

# Bayesian Simulation and Decision Analysis: An Expository Survey

Jason R. W. Merrick

Department of Statistical Sciences and Operations Research

Virginia Commonwealth University

## Abstract

Research into Bayesian analysis of computer simulations has appeared mostly in the simulation literature. Very little of the work has been performed by decision analysts, despite the overlap of the methods used. The hope is that this expository survey on Bayesian simulation will engender 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.

*(Simulation; Bayesian statistics; Decision Analysis)*

## Introduction

From a managerial perspective, the main use of a simulation of a system is to aid in making decisions. Banks et al. (2005) 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, but more recently Bayesian approaches to simulation model construction and inference have been proposed.

The term simulation can mean different things to different people. While many of the methods discussed can be applied in Monte Carlo simulation, Monte Carlo is not our main focus. Simulation techniques have also been applied 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). While such techniques could be referred to as Bayesian simulation, and they are used in the work discussed herein, they are not the simulations of interest in our discussion. Furthermore, some Bayesian work has been performed on modeling parameter uncertainty with deterministic models. In this paper, we discuss stochastic simulation of systems over time and particularly discrete event or system simulation. The models discussed herein include stochastic elements in the model itself, as well as uncertainty in the parameters, and possible model uncertainty. Law and Kelton (2001) and Banks et al. (2005) provide excellent overviews of this type of simulation from the frequentist perspective.

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

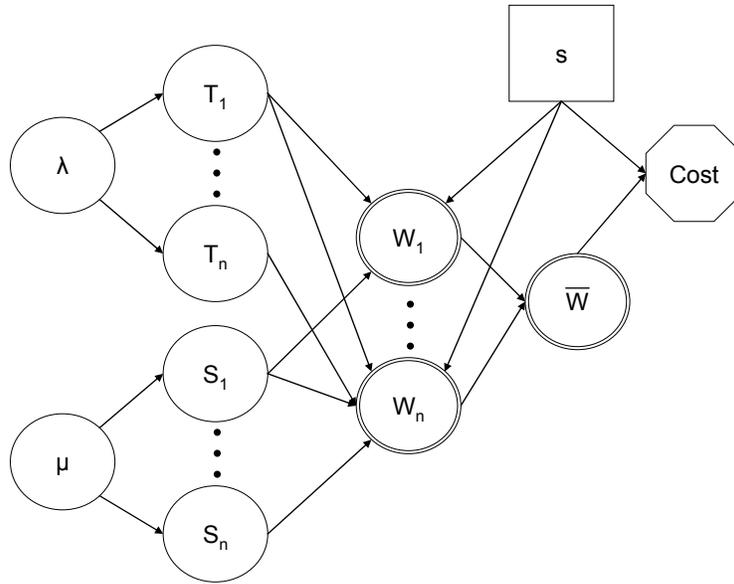
In the next section, we depict a simple Bayesian simulation of a queue using influence diagrams to illustrate the nature of Bayesian simulations from both an input and an output perspective. We then discuss each of the research areas in turn, namely simulation input analysis, output sampling and uncertainty estimation, simulation output analysis and meta-models, using the simulation 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.

Readers with an interest model uncertainty and semiparametric Bayesian analysis should pay particular attention to the Bayesian modeling of simulation inputs and the work that has been performed to handle situations where the distribution of the simulation inputs is not well understood. Readers interested in BMA and MCMC sampling techniques will find applications in the analysis of uncertainty in the outputs and BMA approaches to estimating the sources of such uncertainty, including parameter uncertainty and model uncertainty as well as the uncertainty due to the stochastic nature of the simulation itself. Readers with an interest in Bayesian regression should focus on output analysis using meta-models, both simple and more complex Bayesian response surface approaches. In this section, there is also a discussion of 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 the section on using simulations in the decision process of interest. Readers with an interest in stochastic optimization and algorithms for choosing the best alternative should examine the algorithms for choosing amongst alternative simulation models to choose the best simulated system alternative and particularly the loss function approaches that lead to the best algorithms in empirical testing on benchmark problems. We finish with a discussion of applications of Bayesian simulation, including illustration of one application of Bayesian simulation performed by the author. The final section summarizes the recommendations to decision analysts for further research avenues.

## Influence Diagrams of a simple Bayesian Simulation

Bayesian simulation differs from classical simulation analysis in that probability distributions are used to represent the uncertainty about model parameters, rather than point estimates and confidence intervals. Such treatment is applied 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), while 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. The time between arrivals to a single line,  $T_1, T_2, \dots$ , are assumed to be exponentially distributed with parameter  $\lambda$ .  $s$  servers take entities from the queue on a first-come-first-served basis with the system empty at time 0. Service times,  $S_1, S_2, \dots$ , follow an exponentially distribution with parameter  $\mu$  and entities then leave the system. Neither  $\lambda$  or  $\mu$  are known, so we have a prior distribution on each,  $\pi(\lambda)$  and  $\pi(\mu)$ , and the decision maker has observed some input data for each to obtain posterior distributions on  $\lambda$  and  $\mu$ ,  $\pi(\lambda | t_1, \dots, t_p)$  and  $\pi(\mu | s_1, \dots, s_q)$ .

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



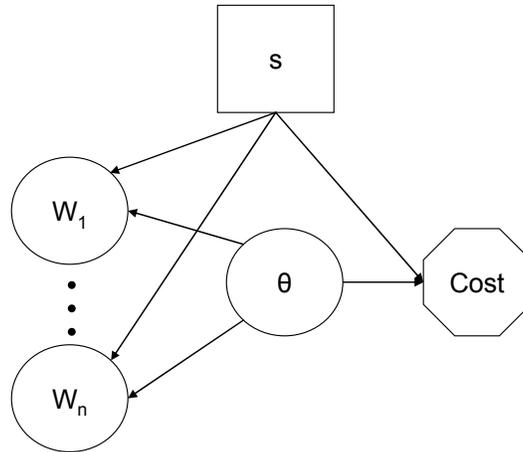
**Figure 1. An influence diagram of a decision about the number of servers in a M/M/s queue including the inputs to the simulation.**

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  $S_1, \dots, S_n$  and  $T_1, \dots, T_n$ . 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  $\lambda$  and  $\mu$ . There is also a possible model uncertainty, usually considered to be uncertainty about the distribution of  $S_1, \dots, S_n$  and  $T_1, \dots, T_n$ , but also including uncertainty about the validity of the calculations used to obtain  $W_1, \dots, W_n$  and  $\bar{W}$  from  $S_1, \dots, S_n$  and  $T_1, \dots, T_n$ . There are two cases of interest here. If we are interested in a fixed  $n$  or if we wish to simulate for a fixed time period (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  $\bar{W} \rightarrow \theta$  as  $n \rightarrow \infty$  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,  $\bar{W} | \lambda, \mu$  (or  $\theta | \lambda, \mu$  in the steady state case) has a distribution that is defined by the distributions of  $S_1, \dots, S_n | \mu$  and  $T_1, \dots, T_n | \lambda$  and  $W_1, \dots, W_n$  given  $S_1, \dots, S_n$  and  $T_1, \dots, T_n$  through the law of total probability, namely

$$\int \dots \int \int \dots \int (\bar{W} | S_1, \dots, S_n, T_1, \dots, T_n)(T_1, \dots, T_n | \lambda)(S_1, \dots, S_n | \mu) dT_n \dots dT_1 dS_n \dots dS_1 \quad (1)$$

This expression may then be integrated over  $\pi(\lambda | t_1, \dots, t_p)$  and  $\pi(\mu | s_1, \dots, s_q)$ . However, this calculation cannot be performed 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  $S_1, \dots, S_n$ ,  $T_1, \dots, T_n$ ,  $\lambda$ , and  $\mu$  have been integrated into the distributions of  $W_1, \dots, W_n$ , thus we are only including the outputs of the simulation. Bayesian inference is made difficult without a known probability model to define a likelihood, thus one may assume a distribution on  $\bar{W} | \lambda, \mu$  (or  $\theta | \lambda, \mu$ ), called a meta-model. Prior distributions are then defined on the parameters of this distribution and the priors are updated given samples from the simulation. This is akin to a meta-model from classical simulation analysis (Chick 1997a; Law and Kelton 2001).



**Figure 2. An influence diagram of a decision about the number of servers in a M/M/s queue considering only outputs from the simulation.**

We will discuss different forms of meta-models later, but for now it is critical to realize that the marginal distribution of  $W_1, \dots, W_n$  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.

### Bayesian Input Analysis

In Bayesian simulation, input uncertainty is incorporated 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 inter-arrival times. Historical inter-arrival times are calculated from data. Let  $T_1^k, \dots, T_{m^k}^k$  be the  $m^k$  independent inter-arrival times for the  $k$ -th random input process ( $k = 1, \dots, n$ ). In our example of the previous section,  $k = 1$  could represent the arrival time inputs and  $k = 2$  could represent the service time inputs.

In a classical simulation approach, the probability model is usually chosen by determining 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 Andersen-Darling, Chi-square or Kolmogorov-Smirnov statistics (Law and Kelton 2001). Suppose

$F_1^k(t | \Theta_1^k), \dots, F_p^k(t | \Theta_p^k)$  are  $p$  families of probability distribution, such as the exponential, Weibull, gamma or log-normal distributions. The superscript  $k$  is included throughout as each arrival process can be modeled by a different probability distribution and will certainly have different parameter values.

$\Theta_1^k, \dots, \Theta_p^k$  are the parameters of the  $p$  potential probability distributions for the  $k$ -th input process.

Best estimates of each set of parameters,  $\hat{\Theta}_j^k$ , are obtained from the data  $D^k = \{T_1^k = t_1^k, \dots, T_{m^k}^k = t_{m^k}^k\}$ ,

using maximum likelihood, method of moments or other estimation procedures. The best fit distribution is then chosen by taking either the fitted distribution with the lowest appropriate fit statistic or at least a fitted distribution that is not rejected by the corresponding hypothesis test and that has desirable

properties, such as simple manipulation of the mean or variance. Thus aleatory uncertainty is modeled by the renewal process, but as only best estimates are used for the parameters of the generating probability distribution, the uncertainty about their true values is not included in the model.

Under the Bayesian paradigm, prior distributions are specified for the parameters of the postulated distributions, denoted by  $\pi_1^k(\Theta_1^k), \dots, \pi_p^k(\Theta_p^k)$ , and the data is used to update these priors using the standard Bayesian machinery to obtain posterior distributions denoted by

$$\pi_1^k(\Theta_1^k | D^k), \dots, \pi_p^k(\Theta_p^k | D^k).$$

One approach here is to use a model selection criterion to choose between potential probability distributions given input data, such as the DIC (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 (2005a). 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 and each distribution is then conditioned on that distribution being the correct one. Prior probabilities are then assigned on the probability of each distribution being correct for the  $k$ -th input process,  $\pi^k(m)$  for  $m = 1, \dots, p$ , as well as prior distributions on each set of parameters given that distribution is correct,  $\pi_m^k(\Theta_m^k | m)$  for  $m = 1, \dots, p$ . The probability that a given model  $m$  is correct given input data can then be found using

$$\pi(m | D^k) = \frac{P(D^k | m)\pi(m)}{\sum_{l=1}^p P(D^k | l)\pi(l)}. \quad (1)$$

The expression  $P(D^k | m)$  can be found using  $\int F_m(D^k | \theta_m^k)\pi(\theta_m^k)d\theta_m^k$  and is the same term used 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.

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 one parametric family of distributions. 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). Thus rather than assuming that the input distribution must take a simple, parametric form, but that which one is unknown, Cheng and Currie assume that form is more complex, but can be represented by a mixture of a finite number of one parametric distribution family. Such a finite mixture covers a wider space of possible distributions than any one single parametric distribution.

This work depicts two approaches to model 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 priors are placed on its parameters. There are other approaches that have been used by decision analysts and Bayesian statisticians when there is uncertainty about the probability model. Bayesian nonparametric approaches include Dirichlet and Gaussian processes, where the prior distribution is assumed 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.

### **Analyzing Sources of Output Uncertainty**

The problem of sampling from input models to obtain correct output samples is first discussed in Chick (1997a). Input uncertainty is propagated to the outputs through multiple replications of the simulation

using the algorithm in Figure 3. In our example of the M/M/s simulation, a value for  $\lambda$  and  $\mu$  are sampled for each replication and then values for  $S_1, \dots, S_n$  and  $T_1, \dots, T_n$  are sampled within each replication given the values of  $\lambda$  and  $\mu$  sampled for that replication. In this manner, aleatory uncertainty, or the randomness of the system itself, is represented within each replication, while epistemic uncertainty, or uncertainty about the parameter values, is represented across multiple replications. This sampling approach becomes obvious when examining Figure 1.  $\lambda$  is relevant to the distributions of all  $S_1, \dots, S_n$ , so one sampled value of  $\lambda$  must be used for sampling one set of values for  $S_1, \dots, S_n$ . The same is true for  $\mu$  and  $T_1, \dots, T_n$ . These values of  $S_1, \dots, S_n$  and  $T_1, \dots, T_n$  are then used to obtain one replication of  $W_1, \dots, W_n$  and  $\theta$ . This is more complex than frequentist simulation, where one estimate of  $\lambda$  and  $\mu$  are used for all replicates of the simulation, meaning that parameter 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 model uncertainty framework from Chick (1999) and provides a general sampling algorithm for sampling from the Chick (1999) framework and discusses standard methods for determining prior distributions, namely non-informative priors, data-driven priors, and moment-matching. While 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 in to 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).

For  $r=1, \dots, n$  replications of the simulation:

1. Sample values from the posterior distributions for the parameters,  $\pi^k(\Theta^k | D^k)$ , to be used in the  $r$ -th replication for each input process  $k$ .
2. For the  $r$ -th replication of the simulation:
  - a. Sample random variates for each simulation input given the parameter values sampled in Step 2;
  - b. Generate the simulation output  $y_r$  for these random variates,.

End loop

**Figure 3. An algorithm from Chick (1997a) for sampling from outputs of Bayesian simulations.**

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. The posterior probability of each input model given input data can be calculated in closed form for exponential family distributions given conjugate priors and approximated otherwise. Components of the output variance are then calculated by considering variations for a given model and variations between models. The output variance due to model uncertainty is estimated by

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

where  $\bar{y}_m = \frac{1}{nt} \sum_{r=1}^n \sum_{s=1}^t y_{rms}$  and  $\pi(m | D)$  is the posterior probability of model  $m$  given the input data.

The output variance due to the stochastic nature of the simulation is estimated by

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

and the output variance due to parameter uncertainty is estimated by

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

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

For  $r = 1, \dots, n$  iterations:

1. For each model  $m = 1, \dots, p$ .

a. Sample values from the posterior distributions for the parameters for the  $m$ -th model,

$\pi_m^k(\Theta_m^k | m, D^k)$ , to be used in the  $r$ -th replication for each input process  $k$ .

b. For  $s = 1, \dots, t$  iterations:

I. Sample random variates for each simulation input from the model  $m$  given the parameter values sampled in Step 1a;

II. Generate the simulation output  $y_{rms}$  for these random variates.

End loop

End loop

End loop

**Figure 4. An algorithm from Zouaoui And Wilson (2004) for estimating stochastic, model, and parameter variance of simulation outputs.**

This area of work appears quite complete for the model uncertainty approaches discussed in the previous section. However, if developments are made using Bayesian nonparametric or semiparametric approaches, then these algorithms will need to be updated to allow for these more complex input set-ups.

### **Bayesian Output Analysis**

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. This can be exploited 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 this can be thought of as a Bayesian version of meta-modeling (Law and Kelton 2001). Chick (1997a) provides a general framework for such a meta-model, including Bayesian expressions for calculating the posterior distribution of the parameters of the output meta-model 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  $r$ -th replication ( $r = 1, \dots, s$  for  $s$  replications). As our example output data is in the form of a count, one model could be a Poisson distribution with rate  $\mu$ , with a conjugate gamma distributed prior on  $\mu$ . The predictive distribution of the number of entities passing through our simulated process is then a Poisson-gamma distribution in the sense of Bernardo and Smith (2000).

Cheng (1998) fits a Bayesian regression model to simulation output to form a meta-model to be used in prediction and decision-making. The independent variables in the regression model can either be input model parameters, such as  $\lambda$  and  $\mu$  in our M/M/s example, or decision or control variables, such as  $s$  in our M/M/s example. Cheng allows for an unknown number of non-linear components in the

regression and assumes that the error terms are normally distributed and independent. Thus we can write the meta-model as

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

where  $Y_j$  is the simulation output for the  $j$ -th replication,  $\theta$  is a vector of the control variables for the simulation (either input model parameters or decision variables),  $f_i(\theta)$  is the  $i$ -th regression term,  $\beta_i$  is the regression parameter for the  $i$ -th regression term and  $z_j$  is the normally distributed error term.  $Q$  is the number of terms in the model and is assumed to be uncertain in this work, but can also be fixed. A Metropolis Hastings algorithm is used for inference with the assumption that any regression parameter taking a value within a specified value  $\delta$  of zero represents a regression component that need not be included; this allows the number of components to be effectively estimated by the derived parameter sample. This is called a derived chain MCMC method as the number of components is not handled by sampling from its posterior distribution. Cheng (1999) extends this approach to allow for non-normal 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 propose a regression meta-model where the simulation output is normally distributed with a mean that is a function of the input distribution parameters. The predictive distribution of the simulation output is then found by integrating the regression meta-model 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 the approach to cases where there is sufficient input data. Similarly the simulation output is asymptotically normal with a mean that can be approximated 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 replicates

performed thus far and written as a function of the number of additional input data points that are to be collected, denoted  $V_{par}(m_1, \dots, m_K)$  where  $m_k$  is the number of additional input data points to be collected for the  $k$ -th input process. Again, though, the asymptotic approximation limits use to cases here a sufficiently large number of initial simulation replications have been performed prior these calculations. The optimization problem is then defined as determining the optimal number of additional data points for each input distribution to minimize the output variance subject to a data collection budget constraint,

$$\begin{aligned} \min \quad & V_{par}(m_1, \dots, m_K) \\ \text{s.t.} \quad & m_k \geq 0 \\ & \sum_{k=1}^K c_k m_k \leq B \end{aligned}$$

where  $c_k$  is the cost of each additional input data point for the  $k$ -th input process and  $B$  is the total budget available for additional input data collection. Ng and Chick (2006) extend this approach to non-normal distributions from the exponential family using Bayesian Model Averaging.

While Ng and Chick (2001 2006) address collecting additional input data to reduce output uncertainty, 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. The aim of the allocation is both differentiation between the simulation models that result from different parameter or control variable values and to enhance parameter estimation for the meta-model. The meta-model 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 (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 design of

experiments; this utility function is often denoted by  $MD$ . Another version averages the entropy across all possible model, a utility function that leads to good parameter estimation; this utility function is often denoted  $SP$ . The utility function proposed by Chick and Ng can then be written as

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

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. Each attribute is scaled 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 provide a closed form approximation of the expected utility for normal errors which can be numerically maximized to find the optimal design.

The introduction of multi-attribute 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 multi-attribute 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 meta-model. There has been significant frequentist work in this area and work for deterministic models using a technique called Kriging (see van Beers and Kleijnen 2004). This is a promising area for future research.

There are several additional areas that have not been addressed thus far. Andradóttir and Bier (2000) briefly discuss the issue of model validity from the Bayesian perspective. Suppose  $Y$  is a simulation output for which we have posterior predictive distribution  $\pi(Y | C)$  after observing some number of simulation replications in  $C$  and suppose  $Z$  is the same output for the real system that has been simulated and  $\pi(Z | D)$  is its posterior predictive distribution after observing real world data  $D$ . What does it mean if  $\pi(Y | C)$  and  $\pi(Z | D)$  do not agree? 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 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 flow of information. One possibility is the use of meta-models, 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 which provide additional information to update the decision maker's knowledge about  $Z$ . The work in aggregation of expert forecasts allows such updating (see Clemen and Winkler (1999) for a survey of such approaches) 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 could be made by decision analysts.

Lastly, we point out that just as input model uncertainty can be approached using Bayesian nonparametric and semiparametric approaches, so can output model uncertainty. Dirichlet process, gamma process, and Dirichlet process mixtures, for example, are also suitable output meta-models, 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 2005b) discuss approaches for nonparametric and semiparametric Bayesian regression.

## **Making Decisions with Bayesian Simulations**

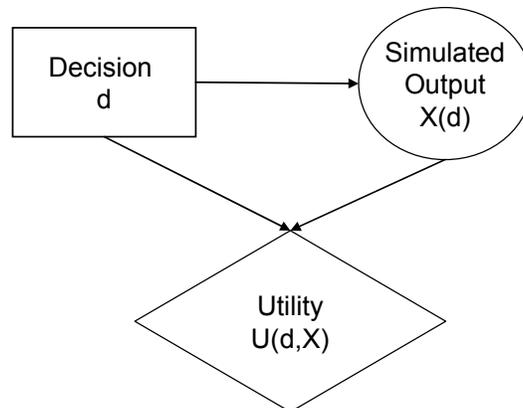
Thus far we have discussed the construction of a Bayesian simulation and several methods for making decisions *about* a simulation, specifically what input data to collect and what simulation replications to perform to meet various objectives. In this section, we will discuss making decisions *with* simulations. In recent years, considerable work has been performed in selecting the best simulated system. Given a set of

alternatives, simulation runs are performed for each alternative and statistical procedures have been developed to assess whether there is sufficient evidence in favor of any one alternative (Matejck and Nelson 1995; Bechhofer et al. 1995; Goldsman and Nelson 1998; Chick and Inoue 2001a 2001b). 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. One concentrates on using Bayesian machinery to develop algorithms that are often better than the frequentist algorithms at finding the simulated system with the highest mean for a given output. The other concentrates on finding the simulated system with the highest expected utility, for both single attribute and multi-attribute problems. In this section, we discuss the latter truly decision analytic work. In the next section, we discuss the former Bayesian versions of frequentist simulated system selection algorithms.

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

$$\max_{d \in \Delta} \left[ V(d) = \int u(d, X) p_d(X) dX \right] \quad (2)$$

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.



**Figure 5. An influence diagram for selecting the best simulated system.**

In a more general decision analytic setting, conditional distribution  $p_d(X)$  has a known functional form. However, in the case of simulation decision problems,  $p_d(X)$  cannot be pointwise calculated, but can be sampled from by running an iteration of the simulation. Thus,  $V(d)$  cannot be calculated 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 (2008) 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 minimize the cost of the 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 being simulated. In this manner, work to improve the validity of the simulation can be considered by its impact on the eventual decision, or its value of information. While the decision maker is assumed to be risk-neutral in Chick and Gans (2008), such a purely economic analysis would be appealing to a decision maker if all costs can be accurately estimated.

Chick (1997b) provides the first formulation of selecting the best simulated system by maximizing expected utility. Chick 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 assumes a normal meta-model for the simulation output for independent replications, then a multivariate normal meta-model 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 compares the Bayesian dominance probability to the frequentist p-value under the assumption of equal means.

Often simulations have multiple outputs and the decisions involved have multiple attributes that are represented by these outputs. This situation suggests the use of multi-attribute utility functions to represent such trade-offs and the associated attitude to multi-attribute risk. Morrice, Butler, Mullarkey (1998) first propose such an approach, with later extensions and elaborations in Morrice, Butler, Mullarkey, Gavirneni (1999) and Butler, Morrice, Mullarkey (2001). This work extends the approach of Chick (1997b) to consider multi-attribute 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, Butler, Mullarkey (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 set-up using Bayesian simulation and maximizing expected utility, one may obtain samples from  $x_1, \dots, x_n$  from  $p_d(X)$  by running the simulation for decision  $d$ , calculate

$$u(d, x_1), \dots, u(d, x_n), \text{ and estimate } V(d) \text{ using } \bar{u}(d) = \frac{1}{n} \sum_{i=1}^n u(d, x_i).$$

Andradóttir and Kim (2007) criticize Butler, Morrice, Mullarkey (2001) because it 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 believe it is difficult to find an appropriate multi-attribute 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 multi-attribute utility functions in all this work does not require the use of the ranking and selection procedure from Butler, Morrice, Mullarkey (2001); any ranking and selection procedures can be used, including the decision analytic methods discussed in the next section.

Again there appear to be several directions that decision analysts could continue this research. The simulation is just one part of an actual decision that must be made. The work of Chick and Gans

shows that decisions about the simulation are just the same as decisions about obtaining other sources of information. This suggests approaches using expected value of sample (simulation) information. As we discussed at the end of the previous section, this also suggests that simulations could be considered to be experts that provide samples rather than full distributions. The work by Morrice and Butler in multi-attribute methods may also lead to further avenues of research for decision analysts.

### **Ranking and Selecting the Simulated System with the Highest Mean Output**

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. Which is best?

All approaches discussed in this section are based on the same basic set-up. Suppose we have to decide between  $K$  simulated systems. Let  $X_{i,j}$  be the  $j$ -th replicate of the output of interest for simulated system  $i$ , for  $i = 1, \dots, K$  and  $j = 1, 2, \dots$ . We assume that  $X_{i,j}$  is normally distributed with mean  $w_i$  and precision  $\lambda_i$ . We denote the ordering of the means from lowest to highest by

$w_{[1]} < \dots < w_{[K]}$ . 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  $\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{i,j}$ . We denote the current ordering of

the sample means by  $\bar{x}_{(1)} < \dots < \bar{x}_{(K)}$ . Thus  $(i)$  denotes the  $i$ -th largest sample mean and  $[i]$  denotes the

$i$ -th largest true mean. At the end of any given algorithm, we have selected the best simulated system if

the system selected, denoted  $D = (K)$ , has the highest true mean,  $D = [K]$ . Let

$C = \{X_{i,j}, i = 1, \dots, K, j = 1, \dots, n_i\}$  be the set of all simulation replicates performed thus far.

Inoue and Chick (1998) define a Bayesian probability of correct selection, defined as

$P(w_{(i)} < w_{(K)}, \forall i \neq K | C)$  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. The lower bound is defined as

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

or the probability that the true mean for the system currently chosen as the best so far does have a higher mean than all other systems, given all simulation replicates observed thus far. Chen, Chen, Dai (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 and 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, Dai, Chen, Yucesan (1997) improves on Chen, Chen, Dai (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,

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

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

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

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

$$E_{w|C}[L_{OC}(D = (K)) | C] = E_{w|C}[w_{[k]} - w_{(K)} | C],$$

the expected difference between the best system and system  $i$ .

Chick and Inoue develop approximations for estimating the expected loss in each of these cases given each possible number of additional replications of each possible system. Replications can then be allocated 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 two versions of the algorithms for the 0-1 loss function are given, (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. A Bonferoni type approximation of the Bayesian probability of correct selection is used for this dependent case. Again, Chick and Inoue develop versions of their algorithms with both linear loss and 0-1 loss functions.

Branke, Chick, and Schmidt (2007) compare a frequentist procedure called  $RN++$  (Goldman 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 only allocates one additional replication. Each procedure is implemented 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  $\hat{\delta}^*$  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 included:

- a slippage configuration: where all systems have equal means except the best which is better by a parameter  $\hat{\delta}^*$  divided by the noise;
- monotone decreasing means: where each successive mean has a linearly decreasing mean; and
- 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, the 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 increased. Branke, Chick, and Schmidt (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 multi-attribute work of Butler and Morrice and the economic work of Chick and Gans have taken steps in the direction of decision analysis, but considerable further work is possible. As the simulation decision problem can be depicted as an influence diagram, algorithms for solving influence diagrams may be applicable, such as augmented probability sampling (Bielza et al. 1998). Extensions to the multi-attribute 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 multi-attribute 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 in to continuous decision variables.

## **Applications**

There have been a number of applications of Bayesian simulation methods in recent years. Popova and Morton (1998) used stochastic optimization for manpower 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 (2001a). McCabe (2003) modeled schedule risk with a Bayesian belief network version of a PERT network and used Monte Carlo simulation to obtain posterior schedule completion probabilities. Chick, Soorapanth, and Koopman (2003) modeled the spread of water-borne infectious diseases using Bayesian inference from simulations, expert judgment, and field data. The applications of the techniques discussed herein are widespread and varied. To provide a more detailed application example, we will discuss the author's use of Bayesian methods to assess uncertainty in simulation-based maritime risk assessment. A full description of this work can be found in Merrick et al (2005a) and Merrick and van Dorp (2006); 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 leads 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 (PWS) Risk Assessment (Merrick et al., 2000, 2002), Washington State Ferries (WSF) 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. Their results have been used in major investment decisions and have played a significant role in

the management of maritime transportation in the US. 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 condition, mean that risk levels change over time. This requires the use of simulation to accurately model the impact of changes that affect the traffic patterns, such as introducing new traffic rules and increases or decreases in the volume of traffic in a given port. However, for the decision maker to make a complete decision, they must understand not only risk estimates of the alternative simulated systems, but also the stochastic and parameter uncertainty in such estimates.

The simulation is used to count the frequency of situations in which the ferries could be involved in an accident. These situations are defined by 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

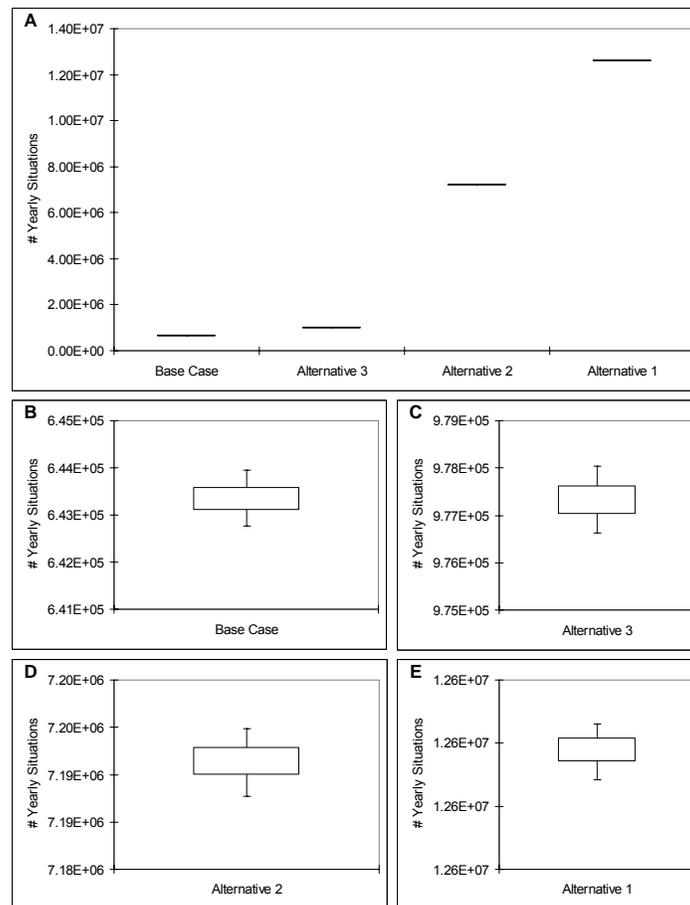
$$P(\text{Collision}) = \sum_{j=1}^k P(\text{Collision} | \text{Situation}_j) P(\text{Situation}_j) \quad (2)$$

where  $\text{Situation}_j$  denotes the  $j$ -th of all possible situations for  $j = 1, \dots, k$  and  $k$  is the total number of possible combinations. The expected yearly frequency of collisions can then be calculated by multiplying the probability in (2) by the number of simulated time periods in a year. A simulation of the maritime system is used to estimate  $P(\text{Situation}_j)$ , while  $P(\text{Collision} | \text{Situation}_j)$  is estimated using a combination of expert judgment and accident data (Merrick et al. 2006).

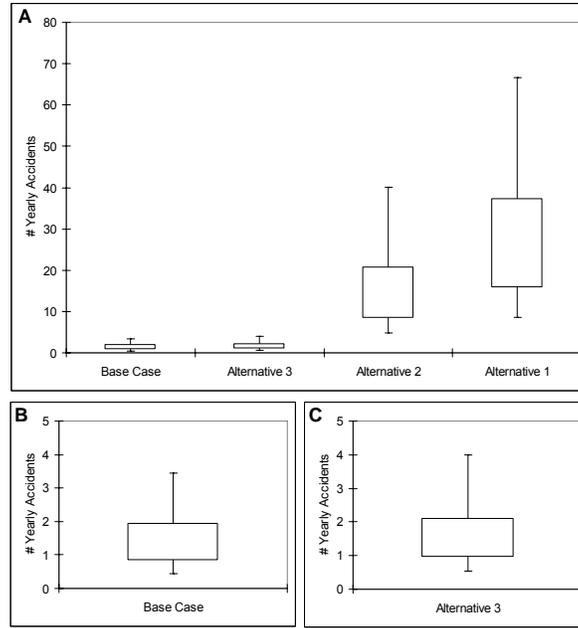
In this risk assessment methodology, the data obtained from the simulation in each replication will be a count of the number of vessel interactions occurring in each replication of the simulation. Following the Bayesian approach, Merrick et al. (2005a) hypothesize a Poisson meta-model for the random output data and specify our prior beliefs about the parameters of this output model using the conjugate gamma distribution. The predictive distribution of  $P(\text{Situation}_j)$  is then a Poisson-gamma distribution in the sense of Bernardo 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. (2005a). Figure 6A 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 6A are actually box plots showing the predictive distribution with the interquartile range as the box and the 5<sup>th</sup> and 95<sup>th</sup> percentiles of the distribution as the whiskers. However, as the remaining uncertainty in these estimates is low, they do not show up on this comparison scale and are repeated Figures 6B through 6E. 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.



**Figure 6. Expected Yearly Situations Comparison.**



**Figure 7. Expected Yearly Collisions Comparison.**

As there is a wide variability and significant uncertainty about  $P(\text{Collision} | \text{Situation}_j)$ , the analysis from Merrick et al. (2005a) is useful, but not definitive. Instead we must examine the collision probability itself,  $P(\text{Collision})$ . Figure 7A 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 this uncertainty cannot be removed 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. As the comparison is not clear on a scale that includes Alternatives 2 and 1, Figures 7B and 7C 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.

## **6. Conclusions and Areas for Research by Decision Analysts**

We have reviewed the literature on Bayesian simulation and covered many types of decisions that can be addressed using Bayesian methods. These decisions have included what input data to collect to minimize output uncertainty, what simulation replications to run to differentiate between alternative systems, what simulated system has the highest mean output, and what simulated system has the highest expected utility, including both single-attribute and multi-attribute 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 work on entropy utility functions and alternative multi-attribute 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, and 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 multi-attribute simulation decision making and in treating decisions about the simulation as part of the overall real-world decision that the simulation has been built to help with. In choosing the best simulated system, algorithms used in choosing the best alternative for influence diagrams may prove fruitful and specific algorithms for multi-attribute simulation problems can be developed.

The aim in writing this review was to introduce the area of Bayesian simulation to the decision analysis community in the hope that more decision analytic tools and approaches can be applied in the area. We hope that enough background has been provided to give a general understanding of the area and

that enough open problems have been identified to wet the appetites of decision analysts to work in this area. Readers will, no doubt, have identified opportunities beyond those mentioned. We look forward to seeing work on Bayesian simulation in the decision analysis literature.

## References

- Abbas, A. E. 2004. Entropy methods for adaptive utility elicitation. *IEEE Transactions on Systems, Man, and Cybernetics, Part A* **34**(2) 169-178.
- Abbas, A. E. R. A. Howard. 2005. Attribute dominance utility. *Decision Analysis* **2**(4) 185-206.
- Andradóttir, S., V. M. Bier. 2000. Applying Bayesian ideas in simulation. *Simulation Theory and Practice* **8** 253–280.
- Andradóttir, S., S. H. Kim. 2007. Fully sequential procedures for comparing constrained systems via simulation. Submitted.
- 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*, New York: John Wiley & Sons.
- Bernardo, J., A. Smith. 2000. *Bayesian Theory*. Second edition. Chichester, UK: Wiley.
- Bielza, C., P. Müller, D Rios Insua. 1999. Decision analysis by augmented probability simulation. *Management Science* **45**(7) 995-1007.
- Branke, J., S.E. Chick, C. Schmidt. 2007. Selecting a Selection Procedure. *Management Science* **53**(12) 1916–1932.
- Butler, J., D. J. Morrice, P. W. Mullarkey. 2001. A multiple attribute utility theory approach to ranking and selection. *Management Science* **47**(6) 800-816.
- 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. In *Proceedings of the 28th Conference on Winter Simulation*, J. M. Charnes, D. J. Morrice, D. T. Brunner, and J. J. Swain, eds. 398-405.
- Chen, C.-H., H.-C. Chen, L. Dai, E. Yucesan. 1997. New development of optimal computing budget allocation for discrete event simulation. In *Proceedings of the 29th Conference on Winter Simulation*, S. Andradóttir, K. J. Healy, D. H. Withers, and B. L. Nelson, eds., 334-341.
- Cheng, R. 1998. Bayesian model selection when the number of components is unknown. In *Proceedings of the 30th Conference on Winter Simulation*, D. J. Medeiros, E. F. Watson, J. S. Carson, and M. S. Manivannan, eds. 653-660.
- Cheng, R. 1999. Regression metamodeling in simulation using Bayesian methods. In *Proceedings of the 31st Conference on Winter Simulation: Simulation---A Bridge To the Future - Volume 1 P*. Farrington, H. Nembhard, D. Sturrock and G. Evans, eds., 330-335.
- Cheng, R. C. and Currie, C. S. 2003. Simulation input modeling: prior and candidate models in the Bayesian analysis of finite mixtures. In *Proceedings of the 35th Conference on Winter Simulation: Driving innovation*, 392-398.
- Chick, S. 1997a. Bayesian analysis for simulation input and output. In *Proceedings of the 29th Conference on Winter Simulation*, S. Andradóttir, K. J. Healy, D. H. Withers, and B. L. Nelson, eds., 253-260.
- Chick, S. 1997b. Selecting the Best System: A Decision Theoretic Approach. In *Proceedings of the 29th Conference on Winter Simulation*, S. Andradóttir, K. J. Healy, D. H. Withers, and B. L. Nelson, eds., 326-333.
- Chick, S. 1999. Steps to implement Bayesian input distribution selection. In *Proceedings of the 31st Conference on Winter Simulation: Simulation---A Bridge To the Future - Volume 1 P*. Farrington, H. Nembhard, D. Sturrock and G. Evans, eds., 317-324.
- Chick, S. 2000. Bayesian methods for simulation. In *Proceedings of the 2000 Winter Simulation Conference*, J. A. Joines, R. R. Barton, K. Kang, P. A. Fishwick, eds., 109-118.

- Chick, S. 2001. Input distribution selection for simulation experiments: accounting for input uncertainty. *Operations Research* **49**(5) 744-758.
- Chick, S. 2004. Bayesian methods for discrete event simulation. In *Proceedings of the 2004 Conference on Winter Simulation*, R. G. Ingalls, M. D. Rossetti, J. S. Smith, B. A. Peters, eds., 89-100.
- Chick, S. 2006a. Subjective probability and Bayesian methodology. In *Simulation*, S. G. Henderson, B. L. Nelson, eds. New York: Elsevier.
- Chick, S. 2006b. Bayesian ideas and discrete event simulation: why, what and how. In *Proceedings of the 38th Conference on Winter Simulation*, L. F. Perrone, B. G. Lawson, J. Liu, and F. P. Wieland, eds., 96-105.
- Chick, S. E., N. Gans. 2008. Economic analysis of simulation selection problems. *Management Science*, to appear.
- Chick, S. E., K. Inoue. 2001a. New procedures to select the best simulated system using common random numbers. *Management Science* **47**(8) 1133-1149
- Chick, S. E., K. Inoue. 2001b. New two-stage and sequential procedures for selecting the best simulated system. *Operations Research* **49**(5) 732-743.
- Chick, S. E., S. Soorapanth, J. S. Koopman. 2003. Inferring infection transmission parameters that influence water treatment decisions. *Management Science* **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*. New York: Springer.
- Draper, D. 1995. Assessment and propagation of model uncertainty (with discussion). *J. of the Royal Statist. Soc., Series B.* **57**(1) 45-97.
- Escobar, D. M., M. West. 1995. Bayesian density estimation and inference using mixtures. *J. Amer. Stat. Assoc.* **90** 577-588.

- Gamerman, D., H. F. Lopes. 2006. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Second Edition. Chapman Hall, Florida.
- Gelfand, A. 1996. Model determination using sampling based methods. In *Markov Chain Monte Carlo in Practice*, W. Gilks, S. Richardson, D. J. Spiegelhalter, eds., 145–161, London: Chapman and Hall.
- Gilks, W., Richardson, S., Spiegelhalter, D. (1996). *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall.
- 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.
- Goldsman, D. B.L. Nelson. 1998. Comparing systems via simulation. In *Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice*, ed. J. Banks. New York: John Wiley & Sons.
- Herzog, T. N. 2002. Bayesian statistics and the Monte Carlo method. In *Proceedings of the 2002 Winter Simulation Conference*, E. Yücesan, C.-H. Chen, J. Snowdon, and J. Charnes, eds., 136–146.
- Inoue, K., S. E. Chick. 1998. Comparison of Bayesian and frequentist assessments of uncertainty for selecting the best system. In *Proceedings of the 30th Conference on Winter Simulation*, D. J. Medeiros, E. F. Watson, J. S. Carson, and M. S. Manivannan, eds., 727-734.
- Kass, R., A. Raftery. 1995. Bayes factors. *Journal of the American Statistical Association* **90**(430) 773–795.
- Keeney, R.L., H. Raiffa. 1993. *Decisions with multiple objectives-preferences and value tradeoffs*. Cambridge & New York: Cambridge University Press.
- Kuo, L. 1986. Computation of mixtures of Dirichlet processes. *SIAM J. Sci. and Stat. Comp.* **7** 60–71.
- Law, A., W. Kelton. 2001. *Simulation Modeling and Analysis*, 3rd ed. New York: McGraw Hill.
- Matejckik, F. J., B. L. Nelson. 1995. Using common random numbers for indifference-zone selection and multiple comparisons in simulation. *Management Science* **41**(12) 1935-1945.

- McCabe, B. 2003. Monte Carlo simulation for schedule risks. In *Proceedings of the 2003 Winter Simulation Conference*, S. Chick, P. Sánchez, D. Ferrin, D. Morrice, eds., 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., J. R. van Dorp. 2006. Speaking the truth in maritime risk assessment. *Risk Analysis* **26**(1) 223-237.
- Merrick, J. R. W. 2007. Getting the right mix of experts. *Decision Analysis* **5**(1) 43-52.
- 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 Engineering and 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 Engineering* **3**(3) 128-142.
- Merrick, J. R. W., J. R. van Dorp, V. Dinesh. 2005a. Assessing uncertainty in simulation based maritime risk assessments. *Risk Analysis* **25**(3) 731-743.
- Merrick, J. R. W., R. Soyer, T. Mazzuchi. 2003. A Bayesian semi-parametric analysis of the reliability and maintenance of machine tools. *Technometrics* **45**(1) 58-69.
- Merrick, J. R. W., R. Soyer, T. Mazzuchi. 2005b. Are maintenance practices for railroad tracks effective? *Journal of the American Statistical Association* **100**(469) 7-25.
- Morrice, D. J., J. Butler, P. Mullarkey. 1998. An approach to ranking and selection for multiple performance measures. In *Proceedings of the 30th Conference on Winter Simulation*, D. J. Medeiros, E. F. Watson, J. S. Carson, and M. S. Manivannan, eds., 719-725.
- Morrice, D. J., J. Butler, P. Mullarkey, S. Gavirneni. 1999. Sensitivity analysis in ranking and selection for multiple performance measures. In *Proceedings of the 31st Conference on Winter Simulation:*

- Simulation---A Bridge To the Future - Volume 1*, P. A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. W. Evans, eds., 618-624.
- Morrice, D. J., J. Butler. 2006. Ranking and selection with multiple targets. In *Proceedings of the 38th Conference on Winter Simulation*, L. F. Perrone, B. G. Lawson, J. Liu, and F. P. Wieland, eds., 222-230.
- Muñoz, D. F. 2003. A Bayesian framework for modeling demand in supply chain simulation experiments. In *Proceedings of the 2003 Winter Simulation Conference*, S. Chick, P. Sánchez, D. Ferrin, D. Morrice, eds., 1319–1325.
- National Research Council. 2000. Risk Management in the Marine Transportation System. *Conference Proceedings 22, National Academy Press, Washington, DC*.
- Ng, S. H., S. E. Chick. 2001. Reducing input parameter uncertainty for simulations. In *Proceedings of the 33rd Conference on Winter Simulation*, 364-371.
- Ng, S. H., S. E. Chick. 2004. Design of follow-up experiments for improving model discrimination and parameter estimation. *Naval Research Logistics* **51**(8) 1129--1148.
- Ng, S. H., S. E. Chick. 2006. Reducing parameter uncertainty for stochastic systems, *ACM Transactions on Modeling and Computer Simulation* 16(1) 26-51
- Popova, E., D. Morton. 1998. Adaptive stochastic manpower scheduling. In *Proceedings of the 1998 Winter Simulation Conference*, D. J. Madeiros, E. F. Watson, J. S. Carson, M. S. Manivannan, eds., 661–668.
- Richardson, D., P. J. Green. 1997. On Bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society: Series 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. *Journal of the Royal Statistical Society: Series B* **64**(4) 583–639.
- van Beers, W. C., J. P. C. Kleijnen. 2004. Kriging interpolation in simulation: A survey. In *Proceedings of the 2004 Winter Simulation Conference*, R. Ingalls, M. Rossetti, J. Smith, B. Peters, eds., 113-121.

van Dorp, J. R., Merrick, J. R. W., J. Harrald, T. Mazzuchi, M. Grabowski. 2001. A Risk Management Procedure for the Washington State Ferries. *Risk Analysis* **21**(1) 127-142.

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

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