Bayesian Simulation and Decision Analysis: An Expository Survey

Jason R. W. Merrick

Department of Statistical Sciences and Operations Research, Virginia Commonwealth University,
Richmond, Virginia 23284, jrmerric@vcu.edu

The aim of this expository survey on Bayesian simulation is to stimulate more work in the area by decision analysts. We discuss the main areas of research performed thus far, including input analysis, propagation and estimation of output uncertainty, output analysis, making decisions with simulations, selecting the best simulated system, and applications of Bayesian simulation methods. Throughout, we offer avenues of future research in Bayesian simulation that may be of interest to decision analysts.

Key words: simulation; Bayesian statistics; decision analysis
History: Received on September 22, 2008. Accepted on March 31, 2009, after 2 revisions. Published online in *Articles in Advance* September 11, 2009.

1. Introduction

From a managerial perspective, the main use of a simulation of a system is to aid in making decisions. Banks et al. (2005, p. 3) define a simulation as "an imitation of the operation of a real-world or proposed process or system over time." Simulation involves "the generation of an artificial history to draw inferences concerning the operating characteristics of the system." Traditionally, such inference has followed the frequentist paradigm. More recently, researchers have proposed Bayesian approaches to simulation model construction and inference.

The term simulation can mean different things to different people. In this paper, we discuss stochastic simulation of systems over time, and particularly discrete event or system simulation where each replication is expensive in computational effort. Researchers have also applied simulation techniques to Bayesian calculations, specifically Markov chain Monte Carlo (MCMC) methods (Gilks et al. 1996, Gamerman and Lopes 2006) and Bayesian model averaging (BMA; Draper 1995). One could refer to such techniques as Bayesian simulation, and we use them in the work discussed herein. However, they are not the simulations of interest in our discussion. Similarly, the Bayesian work on modeling parameter uncertainty with deterministic models is not our focus. Although one can apply many of the methods discussed in Monte Carlo simulation, Monte Carlo is not our focus. The models discussed herein include stochastic elements in the model itself, as well as uncertainty in the parameters, and in some cases model uncertainty. Law and Kelton (2001), Banks et al. (2005), and Henderson and Nelson (2006) provide excellent overviews of the frequentist perspective on this type of simulation.

Stochastic simulations are models of the aleatory uncertainty about a system. In simpler terms, they model the randomness of the system itself. We model the uncertainty about interarrival times, service times, failure times, and other random quantities with probability distributions. In classical simulation, one estimates the parameters of these distributions using point estimates based on available data. In Bayesian simulation, one models knowledge about the parameters using prior distributions that one updates when data is available. The prior (or posterior) distributions represent the epistemic uncertainty, or the lack of knowledge about the system. Thus, both sources of uncertainty are included in a Bayesian simulation model. The analyst can also include uncertainty about the correct family of distributions or the correct model into this framework. A more complete representation of uncertainty is beneficial when we use the

simulation in analyzing a decision. The analyst can estimate how much uncertainty is due to each source and determine whether he or she needs additional input data or additional replications of the simulation. The very nature of Bayesian simulation makes the area fertile for decision analysts.

There have been a number of applications of Bayesian simulation methods in recent years. Popova and Morton (1998) used stochastic optimization for work force scheduling with Bayesian models and inference. Herzog (2002) used Bayesian models for insurance and mortgage problems using Monte Carlo simulation. Muñoz (2003) discussed a model of airline reservations and overbooking costs, applying the model uncertainty framework of Chick (2001). McCabe (2003) modeled schedule risk with a Bayesian belief network version of a PERT (program evaluation and review technique) network and used Monte Carlo simulation to obtain posterior schedule completion probabilities. Chick et al. (2003) modeled the spread of waterborne infectious diseases using Bayesian inference from simulations, expert judgment, and field data. The applications of the techniques discussed herein are widespread and varied.

We aim to introduce the area of Bayesian simulation to the decision analysis community. We hope that they can apply more decision analytic tools and approaches in the area. There have been several fine tutorials on Bayesian simulation written for the simulation audience (Chick 1997a, 2000, 2004, 2006a, b). We review each of the areas of research in Bayesian simulation to provide a thorough background on the research performed thus far, like each of these tutorials. However, we discuss areas of research that decision analysts can develop further.

In the next section, we depict a simple Bayesian simulation of a queue using influence diagrams. This illustrates the nature of Bayesian simulations from both an input and an output perspective. We then discuss each of the research areas in turn. We discuss simulation input analysis, output sampling and uncertainty estimation, and simulation output analysis and metamodels. We discuss using simulations to make decisions and selecting the best from a set of simulated decisions with minimal replications. Each section gives an overview of the work performed in the area and makes suggestions for research in the area that might particularly suit the readers of this journal.

Section 3 discusses Bayesian modeling of simulation inputs. Readers with an interest in model uncertainty and semiparametric Bayesian analysis should pay particular attention to §3. Section 4 discusses the analysis of uncertainty in simulation outputs, including parameter uncertainty and model uncertainty, as well as the stochastic nature of the simulation itself. Readers interested in MCMC sampling techniques and BMA will find applications in this section. Readers with an interest in Bayesian regression should focus on §5 on output analysis using metamodels and Bayesian response-surface methods. In §5, we also discuss how to allocate additional input data collection and simulation replications to reduce output uncertainty. This area could benefit from consideration of the value of information. All decision analysts will find §6.1 on using simulations in the decision process of interest. Section 6.2 concerns the algorithms for choosing among alternative simulation models to choose the best simulated system alternative. This includes loss function approaches that lead to the best algorithms in empirical testing on benchmark problems. Readers with an interest in stochastic optimization and algorithms for choosing the best alternative should examine §6.2. We finish with a discussion of applications of Bayesian simulation in §7, including illustration of one application of Bayesian simulation. The final section summarizes the recommendations to decision analysts for further research avenues.

2. Influence Diagrams of a Simple Bayesian Simulation of a Queue

Bayesian simulation differs from classical simulation analysis in that we use probability distributions to represent the uncertainty about model parameters, rather than point estimates and confidence intervals. We apply such treatment to both the random inputs of the model and the outputs from the model. In the language of uncertainty, classical simulation models only aleatory uncertainty (the randomness of the system itself), whereas Bayesian simulation models both the aleatory and epistemic uncertainty (the lack of knowledge about the system). As an example, let us consider a simulation of an $M/M/s$ queue. We assume

the times between arrivals to a single line, $T_1, T_2, \ldots,$ to be exponentially distributed with parameter λ. s servers take entities from the queue on a first-come, first-served basis with the system starting empty at time 0. Service times, $S_1, S_2, \ldots,$ follow an exponential distribution with parameter μ, and entities then leave the system. Neither λ or μ are known, so we have a prior distribution on each, $\pi(\lambda)$ and $\pi(\mu \mid \lambda)$, and the decision maker has observed some input data for each to obtain posterior distributions on λ and μ, $\pi(\lambda \mid t_1, t_2, \ldots, t_p)$ and $\pi(\mu \mid s_1, s_2, \ldots, s_q, \lambda)$, where $t_1, t_2, \ldots, t_p$ and $s_1, s_2, \ldots, s_q$ are p realizations of $T_1, T_2, \ldots$ and q realizations of $S_1, S_2, \ldots$, respectively.

Suppose for demonstration purposes that we are going to simulate this system for n entities. Figure 1 shows this situation as an influence diagram. We represent random or uncertain quantities by circular nodes, decisions by square nodes, objectives by octagonal nodes, and calculated quantities by double-walled circles. $W_1, W_2, \ldots, W_n$ are the waiting times of the n simulated entities, and $\bar{W}_n$ is their average. $W_1, W_2, \ldots, W_n$ are calculated nodes because if we know $S_1, S_2, \ldots, S_n$ and $T_1, T_2, \ldots, T_n$, then we can calculate $W_1, W_2, \ldots, W_n$. $\bar{W}_n$ is then also a calculated node given $\lambda, \mu, S_1, S_2, \ldots, S_n$, and $T_1, T_2, \ldots, T_n$. The decision here is the number of servers, and the objective is to minimize costs including both server and waiting costs.

Figure 1 is a good example to highlight the nature of simulation from a Bayesian perspective. When one considers all the inputs to the simulation and our uncertainty about them, the simulation output is just a deterministic calculation. The randomness of the system itself, called stochastic uncertainty in the simulation literature or aleatory uncertainty in the decision and risk analysis literature, is represented by the distributions of $S_1, S_2, \ldots, S_n$ and $T_1, T_2, \ldots, T_n$. The uncertainty about the parameters of these models, called parameter uncertainty in the simulation literature or epistemic uncertainty in the decision and risk analysis literature, is represented by λ and μ.

There is also a possible model uncertainty, usually considered to be uncertainty about the distribution of $S_1, S_2, \ldots, S_n$ and $T_1, T_2, \ldots, T_n$, but also including uncertainty about the validity of the calculations used to obtain $W_1, W_2, \ldots, W_n$ and $\bar{W}_n$ from $S_1, S_2, \ldots, S_n$ and $T_1, T_2, \ldots, T_n$. There are two cases of interest here.

Figure 1   **Influence Diagram of a Decision About the Number of Servers in an $M/M/s$ Queue Including the Inputs to the Simulation**
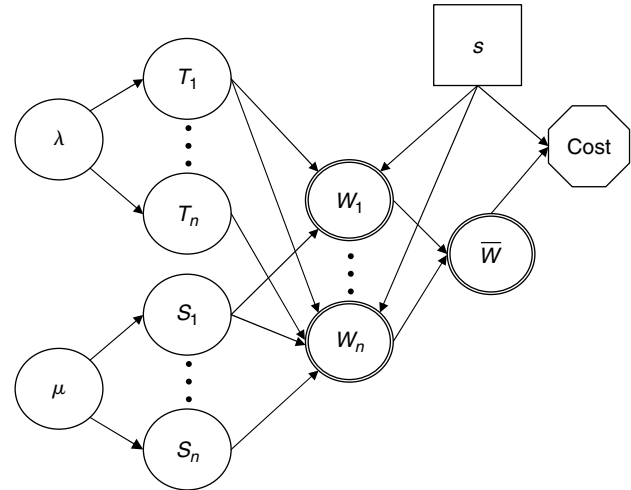

If we are interested in a fixed n or if we wish to simulate for a fixed time (making the number of entities that arrive in the system stochastic), then we have a terminating simulation. If we are interested in the distribution of $\bar{W}_n \to \theta$ as $n \to \infty$, then we have a steady-state simulation. We refer the reader to Banks et al. (2005, Chapter 11) for further discussion of these distinctions.

Evidently, $\bar{W}_n \mid \lambda, \mu$ (or $\theta \mid \lambda, \mu$ in the steady-state case) has a distribution that is defined by the distributions of $S_1, S_2, \ldots, S_n \mid \mu$ and $T_1, T_2, \ldots, T_n \mid \lambda$ and $W_1, W_2, \ldots, W_n$ given $S_1, S_2, \ldots, S_n$ and $T_1, T_2, \ldots, T_n$ through the law of total probability, namely,

$$\int_{dS_1} \cdots \int_{dS_n} \int_{dT_1} \cdots \int_{dT_n} (\bar{W}_n \mid S_1, \ldots, S_n, \ T_1, \ldots, T_n)$$
$$\cdot (T_1, \ldots, T_n \mid \lambda)(S_1, \ldots, S_n \mid \mu) \, dT_n \cdots dT_1 \, dS_n \cdots dS_1. \quad (1)$$

This expression may then be integrated over $\pi(\lambda \mid t_1, t_2, \ldots, t_p)$ and $\pi(\mu \mid s_1, s_2, \ldots, s_q)$. However, we cannot perform this calculation analytically except in the single server case (McGrath and Singpurwalla 1987), so we need to simulate and we may only sample from (1). Figure 2 shows the influence diagram for a simplification of Figure 1 where the distributions of $S_1, S_2, \ldots, S_n, T_1, T_2, \ldots, T_n, \lambda$, and μ have been integrated into the distributions of $W_1, W_2, \ldots, W_n$; thus, we are only including the outputs of the simulation. Bayesian inference is difficult without a known probability model to define a likelihood; thus, one may

Figure 2   Influence Diagram of a Decision About the Number of Servers in an $M/M/s$ Queue Considering Only Outputs from the Simulation
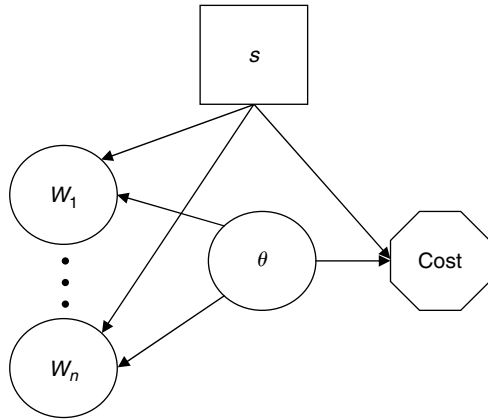
assume a distribution on $\overline{W}_n \mid \lambda, \mu$ (or $\theta \mid \lambda, \mu$), called a metamodel. We define prior distributions on the parameters of this distribution and update the priors given samples from the simulation.

We will discuss different forms of metamodels later, but for now it is critical to realize that the marginal distribution of $W_1, W_2, \ldots, W_n$ is not known, otherwise we would not use simulation. This example shows the nature of Bayesian simulation modeling, but more machinery is necessary to perform such analysis.

3.   Input Analysis

In Bayesian simulation, we incorporate input uncertainty in the analysis to reflect the limited data available to populate the parameters of the arrival processes in a simulation model. Suppose we consider arrivals to the system with n separate arrival processes. These arrival processes can often be modeled by the standard renewal process (Law and Kelton 2001), with a probability distribution chosen to model the interarrival times. We calculate historical interarrival times from data. Let $T_1^k, T_2^k, \ldots, T_{m^k}^k$ be the m^k conditionally independent interarrival times for the kth random input process ($k = 1, 2, \ldots, n$). In our example of the previous section, $k = 1$ could represent the arrival time inputs, and $k = 2$ could represent the service time inputs.

3.1.   Classical Input Analysis
In a classical simulation approach, one usually chooses the probability model by determining the

best estimates of the parameters from the input data for several possible families of distributions and comparing the fit of each distribution to the input data using fit statistics such as the Anderson-Darling, Chi-squared, or Kolmogorov-Smirnov statistics (Law and Kelton 2001). Suppose $F_1^k(t \mid \Theta_1^k), F_2^k(t \mid \Theta_2^k), \ldots,$ $F_p^k(t \mid \Theta_p^k)$ are p families of probability distributions, such as the exponential, Weibull, gamma, or lognormal distributions. The superscript k is included throughout because we can model each arrival process by a different probability distribution and will certainly have different parameter values. $\Theta_1^k, \Theta_2^k, \ldots, \Theta_p^k$ are the parameters of the p potential probability distributions for the kth input process. We obtain best estimates of each set of parameters, $\hat{\Theta}_j^k$, from the data $D^k = \{T_1^k = t_1^k, T_2^k = t_2^k, \ldots, T_{m^k}^k = t_{m^k}^k\}$, using maximum likelihood, method of moments, or other estimation procedures. We choose the best-fit distribution by taking either the fitted distribution with the lowest appropriate fit statistic or at least a fitted distribution that the corresponding hypothesis test does not reject and that has desirable properties, such as simple manipulation of the mean or variance.

3.2.   General Bayesian Input Analysis
Under the Bayesian paradigm, we specify prior distributions for the parameters of the postulated distributions, denoted by $\pi_1^k(\Theta_1^k), \pi_2^k(\Theta_2^k), \ldots, \pi_p^k(\Theta_p^k)$, and we use the data to update these priors using the standard Bayesian machinery to obtain posterior distributions denoted by $\pi_1^k(\Theta_1^k \mid D^k), \pi_2^k(\Theta_2^k \mid D^k), \ldots, \pi_p^k(\Theta_p^k \mid D^k)$. One approach here is to use a model selection criterion to choose between potential probability distributions given input data, such as deviance information criteria (Spiegelhalter et al. 2002), Bayes factors (Kass and Raftery 1995), or posterior predictive densities (Gelfand 1996). This approach mirrors the classical approach and is taken, for instance, in Merrick et al. (2005b).

3.3.   Uncertainty About the Input Model
Chick (1999) offers a different approach to such model uncertainty by placing a probability distribution on the input model itself. Chick's model includes a probability that each given distribution is correct. Chick then conditions each distribution on the event that this distribution is the correct one. Chick assigns prior

probabilities on the probability of each distribution being correct for the kth input process, $\pi^k(m)$ for $m = 1, 2, \ldots, p$, as well as prior distributions on each set of parameters given that distribution is correct, $\pi_m^k(\Theta_m^k \mid m)$ for $m = 1, \ldots, p$. One can then find the probability that a given model m is correct given input data using

$$\pi(m \mid D^k) = \frac{P(D^k \mid m)\pi(m)}{\sum_{l=1}^{p} P(D^k \mid l)\pi(l)}. \qquad (2)$$

One can find $P(D^k \mid m)$ using $\int F_m(D^k \mid \theta_m^k)\pi(\theta_m^k)\,d\theta_m^k$, which is the same term one uses in the calculation of Bayes factors. Chick discusses techniques for finding prior distributions on parameters and calculating posteriors, such as above, including using conjugate priors, numerical approximations, and MCMC techniques.

3.4. Semiparametric Analysis: Mixtures of Known Distributions

Cheng and Currie (2003) provide a similar framework, but with a different approach. Rather than assuming that the p models are different, they assume that the input process is a mixture of distributions from the same parametric family. However, the number of components in this mixture is unknown. Cheng and Currie propose an importance sampling procedure for handling posterior analysis, but suggest that other inference methods such as reversible MCMC jumpers (Richardson and Green 1997) could be more efficient at performing this calculation. Thus, rather than assuming that the input distribution must take a simple, parametric form, but one that is unknown, Cheng and Currie (2003) assume that the form is more complex, but can be represented by a mixture of distributions from the same parametric family. Such a finite mixture covers a wider space of possible distributions than any one single-parametric distribution.

3.5. Other Approaches

We have discussed two approaches above to modeling uncertainty in the Bayesian paradigm, namely, placing a prior distribution on the probability that each of a finite number of models is the correct one, and using finite mixtures of either one or many models. These two approaches differ in the nature of the mixing distribution. In one case, the mixing distribution is the prior. In the other case, the mixing

distribution is part of the model, and we place priors on its parameters.

Decision analysts and Bayesian statisticians have used other approaches when there is uncertainty about the probability model. Bayesian nonparametric approaches include Dirichlet and Gaussian processes, where one assumes that the prior distribution on the cumulative distribution function, hazard function, or some other function that defines a general distribution (Dey et al. 1998). Bayesian semiparametric approaches include unknown mixture distributions, such as an extension of Cheng and Currie (2003), to assume an unknown mixing distribution depicted by a Dirichlet process prior on the cumulative distribution function. This implies that the probability model is a Dirichlet process mixture (Kuo 1986, Escobar and West 1995), an approach used recently in the decision analysis literature by Merrick (2007) for expert judgment aggregation.

4. Analyzing Sources of Output Uncertainty

Chick (1997a) first discussed the problem of sampling from input models to obtain correct output samples. The algorithm in Figure 3 propagates input uncertainty to the outputs through multiple replications of the simulation. In our example of the $M/M/s$ simulation, the algorithm would sample a value for λ and μ for each replication. The algorithm would then sample values for $S_1, S_2, \ldots, S_n$ and $T_1, T_2, \ldots, T_n$ given those values of λ and μ within each replication. In this manner, aleatory uncertainty, or the randomness of the system itself, is represented within each replication, whereas epistemic uncertainty, or uncertainty about the parameter values, is represented across multiple replications. This sampling approach becomes obvious when examining Figure 1. λ is relevant to the distributions of all $S_1, S_2, \ldots, S_n$, so we must use one sampled value of λ for sampling one set of values for $S_1, S_2, \ldots, S_n$.

The same is true for μ and $T_1, T_2, \ldots, T_n$. These values of $S_1, S_2, \ldots, S_n$ and $T_1, T_2, \ldots, T_n$ are then used to obtain one replication of $W_1, W_2, \ldots, W_n$ and θ. This is more complex than frequentist simulation, where one estimate of λ and one estimate of μ are used for all replicates of the simulation, meaning that parameter

Figure 3   **Algorithm from Chick (1997a) for Sampling from Outputs of Bayesian Simulations**

For $r = 1, 2, \ldots, n$ replications of the simulation:
1. For each model $m = 1, 2, \ldots, p$.
   (a) Sample values from the posterior distributions for the parameters, $\pi^k(\Theta^k \mid D^k)$, to be used in the rth replication for each input process k.
2. For the rth replication of the simulation:
   (a) Sample random variates for each simulation input given the parameter values sampled in Step 1;
   (b) Generate the simulation output y_r for these random variates.
End loop

Figure 4   **Algorithm from Zouaoui and Wilson (2004) for Estimating Stochastic, Model, and Parameter Variance of Simulation Outputs**

For $r = 1, 2, \ldots, n$ iterations:
1. For each model $m = 1, 2, \ldots, p$.
   (a) Sample values from the posterior distributions for the parameters for the mth model, $\pi_m^k(\Theta_m^k \mid m, D^k)$, to be used in the rth replication for each input process k.
   (b) For $s = 1, 2, \ldots, t$ iterations:
      (i) Sample random variates for each simulation input from the model m given the parameter values sampled in Step 1(a);
      (ii) Generate the simulation output y_{rms} for these random variates.
      End loop
   End loop
End loop

uncertainty is not represented, and output confidence intervals do not fully cover the range of uncertainty in the output (Inoue and Chick 1998).

Chick (2001) extends this general algorithm to handle the input uncertainty framework from Chick (1999) and provides a general sampling algorithm for sampling from the Chick (1999) framework. Although each of these algorithms obtains correct samples from the simulation outputs that are consistent with equivalent probabilistic calculations for non-simulation-based analysis and that are intuitively correct from examination of Figure 1, they do not allow us to decompose the output uncertainty into its constituent pieces, namely, stochastic, model, and parameter uncertainty. Zouaoui and Wilson (2003) provide an extension of the Chick (1997a) algorithm that allows separation of stochastic and parameter uncertainty using Bayesian model averaging (Draper 1995). Zouaoui and Wilson (2004) further extend this approach to separate stochastic uncertainty, model uncertainty, and parameter uncertainty for the model uncertainty framework in Chick (2001).

Figure 4 shows the extension of Figure 3 proposed by Zouaoui and Wilson (2004), which uses each input model and each sampled input model parameter for a fixed number of simulation replications. The final calculation is a weighted mean where the weights are the posterior probability of each model given the input data. We calculate components of the output variance by considering variations for a given model and variations between models. We estimate the output variance due to model uncertainty by

$$\frac{1}{p}\sum_{m=1}^{p}\pi(m \mid D)\left(\bar{\bar{y}}_m - \sum_{m=1}^{p}\pi(m \mid D)\bar{\bar{y}}_m\right)^2,$$

where $\bar{\bar{y}}_m = (1/nt)\sum_{r=1}^{n}\sum_{s=1}^{t}y_{rms}$ and $\pi(m \mid D)$ is the posterior probability of model m given the input data. We estimate the output variance due to the stochastic nature of the simulation by

$$\frac{1}{p}\sum_{m=1}^{p}\pi(m \mid D)\left(\frac{1}{p(t-1)}\sum_{r=1}^{n}\sum_{s=1}^{t}(y_{rms} - \bar{y}_{rm})^2\right)$$

and the output variance due to parameter uncertainty by

$$\frac{1}{p}\sum_{m=1}^{p}\pi(m \mid D)\left(\frac{1}{n-1}\sum_{r=1}^{n}\sum_{s=1}^{t}(\bar{y}_{rm} - \bar{\bar{y}}_m)^2\right.$$
$$\left. - \frac{1}{n(t-1)}\sum_{r=1}^{n}\sum_{s=1}^{t}(y_{rms} - \bar{y}_{rm})^2\right),$$

where

$$\bar{y}_{rm} = \frac{1}{t}\sum_{s=1}^{t}y_{rms} \quad \text{and} \quad \bar{\bar{y}}_m = \frac{1}{nt}\sum_{r=1}^{n}\sum_{s=1}^{t}y_{rms}.$$

5. Output Analysis and Reducing Output Uncertainty

The presence of input uncertainty means there will be uncertainty in the outputs as well. As we have seen, this will include both the stochastic and parameter uncertainty from the input uncertainty propagated through the simulation, and possibly model uncertainty. The data obtained from the simulation in each replication will be the samples from the distribution for each output of the simulation. For example,

one might collect the total number of entities that leave the simulation or the average waiting time in a given queue as an output. One interesting feature of simulation is the ability to influence the variance, whether through running additional simulation replicates of one or more alternative simulated systems, through common random numbers to reduce variance between alternative simulated systems, or through antithetic variables for reducing variance for a single simulated system. One can exploit this ability for output analysis, as discussed in this section, and decision making, as discussed in the following sections.

Chick (1997a) suggests that the decision maker hypothesize a probability model for the random output data, as in Figure 2, and specify her prior beliefs about the parameters of this output model. As we have mentioned, Chick notes that we can think of this as a Bayesian version of metamodeling (Law and Kelton 2001). Chick (1997a) provides a general framework for such a metamodel, including Bayesian expressions for calculating the posterior distribution of the parameters of the output metamodel given simulation replicates, and the posterior predictive distribution of the simulation output given simulation replicates. As an example, suppose we are interested in a count of the number of entities that pass through a given process in the simulation, then the data is the count occurring in each replication of the simulation, denoted N_r, for the rth replication ($r = 1, 2, \ldots, s$ for s replications). Because our example output data is in the form of a count, one model could be a Poisson distribution with rate μ, with a conjugate gamma distributed prior on μ. The predictive distribution of the number of entities passing through our simulated process is then a Poisson-gamma distribution in the sense of Bernado and Smith (2000).

Cheng (1998) fits a Bayesian regression model to simulation output to form a metamodel that one can use in prediction and decision making. The independent variables in the regression model can either be input model parameters, such as λ and μ in our $M/M/s$ example, or decision or control variables, such as s in our $M/M/s$ example. Cheng (1998) allows for an unknown number of nonlinear components in the regression and assumes that the error terms are normally distributed and independent. Thus, we can write the metamodel as

$$Y_j = \sum_{i=1}^{Q} \beta_i f_i(\theta) + z_j,$$

where Y_j is the simulation output for the jth replication, θ is a vector of the control variables for the simulation (either input model parameters or decision variables), $f_i(\theta)$ is the ith regression term, β_i is the regression parameter for the ith regression term, and z_j is the normally distributed error term. Q is the number of terms in the model, and Cheng (1998) assumes it to be uncertain in this work. Cheng (1998) uses a Metropolis-Hastings algorithm for inference with the assumption that any regression parameter taking a value within a specified value ∂ of zero represents a regression component that need not be included. This allows Cheng to estimate the number of components effectively using the derived parameter sample. Bayesian statisticians call this a derived chain MCMC method, because the algorithm does not handle the number of components by sampling from its posterior distribution. Cheng (1999) extends this approach to allow for nonnormal errors, a situation that will commonly occur in queuing simulations.

In the previous section, we discussed estimates of output variance due to stochastic, model, and parameter uncertainty. Ng and Chick (2001) address the issue of optimally reducing output uncertainty by collecting additional input data. Ng and Chick (2001) propose a regression metamodel where the simulation output follows a normal distribution with a mean that is a function of the input distribution parameters. They then find the predictive distribution of the simulation output by integrating the regression metamodel over the posterior distributions of each input distribution parameter given available input data. The posterior distribution on the simulation input parameters is asymptotically multivariate normal. The use of asymptotic approximations does limit the applicability of the approach to cases where there is sufficient input data. Similarly, the simulation output is asymptotically normal with a mean that one can approximate locally as a linear sum of the input parameters. This allows a closed-form approximation of the parameter uncertainty based on the input data

already available and the simulation replications performed thus far and written as a function of the number of additional input data points that are to be collected, denoted $V_{\text{par}}(m_1, m_2, \ldots, m_K)$, where m_k is the number of additional input data points to be collected for the kth input process. Again, though, the asymptotic approximation limits use to cases where one has performed a sufficiently large number of initial simulation replications prior to these calculations. Ng and Chick (2001) define the optimization problem as determining the optimal number of additional data points for each input distribution to minimize the output variance subject to a constraint on the amount of data collection:

$$\min \quad V_{\text{par}}(m_1, m_2, \ldots, m_K)$$
$$\text{s.t} \quad m_k \geq 0$$
$$\sum_{k=1}^{K} c_k m_m \leq B,$$

where c_k is the cost of each additional input data point for the kth input process and B is the total budget available for additional input data collection. Ng and Chick (2006) extend this approach to nonnormal distributions from the exponential family using Bayesian model averaging.

Beyond reducing output uncertainty by collecting additional input data (Ng and Chick 2001, 2006), Ng and Chick (2004) determine how to optimally allocate simulation replications, specifically through a Bayesian design of follow-on simulation experiments. The approach can be used to allocate additional simulation replications for given parameter values or for given control variable values or decision alternatives. One aim of the allocation is to differentiate between the simulation models that result from different parameter or control variable values. The other aim is to enhance parameter estimation for the metamodel. The metamodel used is again a linear regression with the input parameters and control variables (decisions) as independent variables and the simulation output as the dependent variable. The approach is fully decision analytic with a Bayesian analysis of the regression model and the design decision made by maximizing expected utility. The utility function used is interesting from a decision analysis point of view.

Previous work has used Shannon's information measure, or entropy, as an experimental design criterion or utility function (Abbas 2004). One version averages the entropy difference between each possible pair of models, a utility function that leads to good model discrimination in Bayesian experimental design; this utility function is often denoted by MD. Another version averages the entropy across all possible models, a utility function that leads to good parameter estimation; this utility function is often denoted by SP. The utility function proposed by Ng and Chick (2004) can be written as

$$SQ = w_{MD} \frac{MD - MD_{\min}}{MD_{\max} - MD_{\min}} + w_{SP} \frac{SP - SP_{\min}}{SP_{\max} - SP_{\min}},$$

where w_{MD} and w_{SP} are the weights for MD and SP, respectively, and $w_{MD} + w_{SP} = 1$. The utility function SQ is a standard additive linear utility function combining two attributes: MD to measure the objective of maximizing model discrimination and SP to measure the objective of maximizing parameter estimation. We scale each attribute to the $[0, 1]$ interval, and the weights represent the relative importance of each objective to the decision maker. From a decision analysis perspective, this obviously assumes a linear preference scale for each attribute and additive utility independence of the two attributes (Keeney and Raiffa 1993). Ng and Chick (2004) provide a closed-form approximation of the expected utility for normal errors that numerically maximize this term to find the optimal design.

The introduction of multiattribute techniques and entropy methods in designing simulation experiments suggests an avenue for research by decision analysts. Are MD and SP additive utility independent, or is another multiattribute utility form more suitable? Are MD and SP the best entropy formulations to use? Chick (2004) also points out that another interesting problem is determining optimal simulation replication plans for reducing uncertainty in the output metamodel. There has been significant frequentist work in this area and work for deterministic models using a technique called Kriging (see van Beers and Kleijnen 2008). Bayesian experimental design (Chaloner and Verdinelli 1995) for simulations is also a promising area for future research.

We have not addressed several additional areas thus far. Andradóttir and Bier (2000) briefly discuss

the issue of model validity from the Bayesian perspective. Suppose Y is an output of a simulation for which we have posterior predictive distribution $\pi(Y \mid C)$ after observing some number of simulation replications, denoted by C. Now suppose Z is the same output for the real system, and $\pi(Z \mid D)$ is its posterior predictive distribution after observing real-world data, denoted by D. What does it mean if $\pi(Y \mid C)$ and $\pi(Z \mid D)$ do not have posterior mass at similar values of Y and Z, respectively? Is this due to misspecification of the model, either the calculations or the choice of input models, or misspecification of the distributions of the input parameters? Andradóttir and Bier (2000) suggest that we should jointly update the input and output distributions to learn about both possibilities, but the usual flow of information is in the input-to-output direction. To update the priors on the parameters of the input distributions given unexpected output samples would require a reversal of this route of information. One possibility is the use of metamodels, such as those discussed in this section, to flow information about outputs back to the inputs. Work in this area is ongoing and should prove interesting if this issue can be resolved.

An alternative approach that might appeal to decision analysts would be to treat the simulation output Y as expert forecasts that provide additional information to update the decision maker's knowledge about Z. The work in aggregation of expert forecasts allows such updating (for a survey of such approaches, see Clemen and Winkler 1999) and allows for bias in the forecasts. The difference here is that the usual forecast aggregation assumes that the expert provides a full distribution of Y, but a simulation can only provide samples from the distribution of Y. This will require modification of previous work that decision analysts could make.

Last, we point out that just as input model uncertainty can be approached using Bayesian nonparametric and semiparametric approaches, so too can output model uncertainty. Dirichlet process, gamma process, and Dirichlet process mixtures, for example, are also suitable output metamodels, either on individual outputs or as part of response surface models for outputs or regression models on output. Gelfand (1996) and Merrick et al. (2003, 2005a) discuss approaches for nonparametric and semiparametric Bayesian regression.

6. Making Decisions with Bayesian Simulations

Thus far we have discussed the construction of a Bayesian simulation and several methods for making decisions *about* a simulation, specifically what input data to collect and what simulation replications to perform to meet various objectives. In this section, we will discuss making decisions *with* simulations. In recent years, researchers have performed considerable work on algorithms for selecting the best simulated system. Given a set of alternatives, one performs simulation runs for each alternative, and researchers have developed statistical procedures to assess whether there is sufficient evidence in favor of any one alternative (Matejcik and Nelson 1995; Bechhofer et al. 1995; Goldsman and Nelson 1998; Chick and Inoue 2001a, b). The frequentist literature in this area concentrates on choosing the simulated system with the highest (or lowest) mean for a given output. The Bayesian work in this area falls into two camps. In §6.1, we discuss decision analytic work that concentrates on finding the simulated system with the highest expected utility, for both single-attribute and multiattribute problems. In §6.2, we discuss work that concentrates on using Bayesian machinery to develop algorithms that are often better than the frequentist algorithms at finding the simulated system with the best mean for a given output.

## 6.1. Including a Simulation in the Analysis of a Decision

Consider the problem of selecting the best simulated system under the decision analytic paradigm, namely, choosing d^* from a set of alternatives $\Delta = \{d_1, d_2, \ldots\}$ that maximizes

$$\max_{d \in \Delta}\left[V(d) = \int u(d, X)p_d(X)\, dX \right], \qquad (3)$$

where $u(d, X)$ is the decision maker's utility function and $p_d(X)$ is the distribution of simulation output X under decision d that influences the decision maker's utility. We may consider this problem as an influence diagram in Figure 5.

In a more general decision analytic setting, the conditional distribution $p_d(X)$ has a known functional form. However, in the case of simulation decision problems, we cannot calculate $p_d(X)$, but we can sample from it by running an iteration of the simulation.

Figure 5   Influence Diagram for Selecting the Best Simulated System
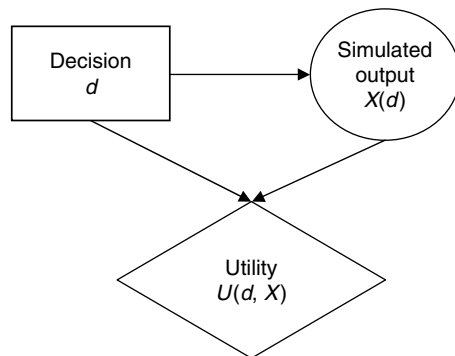
Thus, we cannot calculate $V(d)$ in closed form. Note that this influence diagram depicts the real-world decision to be made as discussed in this section, not the decision of how much to sample from each simulation alternative; the sampling decision is discussed in the next section.

Chick and Gans (2009) perform an analysis that is probably most appealing to practicing decision analysts. Rather than assuming that the objective of minimizing costs applies just to the real system, Chick and Gans (2009) minimize the cost of both the decision situation and the simulation project itself. The cost of input data collection and the cost of running additional simulation replicates are included along with the cost and revenues from the decision about the system that the analyst has simulated. In this manner, we can consider work to improve the validity of the simulation through its impact on the eventual decision, or its value of information. Although Chick and Gans (2009) assume the decision maker to be risk neutral, such a purely economic analysis would be appealing if the decision maker could accurately estimate all costs.

Chick (1997b) provides the first formulation of selecting the best simulated system by maximizing expected utility. Chick (1997b) derives expressions for the probability that the expected utility of System A is greater than the expected utility of System B with a general utility function. In an example, Chick (1997b) assumes a normal metamodel for the simulation output for independent replications, then a multivariate normal metamodel for dependent replications. Dependent replications allow the use of common random numbers for variance reduction (Law and Kelton

2001). This increases the utility dominance probabilities for a given number of simulation replicates and, therefore, discrimination between the expected utilities of the alternative simulated systems. Chick (1997b) compares the Bayesian dominance probability to the frequentist p-value under the assumption of equal means.

Often simulations have multiple outputs that represent multiple attributes involved in the decisions. This situation suggests the use of multiattribute utility functions to represent such trade-offs and the associated attitude to multiattribute risk. Morrice et al. (1998) first proposed such an approach, with later extensions and elaborations in Morrice et al. (1999) and Butler et al. (2001). This work extends the approach of Chick (1997b) to consider multiattribute utility functions. However, the approach uses a frequentist ranking and selection procedure to find the system with the best mean utility score based on the output samples.

Morrice et al. (1998) solve a fully decision theoretic problem using frequentist approximations to find the best alternative. This raises an interesting point, namely, that even with a decision analytic setup using Bayesian simulation and maximizing expected utility, one may obtain samples from $x_1, x_2, \ldots, x_n$ from $p_d(X)$ by running the simulation for decision d, calculate $u(d, x_1), u(d, x_2), \ldots, u(d, x_n)$, and estimate $V(d)$ using $\bar{u}(d) = (1/n) \sum_{i=1}^{n} u(d, x_i)$.

Andradóttir and Kim (2007) criticize Butler et al. (2001) because their approach does not explicitly handle constraints, because the ranking and selection procedure used is inefficient for large numbers of alternative simulated systems, and because Andradóttir and Kim (2007) believe it is difficult to find an appropriate multiattribute utility function. Decision analysts would disagree with this last point, and Morrice and Butler (2006) have extended the approach to model constraints using value functions. A hard constraint value function takes full value when the constraint is met and zero value when it is not. A soft constraint value function smoothly decreases value in a neighborhood around the constraint boundary. We should also note that the use of multiattribute utility functions in all this work does not require the use of the ranking and selection procedure from Butler et al.

(2001); one can use any ranking and selection procedures, including the decision analytic methods discussed in the next section.

Again, there appear to be several directions in which decision analysts could continue this research. A simulation is just one tool that a decision maker can use. The work of Chick and Gans (2009) shows that decisions about the simulation are just the same as decisions about obtaining other sources of information. This suggests approaches using the expected value of sample (simulation) information. As we discussed at the end of the previous section, this also suggests that one can consider simulations as analogous to experts, although a simulation provides samples rather than full distributions. The work by Morrice and Butler (2006) in multiattribute methods may also lead to further avenues of research for decision analysts.

6.2. Ranking and Selecting the Best Simulated System

The fact that the expected utility can be estimated by a mean output implies that any of the methods for selecting the highest mean output can be used to find the best simulated alternative, whether frequentist or Bayesian. Thus, we now turn our attention to such methods that rely on Bayesian machinery and their comparison to frequentist methods.

All approaches discussed in this section are based on the same basic setup. Suppose we have to decide between K simulated systems. Let $X_{i,j}$ be the jth replicate of the output of interest for simulated system i, for $i = 1, 2, \ldots, K$ and $j = 1, 2, \ldots$. We assume that $X_{i,j}$ follows a normal distribution with mean w_i and precision λ_i. We denote the ordering of the means from lowest to highest by $w_{[1]} < w_{[2]} < \cdots < w_{[K]}$. At any given stage of performing simulation replicates, we have n_i replicates of simulation i and can estimate w_i by the sample mean $\bar{x}_i = (1/n_i) \sum_{j=1}^{n_i} x_{i,j}$. We denote the current ordering of the sample means by $\bar{x}_{(1)} < \bar{x}_{(2)} < \cdots < \bar{x}_{(K)}$. Thus, (i) denotes the ith largest sample mean, and $[i]$ denotes the ith largest true mean. At the end of any given algorithm, we have selected the best simulated system if the system selected, denoted by $D = (K)$, has the highest true mean, $D = [K]$. Let $C = \{X_{i,j}, i = 1, 2, \ldots, K, j = 1, 2, \ldots, n_i\}$ be the set of all simulation replicates performed thus far.

Inoue and Chick (1998) define a Bayesian probability of correct selection, defined as $P(W_{(i)} < W_{(K)}, \forall i \neq K \mid C)$, or the probability that the currently chosen system $D = (K)$ has a true mean that is greater than all other systems given all current simulation replications. They show that the Bayesian probability of correct selection is greater than the frequentist equivalent based on Bonferroni approximations. However, the calculation for the Bayesian probability is based on Monte Carlo simulation from the posterior distributions of the w_i, a computationally expensive calculation. Chen (1996) provides a lower bound for the probability of correct selection. Chen defines the lower bound as

$$\prod_{i=1}^{K-1} P(W_{(i)} < W_{(K)} \mid C),$$

or the probability that the true mean for the system currently chosen as the best so far does have a higher mean that all other systems, given all simulation replicates observed thus far. Chen et al. (1996) uses this lower bound for the probability of correct selection to allocate a fixed number of simulation replications across the set of alternative systems. The procedure has two stages. The first stage runs a fixed and equal number of simulation replications to allow estimation of the lower bound on the probability of correct selection. The procedure then optimally allocates simulation replications for a second stage to alternative simulated systems using the steepest ascent optimization method and an approximation of the gradient of the objective function, the lower bound on the probability of correct selection. Chen et al. (1997) improves on Chen et al. (1996) by using a simpler gradient approximation. The authors show that the simpler method provides superior results and call the method optimal computing budget allocation (OCBA).

Chick and Inoue (2001a) propose a procedure with a similar first round of an equal number of simulations of each system and show that the posterior of each mean is a Student t-distribution. They propose the use of Bayesian loss functions with two specific forms. The 0–1 loss function,

$$L_{0\text{-}1}(i) = \begin{cases} 0 & w_{(K)} = w_{[K]}, \\ 1 & \text{otherwise}, \end{cases}$$

takes the value 0 for the best system and 1 for all other systems. Using the 0–1 loss function, the expected loss for choosing the system with the highest current estimated mean is

$$E_{\mathbf{w}|C}[L_{0\text{-}1}(D=(K)) \mid C] = 1 - P(W_{(i)} < W_{(K)}, \; \forall i \neq K \mid C),$$

or the probability that system i is not the best system. The linear loss function, also called the opportunity cost and written

$$L_{OC}(i) = w_{[k]} - w_i,$$

takes a value equal to the mean of the best system minus the mean of the system considered. In this case, the expected loss for choosing the system with the highest current estimated mean is equal to

$$E_{\mathbf{W}|C}[L_{OC}(D=(K)) \mid C] = E_{\mathbf{W}|C}[W_{[K]} - W_{(K)} \mid C],$$

the expected difference between the best system and system i.

Chick and Inoue (2001a) develop approximations for estimating the expected loss in each of these cases given each possible number of additional replications of each possible system. They then allocate replications to systems by minimizing the expected loss when the eventual decision will be to pick system $D = (K)$ with the highest overall mean. One version of the linear loss algorithm adds the cost of replications to the expected loss in the minimization (called LL) and another adds a constraint on the total cost of replications (called LL(B)). Similarly, Chick and Inoue (2001a) give two versions of the algorithms for the 0–1 loss function, called 0–1 and 0–1(B).

Chick and Inoue (2001b) assume a multivariate normal distribution of the mean output of each alternative system. This allows the use of common random numbers in the simulation replications to reduce output uncertainty and, therefore, sharpen comparisons. Chick and Inoue (2001b) use a Bonferroni-type approximation of the Bayesian probability of correct selection for this dependent case. Again, Chick and Inoue (2001b) develop versions of their algorithms with both linear loss and 0–1 loss functions.

Branke et al. (2007) compare a frequentist procedure called $KN + +$ (Goldsman et al. 2002) that has performed well in empirical testing to Bayesian procedures based on 0–1(B), LL(B), and OCBA. They make

each procedure purely sequential where each stage allocates only one additional replication. Branke et al. (2007) implement each procedure with three different stopping rules. One stopping rule is reaching the total allowed budget, which is the same as considered before in 0–1(B), LL(B), and OCBA. However, the authors add two additional stopping rules, one where the probability that the best system is ∂^* better than that currently chosen (called the probability of bad selection) reaches a sufficiently small level, and one where the expected opportunity cost lost reaches a sufficiently small level. The test problems include the following:

• a slippage configuration, where all systems have equal means except the best, which is better by a parameter ∂^* divided by the noise;

• monotone decreasing means, where each successive mean has a linearly decreasing mean;

• random problem instances, where the means and precisions of the systems are sampled from a normal-gamma distribution in one version and an exponential-gamma distribution in another version.

For the slippage configuration and the monotone decreasing means, LL and OCBA (modified to consider linear loss) are superior when used with the expected opportunity cost stopping rule. For the random problem instances, the LL, OCBA, and the frequentist procedure all work well with the probability of bad selection stopping rule. However, the frequentist procedure is not as good as the number of systems considered increases. Branke et al. (2007) conclude by recommending LL and OCBA with linear loss as the best overall procedures, and using either the probability of bad selection or the expected opportunity cost stopping rules depending on the desired goal. Thus, based on the latest testing of procedures, decision theoretic methods provide the best procedures for finding the simulated system with the highest mean output.

As can be seen in this section, algorithms for choosing optimal decisions from simulations have focused on statistical loss (utility) functions, such as 0–1 and LL. The multiattribute work of Butler et al. (2001) and the economic work of Chick and Gans (2009) have taken steps in the direction of decision analysis, but considerable further work is possible. Because we can depict the simulation decision problem as

an influence diagram, algorithms for solving influence diagrams may be applicable, such as augmented probability sampling (Bielza et al. 1999). Extensions to the multiattribute case may also be possible by depicting utility dependence using the utility diagrams of Abbas and Howard (2005) along with influence diagrams. Algorithms specifically designed for the multiattribute case may work better for such problems than simply applying the single-attribute algorithms. It is also noteworthy that methods developed thus far have concentrated on discrete sets of alternatives and not ventured into continuous decision variables.

7. A More Detailed Application

To provide a more detailed application example, we will discuss our use of Bayesian methods to assess uncertainty in simulation-based maritime risk assessment. Merrick et al. (2005b) and Merrick and van Dorp (2006) provide a full description of this work; here we will give an illustrative summary.

Risk management has become a major part of operating decisions for companies in the maritime transportation sector, and thus an important research domain (National Research Council 2000). The Exxon Valdez disaster cost Exxon $2.2 billion in clean up costs alone. This led to the immediate questions of how to prevent such accidents in the future and how to mitigate their consequences if they should occur. The Prince William Sound Risk Assessment (Merrick et al. 2000, 2002), Washington State Ferries Risk Assessment (van Dorp et al. 2001), and an exposure assessment for ferries in San Francisco Bay (Merrick et al. 2003) are three examples of successful risk studies in this domain. Decision makers have used their results in major investment decisions and have played a significant role in the management of maritime transportation in the United States. In a maritime port system, traffic patterns change over time in a complex manner. The dynamic nature of these traffic patterns and, indeed, other situational variables, such as wind, visibility, and ice conditions, mean that risk levels change over time. This requires the use of simulation to model the impact of changes that affect the traffic patterns accurately, such as introducing new traffic rules and increases or decreases in the volume of traffic in a given port. However, for decision makers to make well-informed decisions,

they must understand not only risk estimates of the alternative simulated systems, but also the stochastic and parameter uncertainty in such estimates.

Merrick et al. (2005b) use the simulation to count the frequency of situations in which the ferries could be involved in an accident. They define these situations through such factors as the proximity of other vessels, what type of vessels they are, the presence of fog or high wind, and the training and experience of the crew. We may then define the probability of a collision using the law of total probability,

$$P(Collision) = \sum_{j=1}^{k} P(Collision \mid Situation_j)P(Situation_j), \tag{4}$$

where $Situation_j$ denotes the jth of all possible situations for $j = 1, \ldots, k$, and k is the total number of possible combinations. One can then calculate the expected yearly frequency of collisions by multiplying the probability in (2) by the number of simulated periods in a year. A simulation of the maritime system is used to estimate $P(Situation_j)$, whereas $P(Collision \mid Situation_j)$ is estimated using a combination of expert judgment and accident data (Merrick et al. 2006).

In this risk assessment methodology, the data obtained from the simulation in each replication will be a count of the number of vessel interactions occurring in each replication of the simulation. Following the Bayesian approach, Merrick et al. (2005b) hypothesize a Poisson metamodel for the random output data and specifies our prior beliefs about the parameters of this output model using the conjugate gamma distribution. The predictive distribution of $P(Situation_j)$ is then a Poisson-gamma distribution in the sense of Bernado and Smith (2000). Note that the aleatory uncertainty here can be reduced by running longer simulations, the epistemic uncertainty cannot; this would require additional traffic data.

We start by examining the number of situations that could occur under each alternative, a result from Merrick et al. (2005b). Figure 6(a) shows an aggregate comparison of the alternatives by the total expected yearly number of situations, in this case when a ferry is close enough to other vessels that the situation could lead to a collision.

The lines in Figure 6(a) are actually box plots showing the predictive distribution with the interquartile
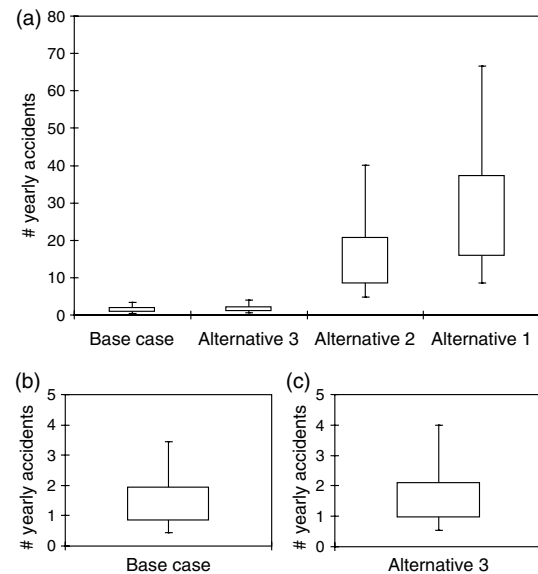
Figure 6   Expected Yearly Situations Comparison

Figure 7   Expected Yearly Collisions Comparison
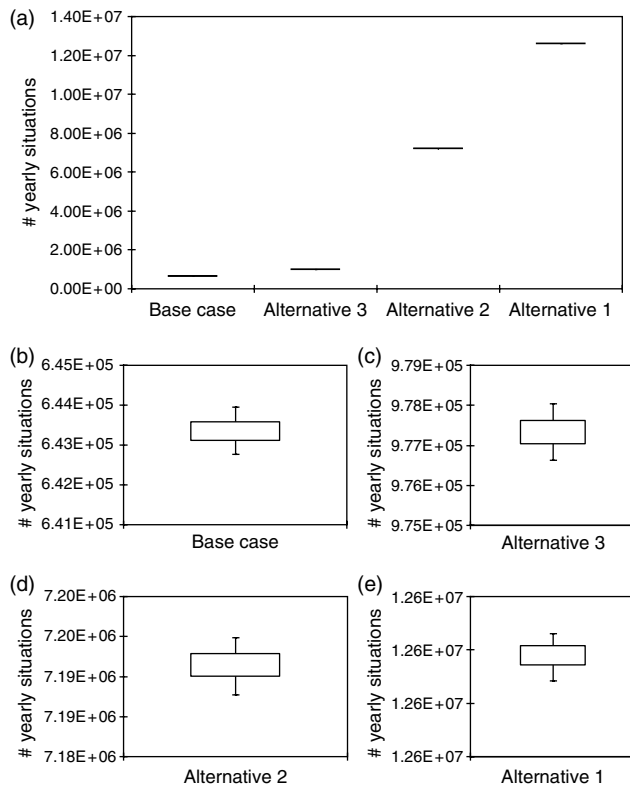
range as the box and the 5th and 95th percentiles of the distribution as the whiskers. However, because the remaining uncertainty in these estimates is low, they do not show up on this comparison scale, and we repeat them in Figures 6(b)–6(e) using different scales. It is evident from Figure 6 that there is an increase in the number of situations across the alternatives and that the amount of uncertainty is small relative to the size of the differences between the alternatives.

Because there is wide variability and significant uncertainty about $P(Collision \mid Situation_j)$, the analysis from Merrick et al. (2005b) is useful, but not definitive. Instead, we must examine the collision probability itself, $P(Collision)$. Figure 7(a) shows a similar pattern of increase for the expected yearly number of collisions as seen for the expected yearly situations. However, with the introduction of estimated collision probabilities based on expert judgments following Merrick et al. (2006), there is significantly more uncertainty evident in these results, and we cannot remove this uncertainty by simply running more simulations. The largest uncertainty remains about

Alternatives 2 and 1. However, there are almost certainly a higher expected number of collisions in Alternative 1 than Alternative 2. There is not such certainty when comparing the base case to Alternative 3.

Whereas there was an almost certain ranking in terms of the expected yearly number of situations, this is not true for the expected yearly number of collisions. Because the comparison is not clear on a scale that includes Alternatives 2 and 1, Figures 7(b) and 7(c) show the box plots for the base case and Alternative 3, respectively; the 90% credibility intervals for the two alternatives are $(0.45, 3.44)$ for the base case and $(0.54, 3.99)$ for Alternative 3. These distributions do indeed overlap, and the best we can say is that Alternative 3 stochastically dominates the base case in the sense that their cumulative distribution functions do not cross.

8.   Conclusions and Areas for Research by Decision Analysts

We have reviewed the literature on Bayesian simulation and covered many types of decisions that one can address using Bayesian methods. These decisions have included which input data to collect to minimize output uncertainty, which simulation replications to run to differentiate between alternative systems, which simulated system has the highest mean output, and which simulated system has

the highest expected utility, including both single-attribute and multiattribute cases.

In each section, we have discussed possible avenues for further research by decision analysts. In the area of simulation input modeling, we pointed out opportunities for Bayesian nonparametric and semiparametric approaches for handling model uncertainty. This would also require extensions of the BMA algorithms for analyzing sources of uncertainty in simulation outputs. In the area of output analysis, possible developments include the following:

• work on entropy utility functions and alternative multiattribute formulations for designing simulation experiments to maximize information,

• work on updating simulation inputs when unexpected simulation outputs are observed,

• work on treating simulation outputs as expert forecasts in the form of samples instead of full distributions,

• the use of Bayesian nonparametric and semiparametric approaches for handling model uncertainty for outputs.

In the area of making decisions using simulations, there are opportunities in multiattribute simulation decision making and in treating decisions about the simulation as part of the overall real-world decision that the analyst has built the simulation to help with. In choosing the best simulated system, algorithms used in choosing the best alternative for influence diagrams may prove fruitful, and one can develop specific algorithms for multiattribute simulation problems.

The aim in writing this review was to introduce the area of Bayesian simulation to the decision analysis community in the hope that they can apply more decision analytic tools and approaches in the area. We hope that we have provided enough background to give a general understanding of the area and that we have identified enough open problems to wet the appetites of decision analysts to work in this area. No doubt, readers will have identified opportunities beyond those mentioned. We look forward to seeing work on Bayesian simulation in the decision analysis literature.

Acknowledgments

References

Abbas, A. E. 2004. Entropy methods for adaptive utility elicitation. *IEEE Trans. Systems, Man, Cybern. A* **34**(2) 169–178.

Abbas, A. E., R. A. Howard. 2005. Attribute dominance utility. *Decision Anal.* **2**(4) 185–206.

Andradóttir, S., V. M. Bier. 2000. Applying Bayesian ideas in simulation. *Simulation Theory Practice* **8** 253–280.

Andradóttir, S., S. H. Kim. 2007. Fully sequential procedures for comparing constrained systems via simulation. Technical report, H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta.

Banks, J., J. S. Carson, B. L. Nelson, D. M. Nicol. 2005. *Discrete-Event System Simulation*. Prentice Hall, Upper Saddle River, NJ.

Bechhofer, R. E., T. J. Santner, D. M. Goldsman. 1995. *Design and Analysis of Experiments for Statistical Selection, Screening, and Multiple Comparisons*. John Wiley & Sons, New York.

Bernado, J., A. Smith. 2000. *Bayesian Theory*, 2nd ed. Wiley, Chichester, UK.

Bielza, C., P. Müller, D. R. Insua. 1999. Decision analysis by augmented probability simulation. *Management Sci.* **45**(7) 995–1007.

Branke, J., S. E. Chick, C. Schmidt. 2007. Selecting a selection procedure. *Management Sci.* **53**(12) 1916–1932.

Butler, J., D. J. Morrice, P. W. Mullarkey. 2001. A multiple attribute utility theory approach to ranking and selection. *Management Sci.* **47**(6) 800–816.

Chaloner, K., I. Verdinelli. 1995. Bayesian experimental design: A review. *Statist. Sci.* **10**(3) 273–304.

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

Chen, C.-H., H.-C. Chen, L. Dai. 1996. A gradient approach for smartly allocating computing budget for discrete event simulation. J. M. Charnes, D. J. Morrice, D. T. Brunner, J. J. Swain, eds. *Proc. 28th Winter Simulation Conf.*, IEEE Computer Society, Washington, DC, 398–405.

Chen, C.-H., H.-C. Chen, L. Dai, E. Yucesan. 1997. New development of optimal computing budget allocation for discrete event simulation. S. Andradóttir, K. J. Healy, D. H. Withers, B. L. Nelson, eds. *Proc. 29th Winter Simulation Conf.*, IEEE Computer Society, Washington, DC, 334–341.

Cheng, R. 1998. Bayesian model selection when the number of components is unknown. D. J. Medeiros, E. F. Watson, J. S. Carson, M. S. Manivannan, eds. *Proc. 30th Winter Simulation Conf.*, IEEE Computer Society Press, Los Alamitos, CA, 653–660.

Cheng, R. 1999. Regression metamodeling in simulation using Bayesian methods. P. Farrington, H. Nembhard, D. Sturrock, G. Evans, eds. *Proc. 31st Winter Simulation Conf.: Simulation—A Bridge to the Future*, Vol. 1, Association for Computing Machinery, New York, 330–335.

Cheng, R. C., C. S. Currie. 2003. Simulation input modeling: Prior and candidate models in the Bayesian analysis of finite mixtures. *Proc. 35th Winter Simulation Conf.: Driving Innovation, New Orleans*, 392–398.

Chick, S. 1997a. Bayesian analysis for simulation input and output. S. Andradóttir, K. J. Healy, D. H. Withers, B. L. Nelson, eds. *Proc. 29th Winter Simulation Conf.*, IEEE Computer Society, Washington, DC, 253–260.

Chick, S. 1997b. Selecting the best system: A decision theoretic approach. S. Andradóttir, K. J. Healy, D. H. Withers, B. L. Nelson, eds. *Proc. 29th Winter Simulation Conf.*, IEEE Computer Society, Washington, DC, 326–333.

Chick, S. 1999. Steps to implement Bayesian input distribution selection. P. Farrington, H. Nembhard, D. Sturrock, G. Evans, eds. *Proc. 31st Winter Simulation Conf.: Simulation—A Bridge to the Future*, Vol. 1, Association for Computing Machinery, New York, 317–324.

Chick, S. 2000. Bayesian methods for simulation. J. A. Joines, R. R. Barton, K. Kang, P. A. Fishwick, eds. *Proc. 2000 Winter Simulation Conf.*, Society for Computer Simulation International, San Diego, 109–118.

Chick, S. 2001. Input distribution selection for simulation experiments: Accounting for input uncertainty. *Oper. Res.* **49**(5) 744–758.

Chick, S. 2004. Bayesian methods for discrete event simulation. R. G. Ingulls, M. D. Rossetti, J. S. Smith, B. A. Peters, eds. *Proc. 2004 Winter Simulation Conf., Washington, DC*, 89–100.

Chick, S. 2006a. Subjective probability and Bayesian methodology. S. G. Henderson, B. L. Nelson, eds. *Simulation*. Elsevier, New York.

Chick, S. 2006b. Bayesian ideas and discrete event simulation: Why, what and how. L. F. Perrone, B. G. Lawson, J. Liu, F. P. Wieland, eds. *Proc. 38th Winter Simulation Conf., Monterey, CA*, 96–105.

Chick, S. E., N. Gans. 2009. Economic analysis of simulation selection problems. *Management Sci.* **55**(3) 421–437.

Chick, S. E., K. Inoue. 2001a. New procedures to select the best simulated system using common random numbers. *Management Sci.* **47**(8) 1133–1149.

Chick, S. E., K. Inoue. 2001b. New two-stage and sequential procedures for selecting the best simulated system. *Oper. Res.* **49**(5) 732–743.

Chick, S. E., S. Soorapanth, J. S. Koopman. 2003. Inferring infection transmission parameters that influence water treatment decisions. *Management Sci.* **49**(7) 920–935.

Clemen, R. T., R. L. Winkler. 1999. Combining probability distributions from experts in risk analysis. *Risk Anal.* **19**(2) 187–203.

Dey, D., P. Muller, D. Sinha. 1998. *Practical Nonparametric and Semi-Parametric Bayesian Statistics*. Springer, New York.

Draper, D. 1995. Assessment and propagation of model uncertainty (with discussion). *J. Royal Statist. Soc.: Ser. B* **57**(1) 45–97.

Escobar, D. M., M. West. 1995. Bayesian density estimation and inference using mixtures. *J. Amer. Statist. Assoc.* **90** 577–588.

Gamerman, D., H. F. Lopes. 2006. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, 2nd ed. Chapman & Hall, Boca Raton, FL.

Gelfand, A. 1996. Model determination using sampling based methods. W. Gilks, S. Richardson, D. J. Spiegelhalter, eds. *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London, 145–161.

Gilks, W., S. Richardson, D. Spiegelhalter. 1996. *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London.

Goldsman, D., B. L. Nelson. 1998. Comparing systems via simulation. J. Banks, ed. *Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice*. John Wiley & Sons, New York.

Goldsman, D., S.-H. Kim, W. S. Marshall, B. L. Nelson. 2002. Ranking and selection for steady-state simulation: Procedures and perspectives. *INFORMS J. Comput.* **14**(1) 2–19.

Henderson, S., B. Nelson. 2006. *Handbooks in Operations Research and Management Science Volume 13: Simulation*. Elsevier, Amsterdam.

Herzog, T. N. 2002. Bayesian statistics and the Monte Carlo method. E. Yücesan, C.-H. Chen, J. Snowdon, J. Charnes, eds. *Proc. 2002 Winter Simulation Conf., San Diego*, 136–146.

Inoue, K., S. E. Chick. 1998. Comparison of Bayesian and frequentist assessments of uncertainty for selecting the best system. D. J. Medeiros, E. F. Watson, J. S. Carson, M. S. Manivannan, eds. *Proc. 30th Winter Simulation Conf.*, IEEE Computer Society Press, Los Alamitos, CA, 727–734.

Kass, R., A. Raftery. 1995. Bayes factors. *J. Amer. Statist. Assoc.* **90**(430) 773–795.

Keeney, R. L., H. Raiffa. 1993. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. Cambridge University Press, Cambridge, UK.

Kuo, L. 1986. Computation of mixtures of Dirichlet processes. *SIAM J. Sci. Statist. Comput.* **7** 60–71.

Law, A., W. Kelton. 2001. *Simulation Modeling and Analysis*, 3rd ed. McGraw Hill, New York.

Matejcik, F. J., B. L. Nelson. 1995. Using common random numbers for indifference-zone selection and multiple comparisons in simulation. *Management Sci.* **41**(12) 1935–1945.

McCabe, B. 2003. Monte Carlo simulation for schedule risks. S. Chick, P. Sánchez, D. Ferrin, D. Morrice, eds. *Proc. 2003 Winter Simulation Conf., New Orleans*, 1561–1565.

McGrath, M. F., N. D. Singpurwalla. 1987. A subjective Bayesian approach to the theory of queues II—Inference and information in M/M/1 queues. *Queueing Systems* **1** 335–353.

Merrick, J. R. W. 2007. Getting the right mix of experts. *Decision Anal.* **5**(1) 43–52.

Merrick, J. R. W., J. R. van Dorp. 2006. Speaking the truth in maritime risk assessment. *Risk Anal.* **26**(1) 223–237.

Merrick, J. R. W., R. Soyer, T. Mazzuchi. 2003. A Bayesian semiparametric analysis of the reliability and maintenance of machine tools. *Technometrics* **45**(1) 58–69.

Merrick, J. R. W., R. Soyer, T. Mazzuchi. 2005a. Are maintenance practices for railroad tracks effective? *J. Amer. Statist. Assoc.* **100**(469) 7–25.

Merrick, J. R. W., J. R. van Dorp, V. Dinesh. 2005b. Assessing uncertainty in simulation based maritime risk assessments. *Risk Anal.* **25**(3) 731–743.

Merrick, J. R. W., J. R. van Dorp, J. P. Blackford, G. L. Shaw, J. Harrald, T. Mazzuchi. 2003. A traffic density analysis of proposed ferry service expansion in San Francisco bay using a maritime simulation model. *Reliability Engrg. System Safety* **81**(2) 119–132.

Merrick, J. R. W., J. R. van Dorp, T. Mazzuchi, J. Harrald, J. Spahn, M. Grabowski. 2002. The Prince William Sound risk assessment. *Interfaces* **32**(6) 25–40.

Merrick, J. R. W., J. R. van Dorp, J. Spahn, J. Harrald, T. Mazzuchi, M. Grabowski. 2000. A systems approach to managing oil transportation risk in Prince William Sound. *Systems Engrg.* **3**(3) 128–142.

Morrice, D. J., J. Butler. 2006. Ranking and selection with multiple targets. L. F. Perrone, B. G. Lawson, J. Liu, F. P. Wieland, eds. *Proc. 38th Winter Simulation Conf., Monterey, CA*, 222–230.

Morrice, D. J., J. Butler, P. Mullarkey. 1998. An approach to ranking and selection for multiple performance measures. D. J. Medeiros, E. F. Watson, J. S. Carson, M. S. Manivannan, eds. *Proc. 30th Winter Simulation Conf.*, IEEE Computer Society Press, Los Alamitos, CA, 719–725.

Morrice, D. J., J. Butler, P. Mullarkey, S. Gavirneni. 1999. Sensitivity analysis in ranking and selection for multiple performance measures. P. A. Farrington, H. B. Nembhard, D. T. Sturrock, G. W. Evans, eds. *Proc. 31st Winter Simulation Conf.: Simulation—A Bridge To the Future*, Vol. 1, Association for Computing Machinery, New York, 618–624.

Muñoz, D. F. 2003. A Bayesian framework for modeling demand in supply chain simulation experiments. S. Chick, P. Sánchez, D. Ferrin, D. Morrice, eds. *Proc. 2003 Winter Simulation Conf., New Orleans*, 1319–1325.

National Research Council. 2000. Risk management in the marine transportation system. *Transportation Res. Board Conf. Proc. 22*, National Academy Press, Washington, DC.

Ng, S. H., S. E. Chick. 2001. Reducing input parameter uncertainty for simulations. *Proc. 33rd Winter Simulation Conf.*, IEEE Computer Society, Washington, DC, 364–371.

Ng, S. H., S. E. Chick. 2004. Design of follow-up experiments for improving model discrimination and parameter estimation. *Naval Res. Logist.* **51**(8) 1129–1148.

Ng, S. H., S. E. Chick. 2006. Reducing parameter uncertainty for stochastic systems. *ACM Trans. Modeling Comput. Simulation* **16**(1) 26–51.

Popova, E., D. Morton. 1998. Adaptive stochastic manpower scheduling. D. J. Madeiros, E. F. Watson, J. S. Carson, M. S. Manivannan, eds. *Proc. 1998 Winter Simulation Conf.*, IEEE Computer Society Press, Los Alamitos, CA, 661–668.

Richardson, D., P. J. Green. 1997. On Bayesian analysis of mixtures with an unknown number of components. *J. Royal Statist. Soc.: Ser. B* **59**(4) 473–484.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, A. van der Linde. 2002. Bayesian measures of model complexity and fit. *J. Royal Statist. Soc.: Ser. B* **64**(4) 583–639.

van Beers, W. C., J. P. C. Kleijnen. 2008. Customized sequential designs for random simulation experiments: Kriging metamodeling and bootstrapping. *Eur. J. Oper. Res.* **186**(3) 1099–1113.

van Dorp, J. R., J. R. W. Merrick, J. Harrald, T. Mazzuchi, M. Grabowski. 2001. A risk management procedure for the Washington state Ferries. *Risk Anal.* **21**(1) 127–142.

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

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