \theta e^{-y\theta}.$$

2. Shifted gamma. $\lambda_2 = (\xi, \alpha, \beta)$, with $\xi, \alpha, \beta > 0$.

$$f_{Y|m=2,\lambda_2}(y) = \frac{\beta^\alpha}{\Gamma(\alpha)}(y-\xi)^{\alpha-1}\exp\left[-(y-\xi)\beta\right].$$

A natural extension of the ideas in Section 2.1 is to consider the unknown distribution to be a random variable M, posit a prior probability that each model is correct ($\pi_M(m)$, for $m = 1, \ldots, q$), and then specify a prior distribution $f_{\mathbf{\Lambda}_m|m}(\lambda_m)$ for the parameter λ_m of each distribution, conditional on knowing the right distribution m (λ_m can be multivariate). Given service time data $\mathbf{y} = (y_1, \ldots, y_{n'})$, Bayes rule again determines the posterior mass function (pmf) $p_{M|\mathbf{y}}(m)$ that a given distribution $M = m$ is correct, given $\mathbf{y}$; the posterior probability density function $f_{\mathbf{\Lambda}_m|m,\mathbf{y}}(\lambda_m)$ that a given parameter λ_m is correct, given $M = m$ and $\mathbf{y}$; and the predictive probability density function $f_{O|\mathbf{y}}(o)$ of the simulation output O, given $\mathbf{y}$:

$$p_{M|\mathbf{y}}(m) = \frac{f_{\mathbf{Y}|m}(\mathbf{y})\,\pi_M(m)}{\sum_{k=1}^{q} f_{\mathbf{Y}|k}(\mathbf{y})\,\pi_M(k)} \quad (5)$$

$$f_{\mathbf{\Lambda}_m|m,\mathbf{y}}(\lambda_m) = \frac{f_{\mathbf{Y}|m,\lambda_m}(\mathbf{y})\; f_{\mathbf{\Lambda}_m|m}(\lambda_m)}{f_{\mathbf{Y}|m}(\mathbf{y})} \quad (6)$$

$$f_{\mathbf{Y}|m}(\mathbf{y}) = \int f_{\mathbf{y}|m,\lambda_m}(\mathbf{y}) f_{\mathbf{\Lambda}_m|m}(\lambda_m) d\lambda_m$$

$$f_{O|\mathbf{y}}(o) = \sum_{m=1}^{q} p_{M|\mathbf{y}}(m) \int f_{O|m,\lambda_m}(o) \times f_{\mathbf{\Lambda}_m|m,\mathbf{y}}(\lambda_m)\, d\lambda_m \quad (7)$$

Equation 7 gives the distribution of simulation output when input distribution and parameter uncertainty are accounted for by the BMA algorithm of Figure 2. Step 1 of the algorithm samples an input distribution m and parameter λ_m, and Step 2 generates the output o as a function of the variates generated using those inputs.

Equation 7 assumes that $f_{O|m,\lambda_m,\mathbf{y}}(o) = f_{O|m,\lambda_m}(o)$, which reflects the modeling assumption that historical data for Y and simulation output O are conditionally independent, given the distribution m and parameter λ_m. The assumption is reasonable if the simulated variates are generated independent of the values of the historical data, but is violated by trace-driven simulations.

Shifted distributions, like the shifted gamma above, are used more frequently in simulation but are less frequently studied from an analytic perspective. Perhaps this is because regularity conditions can fail to hold (the shifted gamma is not a member of the regular exponential family, so a finite dimensional conjugate prior distribution does not exist; maximum likelihood estimators may be ill-defined). Chick (2000) gives an example of assessing a prior distribution for a shifted gamma distribution, and illustrates the use of the adaptive rejection Metropolis sampler (ARMS) of Gilks, Best, and Tan (1995) to generate the samples from the posterior distributions in Equations 5-6. The ARMS and other MCMC techniques may be required to work with posterior distributions when nonregularity arises in general. Note that improper prior distributions cannot be used when there is more than one candidate input distribution, as the arbitrary proportionality constant for an improper $f_{\mathbf{\Lambda}_m|m}(\lambda_m)$ would cause Equation 5 to be ill-defined.

2.3 Nonparametric Bayes Methods

The previous approaches require that parametric models be used for input distributions. An alternate strategy is to use empirical distributions (Law and Kelton 2000). Barton and Schruben (1993) advocate bootstrapping to quantify uncer-

tainty that occurs from using an empirical distribution. Barton and Schruben (2000) later propose several mechanisms for accounting for uncertainty about empirical distributions, including a Bayesian bootstrapping technique they call a randomized empirical distribution function (EDF).

The standard piecewise continuous EDF, $F_e(\cdot)$ assigns probability according to the order statistics $x_{(1)}, \ldots, x_{(n)}$ of the data $\mathbf{x}$.

$$F_e(x) = \begin{cases} 0 & \text{if } x < x_{(1)} \\ \frac{i-1}{n-1} + \frac{x - x_{(i)}}{(n-1)(x_{(i+1)} - x_{(i)})} & \text{if } x_{(i)} \le x < x_{(i+1)} \\ 1 & \text{if } x > x_{(n)} \end{cases} \tag{8}$$

Rather than assigning $F_e(x_{(i)}) = (i-1)/(n-1)$ deterministically, they note that the marginal distribution of the i-th order statistic has a Beta$(i, n-i+1)$ distribution. Therefore, uncertainty about the EDF for a given random quantity in a simulation can be described by using atoms for the EDF that have the same distribution as the order statistics, rather than using $(i-1)/(n-1)$. This leads to the modification in Figure 3 to the model average in Figure 2 to account for input uncertainty for empirical distributions. Figure 3 describes how to handle a single source of uncertainty. Multiple sources are handled analogously.

1. Generate a sample of n values from a uniform [0,1] distribution with ordered values $u_{[1]}, \ldots, u_{[n]}$.
2. Set $\hat{F}_e(x_{(i)}) = u_{[i]}$, then smooth the EDF, say, by linear interpolation as in Equation 8, and use this as the input distribution (rather than selecting a parametric distribution in step 1 of Figure 2).
3. Conduct a simulation run using this empirical distribution (step 2 of Figure 2).
4. Repeat for all n replications.

Figure 3: A randomized empirical distribution function using a Bayesian bootstrap.

3 SENSITIVITY

Section 2 focuses on propagating input distribution and parameter uncertainty to simulation output in order to account for output uncertainty that is due to input uncertainty. The figure of merit is the mean output $E_O[O \mid \mathcal{E}] = E_\Lambda[E_O[O \mid \mathcal{E}, \Lambda]]$ given the data $\mathcal{E}$, averaged not only over stochastic uncertainty of the system (random output o for a given λ), but also over systemic uncertainty (an unknown λ).

If we are interested in how input uncertainty affects the mean system performance, we may be more interested in properties of the conditional expectation of the output, given the inputs. More formally, if θ is a vector of one or more input parameters that describe sources of randomness in a simulation, the mean response

$$\eta(\theta) = E_O[O \mid \mathcal{E}, \theta], \tag{9}$$

gives information about how the unknown mean depends on each input. This section examines properties of the conditional expectation of the output, when input parameters are unknown.

3.1 Output Response Metamodels

One approach to estimating the mean response $\eta(\theta)$ as a function of an unknown p-dimensional input vector θ is to posit a functional form

$$\eta(\theta) = \sum_{i=0}^{k-1} \beta_i f_i(\theta) \tag{10}$$

where the β_i are unknown coefficients that are to be estimated, and the $f_i(\theta)$ are suitably selected basis functions. To simplify, we presume that θ is a vector of continuous parameters, and do not include decision variables (parameters under our control). When the $f_i(\theta)$ are polynomials, this is a polynomial regression model.

There are at least two distinct approaches. George and McCulloch (1996) propose a stochastic search variable selection (SSVS) technique. They use a normal linear model for the n-dimensional output vector $\mathbf{o}$,

$$\mathbf{o} \sim \mathcal{N}\left(\mathbf{\Lambda}\boldsymbol{\beta}, \sigma^2 I_n\right)$$

where n is the number of replications, $\mathbf{\Lambda}$ is the $n \times k$ input matrix (a function of the inputs), σ^2 is a scalar, and I_n is the identity matrix. Motivated by a desire to identify a subset $\lambda_1^*, \ldots \lambda_\ell^*$ of important factors, they propose a prior distribution for the response parameters $\boldsymbol{\beta}$ that allows for the possibility that each β_i is either large in absolute value, or near zero. This is done as a scale mixture of two normal distributions, with indicator variables $\gamma_i = 0$ or 1,

$$p_{\beta_i|\gamma_i}(\beta_i) = (1-\gamma_i)\mathcal{N}\left(0, \tau_i^2\right) + \gamma_i \mathcal{N}\left(0, c_i^2 \tau_i^2\right)$$

where

$$p(\gamma_i = 1) = 1 - p(\gamma_i = 1) = h_i.$$

The hyperparameters τ_i and c_i are set small and large respectively so that $\mathcal{N}\left(0, \tau_i^2\right)$ is concentrated, and $\mathcal{N}\left(0, c_i^2 \tau_i^2\right)$

is diffuse. The idea is that when $\gamma_i = 0$, the coefficient β_i is probably quite small and can be 'safely' estimated as 0. George and McCulloch (1996) also describe strategies for choosing τ_i, c_i, h_i, and a prior distribution for σ^2. Cheng (1999) extends the SSVS, in part by presuming that the errors have an inverse Gaussian distribution, rather than a normal distribution. He uses the technique to fit a high-order polynomial response model to a computer packet assembly/disassembly device simulation. Chipman, Hamada, and Wu (1997) describe prior distributions for $\boldsymbol{\beta}$ that account for complex aliasing, such effect sparsity and effect heredity.

The approach advocated by Raftery, Madigan, and Hoeting (1997) differs from George and McCulloch (1996) in that there are 2^k distinct candidate response models $\eta_1(\theta), \ldots, \eta_{2^k}(\theta)$, each model corresponding to one of the 2^k combinations of considering each of the β_i to be zero or nonzero. Raftery, Madigan, and Hoeting (1997) provide desiderata to assess prior distributions for the β_i, conditional on the response model. Chick and Ng (2000) run simulations to see if that approach can identify the true response surface $\ln\lambda - \ln(\mu - \lambda)$ for the log of the stationary mean occupancy of the M/M/1 queue. They propose the following 2^4 candidate response models, as functions of the unknown arrival and service rates, $\theta = (\lambda, \mu)$:

- $\eta_1(\lambda, \mu) = \beta_0 + \beta_1 \ln\mu + \beta_2 \ln\lambda + \beta_3 \ln(\mu - \lambda)$
- $\eta_2(\lambda, \mu) = \beta_0 + \beta_1 \ln\mu + \beta_2 \ln\lambda + 0$
- $\vdots$
- $\eta_{13}(\lambda, \mu) = 0 + 0 + \beta_2 \ln\lambda + \beta_3 \ln(\mu - \lambda)$
- $\eta_{14}(\lambda, \mu) = 0 + 0 + \beta_2 \ln\lambda + 0$
- $\eta_{15}(\lambda, \mu) = 0 + 0 + 0 + \beta_3 \ln(\mu - \lambda)$

Given 25 replications each for 4 different sampled inputs (μ_i, λ_i, o_i), they found the posterior marginal probability for the correct functional form (η_{13}) was much higher than for competitor models (see Table 1).

Table 1: Posterior Probability (Prob.) of Several Proposed Functional Forms for the Log of the Stationary Occupancy of an M/M/1 Queue; the Correct Model Is Heavily Favored after 100 Replications

Prob.	Response function
0.780	$\eta_{13} = \beta_2 \ln\lambda + \beta_3 \ln(\mu - \lambda)$
0.130	$\eta_9 = \beta_1 \ln\mu + \beta_2 \ln\lambda + \beta_3 \ln(\mu - \lambda)$
0.073	$\eta_5 = \beta_0 + \beta_2 \ln\lambda + \beta_3 \ln(\mu - \lambda)$
0.013	$\eta_1 = \beta_0 + \beta_1 \ln\mu + \beta_2 \ln\lambda + \beta_3 \ln(\mu - \lambda)$
10^{-29}	$\eta_2 = \beta_0 + \beta_1 \ln\mu + \beta_2 \ln\lambda$
$\vdots$	$\vdots$
10^{-102}	$\eta_8 = \beta_0$

3.2 Distribution of Conditional Expectation

Lee and Glynn (1999) estimate the distribution function of a conditional expectation, rather than its mean.

$$\Pr_\Theta(E_O[O \mid \mathcal{E}, \theta] \leq o)$$

Applications of this type of estimation are risk management portfolios, where an option's current value can be expressed as a conditional expectation; and capacity expansion decisions, when the future demand is unknown and the probability of having sufficient capacity is of central concern. Lee and Glynn (1999) use large deviation results to determine how to allocate computing time to two different activities, sampling outputs for at a fixed λ to estimate $E_O[O \mid \mathcal{E}, \lambda]$, and sampling at different λ to account for input uncertainty. See Lee and Glynn (1999) for details.

4 SELECTION PROCEDURES

Selection procedures provide a means to allocate computer budgets to run simulation replications so that the best system can be identified. Here, best is taken to mean the system with the highest (or lowest) expected value of the output. Since the output is random, the system selected as best is not necessarily best.

Figure 4 gives the general structure of a selection procedure. Selection procedures are distinguished by the way they accomplish each step. Indifference-zone (IZ) procedures have been the standard means to identify the best simulated system (Goldsman and Nelson 1998). They are based on a frequentist analysis that guarantees that the best system will be identifed as best with a probability that exceeds some threshold P^*, subject to an indifference zone constraint (the best system is at least δ^* better than all of the others, where δ^* is the smallest practically significant difference to a decision maker). Many IZ procedures are statistically conservative, and require many replications even for systems that perform poorly during the first stage of replications. Nelson, Swann, Goldsman, and Song (2000) combine a subset selection step with IZ ideas to reduce the expected number of simulation replications required when there are noncompetitive systems.

Two distinct Bayesian formulations have been proposed as an alternate means to improve the efficiency of selection procedures. Rather than suggesting how many replications are required to guarantee a given probability of correct selection P^*, focus is on allocating replications to improve the Bayesian posterior probability of correct selection. The two formulations share features described in Section 4.1, and are distinguished in Section 4.2 and Section 4.3. Inoue, Chick, and Chen (1999) indicate that all the sequential Bayesian procedures here, and the two stage procedures

1. Run an initial stage of replications (say, n_0 replications per system).
2. Determine how many more replications n_i to run for each system, $i = 1, \ldots, k$.
3. Run the replications additional replications.
4. If more replications are desired, go to step 2.
5. Select one system as best based on the simulation output.

Figure 4: One generic selection procedure.

based on ideas in Section 4.3, perform favorably with respect to the two-stage IZ procedure of Rinott (1978).

4.1 Bayesian Selection Procedure Basics

Suppose that the best of k systems is to be selected, that the simulation output $o_{i,j}$ for replication j of system i is normally distributed with an unknown mean w_i and variance σ_i^2 ($i = 1, \ldots, k$; $j = 1, \ldots$). For the moment, presume that the $o_{i,j}$ are jointly independent. A Bayesian approach considers the unknown mean of system i to be a random variable W_i whose realization w_i is to be inferred with Bayes' rule and the output $\mathbf{o}_i = (o_{i,1}, \ldots, o_{i,m_i})$ seen so far. We use w_i rather than μ_i to follow the convention of writing random variables (realizations) in upper (lower) case. We also refer to the precision of a normal distribution, $\lambda_i = 1/\sigma_i^2$, rather than the variance. Set $\mathbf{w} = (w_1, \ldots, w_k)$ and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_k)$.

Information from initial runs is used by assigning a prior distribution and using Bayes' rule to infer the values of the unknown mean W_i and precision Λ_i. Here we use the noninformative prior pdf $f(w_i, \lambda_i) \propto 1/\lambda_i$. After observing the output $\mathbf{o}_i = (o_{i,1}, \ldots, o_{i,m_i})$, determine the sample mean $\bar{o}_i$ and sample variance $\hat{\sigma}_i^2$. The joint posterior distribution for W_i and Λ_i is a normal-gamma distribution (de Groot 1970),

$$\Lambda_i \sim \mathcal{G}\left((m_i - 1)/2, \hat{\sigma}_i^2(m_i - 1)/2\right) \quad (11)$$

$$W_i \mid_{\lambda_i} \sim \mathcal{N}\left(\bar{o}_i, \lambda_i^{-1}/m_i\right). \quad (12)$$

where a gamma distribution $\mathcal{G}(\alpha, \beta)$ has mean α/β and variance α/β^2, and a normal distribution $\mathcal{N}(\mu, \sigma^2)$ has mean μ and variance σ^2. Further, the marginal distribution and frequentist confidence interval for the unknown mean are both described by the same three parameter Student-t distribution,

$$W_i \sim \mathbf{St}\left(\bar{o}_i, m_i/\hat{\sigma}_i^2, m_i - 1\right), \quad (13)$$

where the three parameter Student-t distribution $\mathbf{St}(\mu, \kappa, \nu)$ has mean μ, precision κ, and degrees of freedom ν. When $\nu > 2$, the variance is $\kappa^{-1}\nu/(\nu - 2)$ (not κ^{-1}). If the variance σ_i^2 is known, the noninformative prior for w_i is $f(w_i) \propto 1$, and the posterior distribution given $\mathbf{o}_i$ is

$$W_i \sim \mathcal{N}\left(\bar{o}_i, \sigma_i^2/m_i\right). \quad (14)$$

Independent replications and independent noninformative prior distributions imply that the W_i are jointly independent. The distributions in Equation 13 ($i = 1, \ldots, k$) determine the Bayesian probability of correct selection ($\mathrm{PCS}_{\mathrm{Bayes}}$), defined as the posterior probability that the mean $w_{(k)}$ of the selected system is the mean $w_{[k]}$ of the 'best', given all output ($\mathcal{E} = \left\{o_{i,j} : i = 1, \ldots, k;\ j = 1, \ldots, m_i\right\}$),

$$\mathrm{PCS}_{\mathrm{Bayes}} \stackrel{\mathrm{def}}{=} \Pr\left(\left\{\mathbf{w} : w_{(k)} = w_{[k]}\right\} \mid \mathcal{E}\right), \quad (15)$$

The quantity $\mathrm{PCS}_{\mathrm{Bayes}}$ measures the *probabilistic evidence for correct selection* after a given set of replications have been run, *not* the probability that a procedure will select a known best system in repeated applications of a procedure to the same selection problem.

The procedures in Section 4.2 and Section 4.3 are motivated in different ways by the Bayesian formulation. The procedures allocate replications in an attempt to improve $\mathrm{PCS}_{\mathrm{Bayes}}$, and do not yet make claims about prespecified frequentist guarantees for P^*.

4.2 OCBA Procedures

The $\mathcal{OCBA}$ algorithms of Chen (1996) and Chen, Chen, Lin, and Yücesan (1999) are motivated by a thought experiment regarding the distribution of the unknown mean. The idea is to approximate how a few additional replications might change $\mathrm{PCS}_{\mathrm{Bayes}}$, assuming that (a) the system with the best sample mean based on replications seen so far is to be selected, (b) the effect of an additional small number n_i of replications for system i reduces the variance in the estimate in the unknown mean by changing the factor of $\hat{\sigma}_i^2/m_i$ to a factor of $\hat{\sigma}_i^2/(m_i + n_i)$. This type of 'variance adjustment' is also used in Law and Kelton (2000) to estimate the number of additional replications needed to obtain a confidence interval of a prespecified size.

Chen (1996) makes several approximations that are designed to reduce the CPU time to allocate the additional replications (step 2 of Figure 4) for the next stage of the procedure. One, $\mathrm{PCS}_{\mathrm{Bayes}}$ is approximated by a lower bound called the approximate probability of correct selection (APCS).

$$\mathrm{APCS} \stackrel{\mathrm{def}}{=} \prod_{j:j\neq(k)} \Pr\left(\left\{\mathbf{w} : w_j < w_{(k)}\right\} \mid \mathcal{E}\right) \quad (16)$$

Two, the sample variance is presumed to be a good estimate of the known variance, and Equation 14 is used rather than Equation 13. The thought experiment presumes that if an additional n_i replications are allocated for system i, but none are allocated for the others, then

$$\begin{aligned} \tilde{W}_i &\sim \mathcal{N}\left(\bar{o}_i, \hat{\sigma}_i^2/(m_i+n_i)\right) \\ \tilde{W}_j &\sim \mathcal{N}\left(\bar{o}_j, \hat{\sigma}_j^2/m_j\right) \qquad \text{for } j \neq i. \end{aligned}$$

This induces an *estimated approximate probability of correct selection* (EAPCS) of

$$\text{EAPCS}_i = \prod_{j:j\neq(k)} \Pr\left(\{\tilde{\mathbf{w}} : \tilde{w}_j < \tilde{w}_{(k)}\} \mid \mathcal{E}\right), \qquad (17)$$

where the probability is with respect to the distribution of $\tilde{\mathbf{W}} = (\tilde{W}_1, \ldots, \tilde{W}_k)$ rather than $\mathbf{W}$.

Replications are allocated at each stage to increase an estimate of the improvement in correct selection, $\text{EAPCS}_i - \text{APCS}$. Chen and coauthors provide a number of variations on this central theme that make various asymptotic approximations for allocating replications at each stage. The procedures can stop either when the APCS reaches a satisfactorily high level, or when the computer budget for replications is exhausted. The procedures are empirically effective at identifying the best system, particularly when run sequentially.

4.3 Value of Information Procedures

Another approach attempts to select an additional number of replications for each system to improve the *expected value of information* gained from those replications, rather than using the thought experiment of Section 4.2. Information gains for the probability of correct selection are measured with respect to the 0-1 loss function (the loss is 0 if the best is correctly selected, and is 1 for incorrect selections). Alternately replications can be allocated to reduce the expected opportunity cost loss of a potentially incorrect selection. Gupta and Miescke (1996) provide a formal problem statement and indicate that closed-form solutions are known only for special cases, and that computation of an optimal solution is numerically intensive.

Chick and Inoue (2000b) propose a surrogate objective (the sum over pairwise comparisons between the current best at each other system for the expected value of information) that is readily computed under general conditions, and derive two-stage and sequential procedures for both loss functions, assuming that replications are independent. Chick and Inoue (2000a) derive two stage procedures, one for each loss function, assuming that common random numbers and screening are used.

To compute the expected value of information that will be gained from running more replications, the probability that a given system will be selected as 'best' (the system with the highest posterior mean performance) is required. This requires the *predictive distribution* for the output $\mathbf{Y}_i = (Y_{i,1}, \ldots, Y_{i,n_i})$ to be observed, as well as the predictive distribution for the posterior mean $Z_i = E[O_{i,m_i+1} \mid \mathbf{o}_i, \mathbf{Y}_i]$. Since uncertainty about W_i, Λ_i given $\mathbf{o}_i$ is given by Equations 11-12, and the $Y_{i,j}$ is conditionally independent given that mean and precision, one can show (Bernardo and Smith 1994) that

$$Y_{i,j} \sim \mathbf{St}\left(\bar{o}_i, \frac{m_i}{(m_i+1)\hat{\sigma}_i^2}, m_i - 1\right) \qquad (18)$$

$$Z_i \sim \mathbf{St}\left(\bar{o}_i, \frac{m_i(m_i+n_i)}{n_i\hat{\sigma}_i^2}, m_i - 1\right). \qquad (19)$$

Once $\mathbf{y}_i$ is observed, then $z_i = E[O_{i,m_i+1} \mid \mathbf{o}_i, \mathbf{y}_i]$ is readily computed. The system with the maximal z_i after all output is observed is selected as best.

The 'value of information' procedures select the number of replications to be done $(n_1, \ldots, n_k)$ to alter the distributions of the yet-to-be-observed posterior mean performance of each system $(Z_1, \ldots, Z_k)$ to improve the value of information of those replications. See Chick and Inoue (2000a) and Chick and Inoue (2000b) for theory, and for numerical results that show the procedures compare favorably with some indifference-zone procedures.

5 CLOSING REMARKS

Bayesian methods provide an alternate way of viewing issues that arise during a simulation analysis. This paper described a subset of the ways that Bayesian, decision-theoretic methods can be applied to resolve issues about input uncertainty, to select input distributions, to evaluate how output uncertainty is affected by input uncertainty, and to efficiently identify the best of a set of simulated alternatives. Other issues that may be important to address include (i) data collection plans that reduce input parameter uncertainty in a way that reduces output uncertainty; (ii) a more comprehensive integration of input uncertainty into the selection procedures; (iii) further integration of multiple sources of information about input or output characteristics; (iv) inclusion of multiple attributes in the decision-making process that is served by simulation; (v) a further accounting of different types of correlation in Bayesian inference (via antithetic variates, or from batch to batch, etc.) (vi) the incorporation of derivative information into response model estimates.

ACKNOWLEDGMENTS

The many discussions with members of the discrete-event simulation community about Bayesian methods is greatly appreciated. (You know who you are!) Many thanks go to Max Mendel and Dick Barlow, two important teachers, and to Koichiro Inoue, Szu Hui Ng, and Sada Soorapanth, past and present students.

REFERENCES

Andradóttir, S., and V. M. Bier. 1997. Applying Bayesian ideas in simulation. Dept. of Industrial Engineering, University of Wisconsin-Madison, Technical Report 97-1.

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, NJ: Institute of Electrical and Electronics Engineers, Inc.

Barton, R. R., and L. W. Schruben. 2000. Simulating real systems. *in submission*.

Berger, J. O. 1985. *Statistical decision theory and Bayesian analysis* 2nd ed. New York: Springer-Verlag.

Berger, J. O. 1994. An overview of robust Bayesian analysis. *TEST* 3:5–124.

Berger, J. O., and L. R. Pericchi. 1996. The intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association* 91 (433): 109–122.

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

Carlin, B. P., and T. A. Louis. 1996. *Bayes and empirical Bayes methods for data analysis*. London: Chapman & Hall.

Chen, C.-H. 1996. A lower bound for the correct subset-selection probability and its application to discrete event simulations. *IEEE Transactions on Automatic Control* 41 (8): 1227–1231.

Chen, H., and B. Schmeiser. 1995. Monte Carlo estimation for guaranteed-coverage non-normal tolerance intervals. *Journal of Statistical Computation and Simulation* 51 (2-4): 223–238.

Chen, H.-C., C.-H. Chen, J. Lin, and E. Yücesan. 1999. An asymptotic allocation for simultaneous simulation experiments. In *Proceedings of the Winter Simulation Conference*, ed. P. A. Farrington, H. B. Nembhard, D. Sturrock, and G. Evans, 359–366. Piscataway, NJ: Institute of Electrical and Electronics Engineers, Inc.

Cheng, R. C. H. 1999. Regression metamodelling in simulation using Bayesian methods. In *Proceedings of the Winter Simulation Conference*, ed. P. A. Farrington, H. B. Nembhard, D. Sturrock, and G. Evans, 330–335. Piscataway, NJ: Institute of Electrical and Electronics Engineers, Inc.

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. Input distribution selection for simulation experiments: Accounting for input uncertainty. *Operations Research*. to appear.

Chick, S. E., and K. Inoue. 2000a. New procedures for identifying the best simulated system using common random numbers. *in resubmission*.

Chick, S. E., and K. Inoue. 2000b. New two-stage and sequential procedures for selecting the best simulated system. *Operations Research*. to appear.

Chick, S. E., and S.-H. Ng. 2000. Input distribution selection for simulation experiments: Reducing input uncertainty. Technical report, University of Michigan, Ann Arbor, Dept. of Industrial & Operations Engineering.

Chipman, H. A., M. Hamada, and C. Wu. 1997. A Bayesian variable-selection approach for analyzing designed experiments with complex aliasing. *Technometrics* 39 (4): 372–381.

Cooke, R. M. 1994. Uncertainty in dispersion and deposition accident consequence modelling assessed with performance-based expert judgement. *Reliability Engineering and System Safety* 45:35–46.

de Finetti, B. 1990. *Theory of probability, v. 2*. New York: John Wiley & Sons, Inc.

de Groot, M. H. 1970. *Optimal statistical decisions*. New York: McGraw-Hill, Inc.

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

Edwards, A. W. F. 1984. *Likelihood*. Cambridge: Cambridge University Press.

George, E. I., and R. E. McCulloch. 1996. Stochastic search variable selection. In *Markov chain Monte Carlo in practice*, ed. W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. London: Chapman and Hall.

Gilks, W. R., N. G. Best, and K. K. C. Tan. 1995. Adaptive rejection metropolis sampling. *Applied Statistics* 44:455–472.

Gilks, W. R., S. Richardson, and D. J. Spiegelhalter. 1996. *Markov chain Monte Carlo in practice*. London: Chapman and Hall.

Glynn, P. 1986. Problems in Bayesian analysis of stochastic simulation. In *Proceedings of the Winter Simulation Conference*, ed. J. R. Wilson, J. O. Henriksen, and S. D. Roberts, 376–383. Piscataway, NJ: Institute of Electrical and Electronics Engineers, Inc.

Goldsman, D., and B. L. Nelson. 1998. Statistical screening, selection, and multiple comparisons in computer

simulation. In *Proceedings of the Winter Simulation Conference*, ed. D. J. Madeiros, E. F. Watson, J. S. Carson, and M. S. Manivannan, 159–166. Piscataway, NJ: Institute of Electrical and Electronics Engineers, Inc.

Gupta, S. S., and K. J. Miescke. 1996. Bayesian look ahead one-stage sampling allocations for selecting the best population. *Journal of Statistical Planning and Inference* 54:229–244.

Helton, J. C. 1996. Computational structure of a performance assessment involving stochastic and subjective uncertainty. In *Proceedings of the Winter Simulation Conference*, ed. J. M. Charnes, D. J. Morrice, D. T. Brunner, and J. J. Swain, 239–247. Piscataway, NJ: Institute of Electrical and Electronics Engineers, Inc.

Inoue, K., S. E. Chick, and C.-H. Chen. 1999. An empirical evaluation of several methods to select the best system. *ACM Transactions on Modeling and Computer Simulation* 9 (4): *in press*.

Kahneman, D., P. Slovic, and A. Tversky. 1982. *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.

Kass, R. E., and L. Wasserman. 1996. The selection of prior distributions by formal rules. *Journal of the American Statistical Association* 91 (435): 1343–1370.

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

Lee, S.-H., and P. Glynn. 1999. Computing the distribution function of a conditional expectation via Monte Carlo: Discrete conditioning spaces. In *Proceedings of the Winter Simulation Conference*, ed. P. A. Farrington, H. B. Nembhard, D. Sturrock, and G. Evans, 1654–1663. Piscataway, NJ: Institute of Electrical and Electronics Engineers, Inc.

Lindley, D. V. 1972. *Bayesian statistics, a review*. Philadelphia: Society of Industrial and Applied Mathematics.

Nelson, B. L., B. W. Schmeiser, M. R. Taaffe, and J. Wang. 1997. Approximation-assisted point estimation. *Operations Research Letters* 20 (3): 109–118.

Nelson, B. L., J. Swann, D. Goldsman, and W. Song. 2000. Simple procedures for selecting the best simulated system when the number of alternatives is large. Technical report, Northwestern University, Department of Industrial Engineering and Management Science, Evanston, IL.

O'Hagan, A. 1995. Fractional Bayes factors for model comparison (with discussion). *Journal of the Royal Statistical Society, Series B* 56:99–118.

Raftery, A. E. 1995. Bayesian model selection in social research (with discussion by Andrew Gelman & Donald B. Rubin, and Robert M. Hauser, and a rejoinder). In *Sociological Methodology 1995*, ed. P. V. Marsden. Cambridge, Mass.: Blackwells.

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.

Rinott, Y. 1978. On two-stage selection procedures and related probability-inequalities. *Communications in Statistics* A7:799–811.

Savage, L. J. 1972. *The foundations of statistics*. New York: Dover Publications, Inc.

Scott, E. M. 1996. Uncertainty and sensitivity studies of models of environmental systems. In *Proceedings of the Winter Simulation Conference*, ed. J. M. Charnes, D. J. Morrice, D. T. Brunner, and J. J. Swain, 255–259. Piscataway, NJ: Institute of Electrical and Electronics Engineers, Inc.

Wagner, M. A. F., and J. R. Wilson. 1995, July. Graphical interactive simulation input modeling with bivariate Bézier distributions. *ACM Transactions on Modeling and Computer Simulation* 5 (3): 163–189.

Winkler, R. L. 1972. *Introduction to Bayesian inference and decision*. New York: Holt.

AUTHOR BIOGRAPHY

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 interests include Bayesian statistics, epidemiology models, decision analysis, reliability, and computational methods in statistics. His email and web addresses are <sechick@engin.umich.edu> and <http://www.engin.umich.edu/~sechick/>.