Proceedings of the 2004 Winter Simulation Conference
R. G. Ingalls, M. D. Rossetti, J. S. Smith, and B. A. Peters, eds.

BAYESIAN METHODS FOR DISCRETE EVENT SIMULATION

Stephen E. Chick

INSEAD
Technology Management Area
Boulevard de Constance
77300 Fontainebleau, FRANCE

ABSTRACT

Bayesian methods are now used in a variety of ways in discrete-event simulation. Applications include input modeling, response surface modeling, uncertainty analysis, and experimental designs for field data collection, selection procedures, and response surface estimation. This paper reviews some fundamental concepts of subjective probability and Bayesian statistics that have led to results in simulation applications.

1 INTRODUCTION

Considerable progress in discrete-event simulation has been made over the last 8 years using Bayesian methods. Advances in other fields can also contribute to the advancement of Bayesian methods in simulation. Glynn (1986) noted a role for Bayesian methods in certain applications of simulation, particularly focusing on *uncertainty analysis*. In that paper, the traditional role of estimating

$$\alpha = h(E[Y])$$

is extended to account for statistical input parameter uncertainty, so $\alpha(\theta) = h(E[Y \mid \theta])$ depends upon unknown parameters with distribution $p(\theta)$ that can be updated with data from the modeled system. Three questions he suggests be examined are (i) how to estimate the distribution of $\alpha(\Theta)$ induced by the random variable Θ, (ii) how to estimate the mean $E[\alpha(\Theta)]$, and (iii) estimation of credible sets, e.g. finding θ_{lo}, θ_{hi} so the probability $p([h(\theta_{lo}), h(\theta_{hi})])$ equals a prespecified value, like 0.95. Chick (1997) reviewed few works to that date that applied Bayesian ideas to simulation, then suggested a broader range of application areas than uncertainty analysis, including *ranking and selection*, *response surface modeling*, and *experimental design*.

This tutorial discusses these fundamental concepts for Bayesian reasoning, and identifies several applications to simulation experiments. Modeling issues that arise in the Bayesian framework are the need to specify initial uncertainty about unknown parameters by *specifying prior distributions* for unknown quantities (including unknown outputs, unknown input parameters and unknown metamodel parameters); specifying *likelihood models* to relate unknown parameters to observable data, if applicable; *numerical tools* to update beliefs about unknown quantities as data becomes available using *Bayes' rule* to obtain posterior distributions for unknown quantities. But quantifying uncertainty is insufficient, since simulations are often intended to aid system design *decisions*. Decisions under uncertainty are guided by the principle of minimizing the *expected loss* associated with a choice of design parameters for a simulated system. Loss functions can also lead to experimental design criteria for structuring simulation experiments.

Although several simulation applications of these ideas are provided, the treatment is necessarily incomplete. Further work that applies Bayesian ideas in simulation includes Chen and Schmeiser (1995), Chen (1996), Scott (1996), Nelson et al. (1997), Chen et al. (1999), Cheng (1999), Lee and Glynn (1999), Andradóttir and Bier (2000), Chick and Inoue (2001a), Chick and Inoue (2001b), Chick (2001), Zouaoui and Wilson (2001), Zouaoui and Wilson (2003), Kurowicka and Cooke (2002), Ng and Chick (2002), Cheng and Currie (2003), Steckley and Henderson (2003). WSC and simulation publications with applications of Bayesian methods to scheduling, insurance, finance, traffic modeling, public health, waterway safety, supply chain and other areas include Popova and Morton (1998), Herzog (2002), Muñoz (2003), Merrick, Dinesh, Singh, van Dorp, and Mazzuchi (2003), McCabe (2003), Chick, Soorapanth, and Koopman (2003). Work on deterministic simulation with potentially important implications for stochastic simulation includes (O'Hagan et al. 1999, Craig et al. 2001, Santner et al. 2003). References for subjective probability and Bayesian statistics in general, not just as they apply to simulation, include Lindley (1972), Winkler (1972), Berger (1985), with special mention for de Finetti (1990), Savage (1972), de Groot (1970) and Bernardo and Smith (1994).

2 MAIN CONCEPTS

We represent a stochastic simulation as a deterministic function of several types of inputs, with

$$Y_r = g(\boldsymbol{\theta}_p, \boldsymbol{\theta}_e, \boldsymbol{\theta}_c; \boldsymbol{U}_r), \tag{1}$$

so that Y_r is the output of the r-th replication, $\boldsymbol{\theta}_p$ is a vector of statistical input parameters that can be inferred with data from the real system a model is intended to represent; $\boldsymbol{\theta}_e$ is a potentially unknown parameter that describes a system's operating environment, but that is *not* inferrable from a stream of potentially available data; $\boldsymbol{\theta}_c$ are control parameters (or design variables) for the modeled system; and $\boldsymbol{U}_r$ represents sampling from different portions of a uniform random number generator to provide 'randomness' in the simulation output.

A subjective probabilist represents all uncertain quantities with probability distributions, including $\boldsymbol{\theta}_p$ and $\boldsymbol{\theta}_e$, not only uniform variates $\boldsymbol{U}$ and nonuniform random variates $\boldsymbol{X}$. We add subscripts if additional specificity is needed, e.g. to refer to the j-th variate for the i-th source of randomness during the r-th replication, we write X_{rij}. See Fig. 1.

One reason a simulation experiment may be run is to estimate the function g because it's exact form is not known. Because of this, it is often useful to study *metamodels* of simulation models to predict the outputs of a simulation model (or the simulated system) when a full simulation takes a long time to run. Examples of metamodels are linear regression models and Gaussian random fields (GRFs). Metamodels may have parameters $\boldsymbol{\psi}$. Sec. 3.4 describes some Bayesian methods to describe uncertainty about $\boldsymbol{\psi}$.

The specification of prior distributions and Bayesian inference with data is discussed in Sec. 2.1 and 2.2. Asymptotic theorems are presented in Sec. 2.3, followed by a discussion of loss functions that can be used to design sampling allocations for simulation experiments.

2.1 Exchangeability and Input Parameters

An important simulation design issue is the selection of appropriate input distributions to characterize the stochastic behavior of the modeled system. This section reviews basic ideas and important theorems for inferring input parameters with the Bayesian formalism. The presentation is in the context of selecting a parameter θ for a specific candidate distribution (e.g. *one* of the Bernoulli, exponential, or gamma distributions) for input into a computer simulation. Sec. 3.2 explores input selection if multiple candidate models for a given source of randomness is proposed.

For a Bayesian, the idea of exchangeability is preferred to the idea of independent and identically distributed (i.i.d.) random variables. Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_N)$ be a generic vector of random variables on an outcome space Ω. A probability p on Ω is said to be *exchangeable* if it is invariant with respect to permutations of the coordinates (e.g. $p(x_1, x_2, \ldots, x_n) = p(x_{s_1}, x_{s_2}, \ldots, x_{s_n})$ for permutations s on $\{1, 2, \ldots, n\}$ for arbitrary n. Exchangeability is a weaker assumption than independence (e.g. for $N = 2$ coin flips, the subjective probability assessment $p((0, 1)) = p((1, 0)) = 1/2$ is an exchangeable Bernoulli model, but not independent, because $X_1 + X_2 = 1$).

Simulation is very often concerned with conceptually infinite ($\lim N \to \infty$) exchangeable sequences of random variables (e.g. replications $r = 1, 2, \ldots$ or service times x_{rij} for $j = 1, 2, \ldots$). A key theorem (de Finetti 1990 or Bernardo and Smith 1994) for conceptually infinite exchangeable sequences of Bernoulli random variables says that outcomes are conditionally independent, given the limiting fraction of heads, $\Theta = \lim_{N\to\infty} \sum_{i=1}^{N} X_i/N$, with some mixture distribution $\pi(\theta)$,

$$\lim_{N\to\infty} p(\boldsymbol{x}_n) = \int \left\{ \prod_{i=1}^{n} f(x_i \mid \theta) \right\} d\pi(\theta), \tag{2}$$

where $p(x_i \mid \theta) = f(x_i \mid \theta) = \theta^{x_i}(1-\theta)^{1-x_i}$ is a conditional probability when considered as a function of x_i and a likelihood when written as a function of θ. A mixture of conceptually infinite sequences of exchangeable random variables written in the form Eq. 2 for an arbitrary parameter θ, distribution $\pi(\theta)$ and likelihood model f is called a *de Finetti-type representation*, although judgments stronger than exchangeability, such as invariance to sums or to an ℓ_p-norm, may be required to justify them (Barlow and Mendel 1992, Chick and Mendel 1998). The notation anticipates the convention of writing a *prior distribution* as $\pi(\cdot)$, representing the *a priori* belief that the parameter takes on a given value. That representation allows for the inference of θ from data $\boldsymbol{x}_n = (x_1, \ldots, x_n)$ via *Bayes rule*,

$$p(\theta \mid \boldsymbol{x}_n) = \frac{\pi(\theta)\, p(\boldsymbol{x}_n \mid \theta)}{p(\boldsymbol{x}_n)}. \tag{3}$$

The *posterior probability* $p(\theta \mid \boldsymbol{x}_n)$ of θ, given $\boldsymbol{x}_n$, summarizes uncertainty about θ via the likelihood model and prior distribution. The assumption of i.i.d. does not allow beliefs about θ to be updated as data is observed.

2.2 Prior Probability Distributions

Bayesian methods require probability distributions to quantify initial uncertainty before data is observed. The selection of a prior distribution is controversial. Bruno de Finetti (1990) argues that a prior distribution is a subjective expression of uncertainty, and that You (yes, You) may justifiably specify a different distribution than I, since we may have different beliefs about the likelihood of a given event. Savage

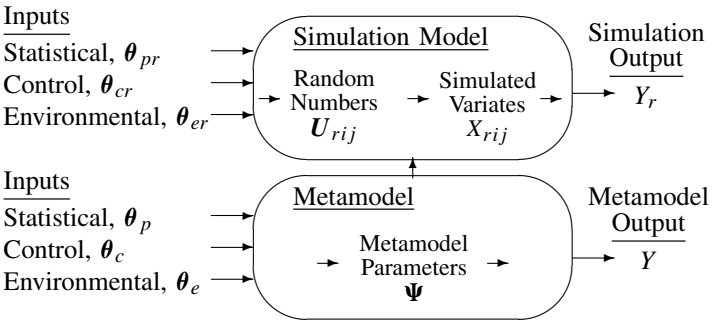

Figure 1: Simulation Takes Multiple Types of Inputs and Metamodels Predict Outputs for Unsimulated Input Values

(1972) suggests a process for eliciting a prior distribution from a modeler through the evaluation of 'fair bets'. Kahneman, Slovic, and Tversky (1982) illustrate potential pitfalls with eliciting probability judgments and present techniques to counter them. While this may seem 'too subjective' and open to biases (Edwards 1984), the ability to include prior information provides necessary flexibility and can be considered an advantage of the approach. Frequentist methods apply only with data, and problems remain (e.g. Sec. 3.2).

To avoid the impression of subjectivity, several 'automated' mechanisms have nonetheless been proposed to support the selection of a prior distribution. When a lot of data are available, the likelihood function will be the dominant term in Bayes' rule, rather than the prior distribution, so these methods may be helpful. The first approach is to obtain a prior distribution for a parameter of an infinite exchangeable sequence as a limiting case of an *indifference judgment* for a finite exchangeable sequences. For the finite exchangeable sequence of $\{0, 1\}$ outcomes, if each of the finite set of alternatives $\theta_N \in \{0/N, 1/N, \ldots, (N-1)/N, 1\}$ is judged equally likely for each N, then $\lim_{N\to\infty} p(\theta_N) \xrightarrow{\mathcal{D}} \texttt{uniform}[0, 1]$, the prior probability model used by Laplace (1812) to assess his prior probability that the sun would come up tomorrow. That approach is coordinate dependent (e.g. indifference for θ versus $\log\theta$). Jeffreys (1946) suggested $\pi(\theta) \propto |H(\theta)|^{1/2} d\theta$, where H is the expected information in one observation,

$$H(\theta) = E_X\left[-\frac{\partial^2 \log p(X \mid \theta)}{\partial\theta^2}\bigg|_\theta\right], \qquad (4)$$

because it has the attractive property of being invariant with respect to coordinate changes in θ. It is 'uniform' with respect to the natural metric induced by the likelihood function (Kass 1989). *Jeffreys' prior* for Bernoulli sampling is a $\texttt{beta}(1/2, 1/2)$ distribution. For some models, Jeffreys' prior is improper (does not integrate to one), but may be useful if the data results in a proper posterior after Bayes' rule is formally applied.

A third approach that is mathematically convenient is to assume a *conjugate prior* distribution, meaning that the posterior distribution has the same functional form as the prior distribution. For Bernoulli(θ) sampling, the $\texttt{beta}(\alpha, \beta)$ distribution with probability density function (pdf) $f(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$ is a conjugate prior. If data $\boldsymbol{x}_n$ is observed, with $s_n = \sum_{i=1}^n x_i$, then the posterior pdf is $f(\theta \mid \boldsymbol{x}_n) \propto \theta^{\alpha+s_n-1}(1-\theta)^{\beta+n-s_n-1}$, a $\texttt{beta}(\alpha+s_n, \beta+n-s_n)$ distribution. Conjugate prior distributions exist for all members of the regular exponential family (Bernardo and Smith 1994), which includes the exponential, normal, gamma, lognormal, Wishart, Bernoulli, geometric, and Poisson distributions, as well as linear regression models with normally distributed error, among others. (The $\texttt{uniform}[0, 1]$ distribution is in the conjugate family for Bernoulli sampling—it is a $\texttt{beta}(1, 1)$ distribution.) Priors selected in this way are often selected to be as *noninformative* as possible, meaning that probability is spread 'evenly' over the space of parameters. Although evenly is subjectively defined, heuristics are available for members of the regular exponential family, whose likelihood function can be written $p(x \mid \theta) = a(x)h_0(\theta)\exp\left[\sum_{j=1}^d c_j\phi_j(\theta)h_j(x)\right]$ for some $a(\cdot), h_0(\cdot), c_j, \phi_j(\cdot), h_j(\cdot)$. The conjugate prior distribution is $p(\theta) = [K(\boldsymbol{t})]^{-1}[h_0(\theta)]^{n_0}\exp\left[\sum_{j=1}^d c_j\phi_j(\theta)t_j\right]$. The posterior distribution given n conditionally independent data points then has parameters $n_0 + n$ and the sum of $\boldsymbol{t} = (t_1, t_2, \ldots, t_d)$ and the sufficient statistics (Bernardo and Smith 1994). The parameter n_0 is therefore interpreted by some to be the 'strength' of the prior, measured in terms of the number of samples. In that case, evenly spreading probability can be taken to mean selecting n_0 close to 0, while insuring that the prior is still proper. Jaynes (1983) suggests a fourth approach that is common in image and signal processing: maximum entropy methods define 'diffuse' prior with respect to a background measure, subject to moment constraints on the parameters. Berger (1994) and Kass and Wasserman (1996) discuss on default prior distributions and sensitivity analysis with respect to them.

Probability modeling is inherently subjective—even so-called 'objective' methods require the subjective specification of a likelihood model. One standard Bayesian practice is to use a slightly informative conjugate distribution for the unknown mean, by choosing it to be proper but diffuse (Gilks, Richardson, and Spiegelhalter 1996). For example, the conjugate prior for an unknown mean of a normal distribution is also a normal distribution. A diffuse prior would be $\texttt{Normal}\left(0, \sigma^2_{big}\right)$ for some large σ^2_{big}. Conjugate prior distributions are mathematically convenient, but care is still required with their use, as with any statistical analysis, Bayesian or otherwise.

2.3 Asymptotic Theorems

Classical asymptotic theorems (laws of large numbers, LLN; central limit theorem, CLT; e.g. Billingsley 1986) have Bayesian interpretations when considered to be conditional on the mean and standard deviation of an infinite exchangeable sequence. A Bayesian extension of the LLN allows for sample averages to converge to random variables rather than to 'true' means.

Theorem 1 (Bayesian LLN) *If $\bar{X}_n$ and $\bar{Y}_m$ are respectively the averages of n and m exchangeable random quantities X_i (the two averages may or may not have some terms in common), the* probability *that*

$$\left|\bar{X}_n - \bar{Y}_m\right| > \epsilon$$

may be made arbitrarily small by taking n and m sufficiently large (de Finetti 1990, p. 216 assumes a finite variance).

Even though the modes of Bayesian posterior distributions may not be the true mean, an asymptotic normality property holds for posterior distributions of parameters.

Theorem 2 (Posterior Normality) *For each n, let $p_n(\cdot)$ be the posterior pdf of the d-dimensional parameter θ_n given $\boldsymbol{x}_n = (x_1, \ldots, x_n)$, let $\tilde{\theta}_n$ be its mode, and define the $d \times d$ Bayesian observed information matrix Σ_n^{-1} by*

$$\Sigma_n^{-1} = -\left.\frac{\partial^2 \log p_n(\theta \mid \boldsymbol{x}_n)}{\partial \theta^2}\right|_{\tilde{\theta}_n}. \tag{5}$$

Then $\phi_n = \Sigma_n^{-1/2}(\theta_n - \tilde{\theta}_n)$ converges in distribution to a standard (multivariate) normal random variable (Bernardo and Smith 1994, Prop 5.14 needs regularity conditions).

Theorem 2 asserts that uncertainty about the value of the unknown parameter value can be approximated asymptotically with a normal distribution. The Bayesian observed information Σ_n^{-1} is a measure of precision of the posterior distribution of θ, and behaves asymptotically like the frequentist observed information (which ignores the prior distribution) under rather general conditions, but the interpretation differs somewhat. The classical analog of Theorem 2 asserts that the MLE is asymptotically normally distributed about a 'true' parameter θ_0 (Law and Kelton 2000), rather than describing uncertainty about θ. The mode $\tilde{\theta}_n$ is often called a MAP (maximum *a posteriori* probability) estimator.

2.4 Value of Information

The fact that input uncertainty is described by probability distributions allows the modeler to assess the *expected value of information* of additional data. The expected value of information is useful in experimental design. It measures the value of resolving uncertainty with respect to a *loss function* $\mathcal{L}(d, \omega)$ that describes the loss when a *decision* d is chosen when the state of nature is ω. The expected improvement in the loss given by the information in an experiment is a Bayesian experimental design criterion.

The value of information idea directly leads to the selection procedures in Sec. 3.3. A simplified version of that problem adapted from de Groot (1970) illustrates the key concepts. Suppose we must decide whether or not the unknown mean W of a normal distribution (known sampling variance σ^2) is smaller (decision $d = 1$) or larger ($d = 2$) than w_0. Conditionally independent samples $\boldsymbol{X}_n = (X_1, X_2, \ldots, X_n)$, with $p(X_i) \sim \texttt{Normal}\left(w, \sigma^2\right)$ given $W = w$, can be used to infer the value of the mean. The decision maker designs an experiment (chooses n) to balance the cost of sampling, cn, and the expected penalty if the wrong answer is chosen. Here the penalty for incorrect selection is the opportunity cost $\mathcal{L}(1, w)$, the difference between the actual value of w and w_0 when the wrong answer is selected, and 0 if the right answer is selected.

$$\mathcal{L}(1, w) = \begin{cases} 0 & \text{if } w \le w_0 \\ w - w_0 & \text{if } w > w_0, \end{cases}$$

$$\mathcal{L}(2, w) = \begin{cases} w_0 - w & \text{if } w \le w_0 \\ 0 & \text{if } w > w_0. \end{cases}$$

Since the mean is not known exactly, there is a potential penalty for incorrectly specifying whether W is smaller or larger than w_0. We model uncertainty about W with a $\texttt{Normal}(\mu, 1/\tau)$ prior distribution, which is conjugate for normal sampling with an unknown mean and known variance (Bernardo and Smith 1994). Here, τ is the *precision* in our uncertainty about W. Observing $\boldsymbol{X}_n = \boldsymbol{x}_n$ would reduce the uncertainty and result in the posterior distribution

$$p(w \mid \boldsymbol{x}_n) \sim \texttt{Normal}\left(z, \tau_n^{-1}\right),$$

where

$$z = \text{posterior mean of } W = E[W \mid \boldsymbol{x}_n] = \frac{\tau\mu + \frac{n}{\sigma^2}\bar{x}_n}{\tau + \frac{n}{\sigma^2}},$$

and τ_n = posterior precision of $W = \tau + n/\sigma^2$.

The variance τ_n^{-1} equals the posterior variance approximation Σ_n in Eq. 5 because Σ_n is based on a normal distribution approximation.

The posterior mean z influences the decision, but it depends upon n, which must be selected before $\boldsymbol{X}_n$ is observed. We need the *predictive distribution* $p(z)$ of the posterior mean $Z = E[W \mid \boldsymbol{X}_n] = (\tau\mu + n\bar{X}_n)/\tau_n$ to see how n samples influences the decision d. The conditional distribution of $\bar{X}_n$ given w is $\texttt{Normal}\left(w, \sigma^2/n\right)$. Mixing over the prior distribution of W implies that the predictive distribution for Z is $\texttt{Normal}\left(\mu, \tau_z^{-1}\right)$, where

$$\tau_z \quad = \quad \tau(\tau + n/\sigma^2)/(n/\sigma^2). \tag{6}$$

The variance τ_z^{-1} of Z is 0 when $n \to 0$ (no new information). If $n \to \infty$ (perfect information about w), then $\text{Var}[Z] \to$ prior variance for W, or σ^2.

The experimental design that minimizes risk (the cost of sampling plus expected losses due to a potentially incorrect decision) is the n that minimizes a nested expectation, an inner expectation corresponds to the expected loss after $\boldsymbol{X}_n$ is observed, an outer expectation averages over $\boldsymbol{X}_n$,

$$\rho(n) \quad = \quad cn + E_{\boldsymbol{X}_n}[E_W[\mathcal{L}(d(\boldsymbol{X}_n), W) \mid \boldsymbol{X}_n]]. \tag{7}$$

One technique to analyze $E_W[\mathcal{L}(d(\boldsymbol{X}_n), W) \mid \boldsymbol{X}_n]$ is to obtain an auxiliary loss function $\mathcal{L}^*$ that has the same optimal decision, but simplifies the loss function by making the loss of one of the decisions equal to 0. Adding a function of w does not change the optimal decision (de Groot 1970). Set $\mathcal{L}^*(d, w) = \mathcal{L}(d, w) - \mathcal{L}(1, w)$, which is 0 if $d = 1$ and is $w_0 - w$ if $d = 2$. Then

$$E_W[\mathcal{L}^*(d(\boldsymbol{X}_n), W) \mid \boldsymbol{X}_n] \quad = \quad \begin{cases} 0 & \text{if } d = 1 \\ w_0 - Z & \text{if } d = 2. \end{cases} \tag{8}$$

The decision that minimizes the loss in Eq. 8 is to assert $d(\boldsymbol{X}_n) = 2$ ('bigger') if the posterior mean exceeds the threshold, $Z > w_0$, and to assert $d(\boldsymbol{X}_n) = 1$ ('smaller') if $Z \le w_0$.

The expectation over the outcomes $\boldsymbol{X}_n$ can be determined with well-known tables because the decision depends upon $\boldsymbol{X}_n$ only through Z, and Z has a normal distribution. The expected loss can be determined by the standard normal loss $\Psi[s] = \int_s (t-s)\phi(t)dt = \phi(s) - s(1 - \Phi(s))$ for expected lost sales in the newsvendor problem if demand is normally distributed (e.g. Porteus 2002).

$$\begin{aligned} E[\mathcal{L}^*(d(\boldsymbol{X}_n), W)] &= E_{\boldsymbol{X}_n}[E_W[\mathcal{L}^*(d(\boldsymbol{X}_n), W) \mid \boldsymbol{X}_n]] \\ &= -\int_{w_0}^{\infty} (z - w_0)\, p(z \mid \boldsymbol{X}_n) dz \\ &= -\tau_z^{\frac{-1}{2}} \Psi[\tau_z^{\frac{1}{2}}(w_0 - \mu)] \end{aligned}$$

The expected loss of the original loss function is recovered by adding back $E[\mathcal{L}(1, W)]$, using the prior distribution of W for the expectation.

$$\begin{aligned} E[\mathcal{L}(d(\boldsymbol{X}_n), W)] &= \tau^{\frac{-1}{2}} \Psi[\tau^{\frac{1}{2}}(w_0 - \mu)] \\ &\quad -\tau_z^{\frac{-1}{2}} \Psi[\tau_z^{\frac{1}{2}}(w_0 - \mu)] \end{aligned} \tag{9}$$

The *expected value of information* for m samples is the difference between Eq. 9 when $n = 0$ and when $n = m$ (τ_z depends on n). Combine Eq. 9 with Eq. 6-7, note that $d\Psi/ds = \Phi(s) - 1$ and $d\tau_z/dn = -\tau^2\sigma^2/n$, and take the derivative with respect to n (relaxing the integer assumption) to obtain an optimality condition for the sample size.

$$\frac{\partial \rho}{\partial n} = \frac{1}{2}\tau_z^{-\frac{3}{2}} \phi[\tau_z^{\frac{1}{2}}(w_0 - \mu)] \cdot \frac{-\tau^2\sigma^2}{n} + c = 0$$

For diminishing costs $c \to 0$, the sample size is large. Since $\tau_z \to \tau$ as $n \to \infty$, the optimal sample size n is (asymptotically) approximately

$$n^* = \left(\tau^{\frac{1}{2}} \sigma^2 \phi[\tau_z^{\frac{1}{2}}(w_0 - \mu)]/(2c) \right)^{1/2}. \tag{10}$$

This argument illustrates the basic ideas of loss functions, and the use of predictive distributions for future samples to infer the expected value of information of sampling. *The technique of adding functions of the unknowns can be useful to simplify the derivation of the optimal solution. Asymptotic approximations are a further tool to identify criteria-based sampling plans.* Extensions of this basic argument justify the value of information based selection procedures summarized in Sec. 3.3 (Chick and Inoue 2001a).

An alternate mechanism to approximate the effect of information on parameter uncertainty is based on a thought experiment for the posterior probabilities of parameters. For members of the regular exponential family, the asymptotic variance approximation Σ_n in Eq. 5 simplifies to the form $H^{-1}(\theta)/(n_0 + n)$, where H is the expected information from one observation (Eq. 4), when a canonical conjugate prior distribution is used (Bernardo and Smith 1994). To approximate the effect of collecting m additional samples on the parameter uncertainty, one could presume that the posterior distribution changes from $\texttt{Normal}\left(\tilde{\theta}_n, \Sigma_n\right)$ to

$$\texttt{Normal}\left(\tilde{\theta}_n, \Sigma_n \frac{n_0 + n}{n_0 + n + m}\right). \tag{11}$$

This transformation reflects an appropriate scaling of the posterior precision, and the idea is used in a frequentist context for estimating how many replications are required to achieve a confidence interval of a given size (Law and Kelton 2000). Chen (1996) uses this type of approximation

for the Bayesian posterior distribution of the unknown means of several simulated systems in order to motivate a class of ranking and selection procedures called the $OCBA$. Ng and Chick (2001), Chick and Ng (2004) use the idea to plan experiments to reduce input uncertainty in a way that reduces output uncertainty.

2.5 Entropy and Kullback-Leibler Discrepancy

Kullback-Leibler discrepancy is a useful measure of the difference between two distributions. For discrete distributions $\tilde{p}$ and p, the discrepancy is $\delta(p \mid\mid \tilde{p}) = \sum \tilde{p}_i \log(\tilde{p}_i/q_i)$. Two continuous distributions for a random variable X with densities $\tilde{f}$ and $f_\theta = f(x \mid \theta)$ have discrepancy

$$\delta(f_\theta \mid\mid \tilde{f}) = \int \tilde{f}(x) \log \frac{\tilde{f}(x)}{f(x \mid \theta)} dx. \qquad (12)$$

One application for discrepancy is as a loss function for a decision maker that must specify a probability distribution. If the decision-maker believes that the distribution is $\tilde{f}$, and loses $\delta(f \mid\mid \tilde{f})$ if he/she provides a distribution f, then the decision-maker should honestly report $\tilde{f}$ to minimize the expected loss (Bernardo and Smith 1994).

A second application of discrepancy is as a design criterion for experiments designed to assist with parameter estimation (Bernardo 1979, Smith and Verdinelli 1980, Ng and Chick 2002). In the simulation context, this could mean selecting a design matrix $\mathbf{d}_\Theta$ of r vectors of inputs $(\boldsymbol{\theta}_{pi}, \boldsymbol{\theta}_{ei}, \boldsymbol{\theta}_{ci})$ for $i = 1, 2, \ldots, r$ with output $\mathbf{Y}$ in order to best differentiate the posterior distribution of the response parameters $\boldsymbol{\psi}$ from the prior distribution for $\boldsymbol{\psi}$. In other words, the goal is to select the $\mathbf{d}_\Theta$ from a set of possible design matrices to maximize the expected divergence of posterior and prior distributions for $\boldsymbol{\psi}$, the expectation over the predictive distribution of the output $\mathbf{Y}$.

$$\int p(\mathbf{Y} \mid \mathbf{d}_\Theta) \left(\int p(\boldsymbol{\psi} \mid \mathbf{Y}) \log \frac{p(\boldsymbol{\psi} \mid \mathbf{Y})}{p(\boldsymbol{\psi})} d\boldsymbol{\psi} \right) d\mathbf{Y} \qquad (13)$$

The predictive distribution $p(\mathbf{Y} \mid \mathbf{d}_\Theta)$ of future outputs depends upon the current uncertainty about $\boldsymbol{\psi}$ and the design matrix $\mathbf{d}_\Theta$. This approach is essentially an expected value of information design criterion, as in Sec. 2.4, except now the loss function is the Kullback-Leibler discrepancy rather than the opportunity cost.

Other applications of discrepancy include the maximum entropy prior distribution mentioned above (Jaynes 1983), and for input distribution selection, as in Sec. 3.2 below.

3 APPLICATIONS

3.1 Uncertainty Analysis

A sensitivity analysis tests how the mean simulation output depends upon one or more input parameters as that parameter is varied (estimating $E[g(\theta) \mid \mathcal{E}]$ as a function of θ, given all information $\mathcal{E}$). Uncertainty analysis entails propogating input parameter uncertainty about Θ through to uncertainty about outputs Y. Even if a simulation has no random number stream, a distribution on unknown inputs means that the output is random.

An unbiased estimator of the mean output $E[Y \mid \mathcal{E}]$ with both stochastic (from $\boldsymbol{u}$) and systemic (or parameter) uncertainty accounted for is obtained from the Bayesian model average (BMA) in Fig. 2, which averages over random inputs sampling according to the distribution $p(\theta \mid \mathcal{E})$ (Draper 1995, Chick 2001). Zouaoui and Wilson (2003) explore the relative magnitude of stochastic and systemic uncertainty with variations on the BMA, and discuss how to update the estimate should new data become available (so the algorithm need not be rerun from scratch). Importance sampling techniques can reweight estimates accordingly (with likelihood ratio determined as the ratio of the 'new' posterior divided by the 'old' distribution). Andradóttir and Glynn (2004) examine the estimation of $E[Y(\Theta) \mid \mathcal{E}]$ when there may be bias in the estimates of Y given θ, when quasi-random sequences are used in place of the pseudo-random sequences assumed by Fig. 2, or when numerical techniques like Simpson's rule are employed to select values of θ.

Other issues for sensitivity analysis include estimation of the distribution of the conditional expectation $E[Y \mid \Theta, \mathcal{E}]$. When Y is a deterministic function of Θ, then naive Monte Carlo simulation can be used with traditional kernel estimation techniques to assess the distribution of $Y(\Theta)$. When the simulation is stochastic (depends on the random number stream $\boldsymbol{u}$), then $E[Y \mid \theta, \mathcal{E}]$ is imperfectly estimated for any given θ. Given several technical conditions (e.g. univariate continuous-valued θ, monotonic mean response), Steckley and Henderson (2003) derive asymptotically optimal ways of selecting by cleverly selecting r and m in Fig. 2 to produce a kernel density estimator based on the output. Their work builds upon earlier work by Lee

for $r = 1, \ldots, R$ replications
  sample parameter θ_r from $p(\theta \mid \mathcal{E})$
  for $i = 1, 2, \ldots, n$
    generate simulation output y_{ri} given input θ_r
  end loop
end loop
Estimate $E_Y[Y \mid \mathcal{E}]$ with $\bar{y} = \sum_{r=1}^{R} \frac{1}{R} \sum_{i=1}^{n} y_{ri}/n$.

Figure 2: Bayesian Model Average (BMA)

and Glynn (1999) to estimate the distribution function of $E[Y \mid \Theta, \mathcal{E}]$ for the case of discrete-valued θ.

3.2 Selecting from Multiple Candidate Distributions

Input selection in simulation practice may consider $q > 1$ candidate distributions for potential input to a model. The usual approach Law and Kelton (2000) is to find the MLE for each candidate distribution, assess the goodness of fit for each model, then select a model that 'fits well'. While this is practically appealing, there are known problems with the approach (Lindley 1957, Berger and Pericchi 1996, Chick 2001). *In the simulation context, input uncertainty can make standard confidence intervals for the mean output almost meaningless if the classical approach is used* (Chick 2001, Barton and Schruben 2001), because an excellent simulation estimate based on point estimates for inputs misses the boat if there is uncertainty about input values, which is typical for simulations of real systems.

A Bayesian approach with model uncertainty is basically the same as for parameter uncertainty alone, as in Sec. 2.1 above, except now that a prior probability distribution needs to be placed on the model/parameter combination, $\pi(m, \theta_m)$, a mixed discrete-continuous model, where $m \in \{1, 2, \ldots, q\}$ indexes the set of candidate distributions, and θ_m is the parameter for the m-th candidate distribution. As data $\boldsymbol{x}_n$ becomes available, the BMA then requires sampling from the joint posterior $p(m, \theta_m \mid \boldsymbol{x}_n)$. This can be accomplished by composition, sampling the input model then the parameter with $p(m \mid \boldsymbol{x}_n)p(\theta_m \mid m, \boldsymbol{x}_n)$.

Chick (2001) first illustrated Bayesian input modeling in a stochastic simulation context, and suggested a method-of-moments approach for assessing prior distributions for the unknown parameters of each candidate model. Zouaoui and Wilson (2001) noted a decoupling of stochastic uncertainty from two types of structural uncertainty (that due to uncertainty about the candidate model, plus that due to uncertainty about the parameters of the candidate models) under special conditions, provided a variance reduction for the BMA and numerical analysis.

Selecting models according to $p(m, \theta_m \mid \mathcal{E})$ is consistent in that if one of the entertained models is actually the true model, then the true model is selected if enough data is observed and some regularity conditions hold. When the true model is not among those being considered, Bayesian model selection chooses the model that is closest to the true model in terms of Kullback-Leibler divergence (Berk 1966, Bernardo and Smith 1994, Dmochowski 1999).

3.3 Ranking and Selection

Simulation procedures to select the best system using Bayesian approaches come in two different flavors.

Chick and Inoue (2001a) used expected value of information ideas that extend Eq. 9-10 to obtain two-stage value of information procedures (VIPs) with independent replications that identify the best of a finite set of simulated systems. Eq. 9 was generalized to allow for unknown and potentially different variances for multiple simulated systems. The loss for $k > 2$ systems was approximated by the sum of losses for $k - 1$ pairwise comparisons between the system selected in the absence of additional replications and each of the other systems. Additional replications for the second stage are allocated so minimize that loss function using an asymptotic (in the number of replications) approximation like that used for Eq. 10. A variation improves the (Bayesian posterior) probability of correct selection with the 0-1 loss function (loss of 1 if the wrong system is selected, 0 if the best is selected), and sequential variations exist (Chick and Inoue 2001a, Chick and Inoue 2002). The procedures, named $\mathcal{LL}(\mathcal{B})$, $\mathcal{LL}(\mathcal{S})$, 0-1$(\mathcal{B})$, 0-1$(\mathcal{S})$ depending upon whether an opportunity cost (aka linear loss) or 0-1 loss function is used, and whether a budget constraint or sequential sampling apply, are empirically quite effective for identifying the best system with respect to several figures of merit and relative to several other procedures (Chick and Inoue 2001a, Inoue et al. 1999). The idea has also been extended to handle common random numbers with screening to improve efficiency. The analysis to justify them also requires missing data techniques (Chick and Inoue 2001b).

Chen (1996) and Chen et al. (1999) use a different tack—the thought experiment in Eq. 11 that supposes that additional replications won't change the estimate of the means of each system, but can reduce the variance associated with the estimate. They provide empirically effective procedures (called $OCBA$) with only a few parameters to tune for identifying the best system with high probability. A variation to reduce the expected opportunity cost remains future work for this approach.

3.4 Metamodels

Metamodels are a model of how simulated responses are believed to behave as a function of input parameters, even for values of parameters not yet input to the simulation. This is particularly useful when the simulation model requires extensive computation. In addition to describing the response, metamodels can be used to predict how reductions in input parameter uncertainty can reduce uncertainty about the mean system performance. Since the metamodel is unknown, uncertainty about it can be modeled from a Bayesian perspective. Here we discuss the normal linear model and Gaussian random function (GRF) metamodels.

The normal linear model is

$$Y = \sum_{\ell=1}^{p} g_\ell(\boldsymbol{\theta})\beta_\ell + Z(\boldsymbol{\theta}; \boldsymbol{U}) = \boldsymbol{g}^T(\boldsymbol{\theta})\boldsymbol{\beta} + Z(\boldsymbol{\theta}; \boldsymbol{U}), \quad (14)$$

for *known* regression functions $g_1, \ldots, g_p$, potentially unknown regression coefficients $\boldsymbol{\beta}$, and a zero-mean random noise $Z(\cdot)$. The conjugate prior is an inverted gamma distribution for the unknown variance σ^2 and a conditionally normal distribution for $\boldsymbol{\beta}$ given σ^2, if all factors are active (Bernardo and Smith 1994). The metamodel parameters are then $\boldsymbol{\psi} = (\boldsymbol{\beta}, \sigma^2)$. Raftery, Madigan, and Hoeting (1997) describe a relatively 'uninformative' prior distribution for $\boldsymbol{\psi}$ for this normal linear model. A vector of simulation output $\mathbf{Y} = \mathbf{y}$ obtained with design matrix $\mathbf{d}_{\boldsymbol{\Theta}}$ with inputs $(\boldsymbol{\theta}_{pi}, \boldsymbol{\theta}_{ei}, \boldsymbol{\theta}_{ci})$ for runs $i = 1, \ldots, r$ can be used to update the posterior distribution of unknown response parameters, $p(\boldsymbol{\psi} \mid \mathbf{y}, \mathbf{d}_{\boldsymbol{\Theta}})$, using Bayes' rule.

Identifying important factors (factors with nonzero β_i) can be formulated as a Bayesian model selection problem, selecting from 2^p different response models that are distinguished by the presence or absence of each factor. George and McCulloch (1996) and Cheng (1999) discuss techniques for estimating which factors are active with what probability. Ng and Chick (2002) describe an entropy-based experimental design criterion (cf. Sec. 2.5) to identify both which factors are active and reduce parameter uncertainty simultaneously.

When the g_i represent the individual dimensions of the unknown parameters $(\boldsymbol{\theta}_p, \boldsymbol{\theta}_e)$, the β_ℓ are gradients with respect to the inputs. If the model has only statistical input parameters $\boldsymbol{\theta}_p$ for which data can be collected (but not parameters $\boldsymbol{\theta}_e$ for which no data is available), Ng and Chick (2001) and Zouaoui and Wilson (2003) indicate that output uncertainty can be decoupled asymptotically or under special conditions.

$$\begin{aligned} \mathrm{Var}[\bar{Y} \mid \mathcal{E}] &\approx \text{stochastic} + \text{parameter uncertainty} \\ &\approx \frac{\hat{\sigma}_0^2}{m} + \frac{\boldsymbol{\beta} \hat{H}_{\hat{\boldsymbol{\theta}}_p}^{-1} \boldsymbol{\beta}}{n}, \end{aligned}$$

where $\hat{\sigma}_0^2$ is the estimate of the variance from m replications, the MLE $\hat{\boldsymbol{\theta}}_p$ and estimate $\hat{H}_{\hat{\boldsymbol{\theta}}_p}^{-1}$ of the information in one observation are based on n data points, and some technical conditions hold. This adapts a frequentist result of Cheng and Holland (1997).

Ng and Chick (2001) use that decoupling, applied to uncertainty due to multiple input parameters,

$$\sum_{i=1}^{n_p} \boldsymbol{\beta}_i \hat{H}_{\hat{\boldsymbol{\theta}}_{pi}}^{-1} \boldsymbol{\beta}_i / n_i,$$

to provide sampling plans to collect further data to reduce input parameter uncertainty to optimally reduce output uncertainty in some sense, assuming that different numbers of data points can be collected to infer the parameters of different sources of randomness (e.g. arrival rates versus service time distributions). Chick and Ng (2004) extend that analysis by accounting for uncertainty about $\boldsymbol{\beta}$; suggest allocations of resources for either running more replications or collecting more data points to reduce the asymptotic output variance approximation; and give a numerical analysis.

To date, much simulation research seeks analytical results for stochastic models, or mechanisms to reduce the variance of estimators due to stochastic noise. Those results need to be complemented with an *understanding of how performance depends on input uncertainty, and methods to reduce input uncertainty to effectively reduce output uncertainty.* The Bayesian approach is a tool that can help.

Gaussian random functions (GRFs) are well-known response models in deterministic simulations, particularly in geostatistics (Cressie 1993, Santner et al. 2003), but are less well known in stochastic simulation. GRFs provide flexibility that the linear model does not, and are useful when g takes a long time to compute. The GRF for an unknown nonstochastic g (no random numbers $\boldsymbol{u}$) is

$$Y(\boldsymbol{\theta}) = \sum_{\ell=1}^{p} g_\ell(\boldsymbol{\theta}) \beta_\ell + Z(\boldsymbol{\theta}) = \boldsymbol{g}^T(\boldsymbol{\theta}) \boldsymbol{\beta} + Z(\boldsymbol{\theta}) \qquad (15)$$

for *known* regression functions $g_1, \ldots, g_p$ of $\mathbb{R}^d$, unknown regression coefficients $\boldsymbol{\beta}$, and the zero-mean random second-order process such that for any distinct inputs $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m$, the vector $(Y_1, \ldots, Y_m)$ has multivariate normal distribution, conditional on the value $\boldsymbol{\beta}$. GRFs are determined by their mean $\boldsymbol{g}^T(\boldsymbol{\theta})\boldsymbol{\beta}$ and (auto)covariance function $C^*(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \mathrm{Cov}(Y(\boldsymbol{\theta}_1), Y(\boldsymbol{\theta}_2))$ It is common to assume strong stationarity ($(Y_1, \ldots, Y_m)$ and $(Y_1 + \boldsymbol{h}, \ldots, Y_m + \boldsymbol{h})$ have the same distribution), so that $C^*(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = C(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2)$.

Inference for $g(\boldsymbol{\theta})$ at values of $\boldsymbol{\theta}_{r+1}$ not yet input to a simulation model is enabled via the correlation function $R(\boldsymbol{h}) = C(\boldsymbol{h})/C(0)$ for $\boldsymbol{h} \in \mathbb{R}^d$. An example is the power exponential function $R(\boldsymbol{h}) = \prod \exp[-|h_i/\gamma_i|^{p_i}]$ for $p_i \in [0, 2]$. Kriging, a geostatistics term, is a best linear unbiased prediction (BLUP) for $g(\boldsymbol{\theta}_{r+1})$. An assessment of the uncertainty in $g(\boldsymbol{\theta}_{r+1})$ can be used as an experimental design technique to choose inputs to reduce response uncertainty (Santner et al. 2003).

The advanced tutorial by van Beers and Kleijnen (2004) in this year's WSC more fully expands upon kriging . See also Sacks et al. (1989), O'Hagan et al. (1999), Kennedy and O'Hagan (2001), Santner et al. (2003), van Beers and Kleijnen (2003). GRFs facilitate experimental designs to reduce the computational effort needed to get a good response estimate by selecting simulation inputs on areas where the mean response has the greatest uncertainty. Most work for GRFs has been in a nonstochastic simulation context. *More work is needed for GRFs in the stochastic simulation context.*

4 IMPLEMENTATION

Three basic computational issues for implementing a Bayesian analysis are maximization (e.g. to find the MLE $\hat{\theta}$ or MAP $\tilde{\theta}$ estimators for a posterior distribution); integration, either to find a marginal distribution (e.g., find $p(\theta_1 \mid \boldsymbol{x}_n)$ from $p(\theta_1, \theta_2 \mid \boldsymbol{x}_n)$) or constant of proportionality for a posterior distribution (e.g. find $c^{-1} = \int f(\boldsymbol{x}_n \mid \theta) d\pi(\theta)$, or $p(m \mid \boldsymbol{x}_n)$ for Sec. 3.2); and simulation (e.g., sample from $p(\theta \mid \boldsymbol{x}_n)$ in order to estimate $E[g(\theta) \mid \boldsymbol{x}_n]$). These techniques are described in a variety of sources (e.g. Evans and Swartz 1995, Tanner 1996, Gilks, Richardson, and Spiegelhalter 1996, Devroye 2004).

For maximization, a number of methods are available including gradient-based methods (e.g. Newton-Raphson), gradient-free methods (e.g. Nelder-Mead), and simulation-based methods. The expectation-maximization (EM) algorithm is a technique for finding the MAP or MLE when there is missing data or nuisance parameters are to be integrated out (e.g. the MAP of $p(\theta_1 \mid \boldsymbol{x}_n)$ when it is 'messy' but $p(\theta_1, \theta_2 \mid \boldsymbol{x}_n)$ is easier to manipulate).

For integration, five general techniques apply when analytical results are not available: asymptotic methods, Markov chain methods, importance sampling, adaptive importance sampling, and multiple quadrature. Quadrature is useful when the number of dimensions is not too large. The Laplace method is an interesting approximation for integrals $\int g(\theta) f(\theta \mid \boldsymbol{x}_n) d\pi(\theta)$ based on asymptotic normality results (like Eq. 5), and it applies even if $f(\theta \mid \boldsymbol{x}_n)$ is only proportional to the posterior distribution. It also applies for integrating out nuisance parameters if regularity conditions hold. Another effective technique for approximating $p(\theta_1 \mid \boldsymbol{x}_n)$ (not just the MLE or MAP) when it is 'messy' but $p(\theta_1 \mid \theta_2, \boldsymbol{x}_n)$ and $p(\theta_2 \mid \theta_1, \boldsymbol{x}_n)$ are easy to manipulate is data augmentation, often called the IP algorithm (for imputation, posterior algorithm). *Importance sampling (IS) remains a powerful methods for efficient integration.*

For simulation of variates, classical methods for generating independent variates from posterior distributions may apply. Posterior distributions are often known only up to a constant of proportionality (the denominator in Bayes rule may be hard to compute). It is therefore important to have a method to simulate variates for arbitrary functions proportional to posterior distributions. *Markov Chain Monte Carlo (MCMC) is the most important of those methods at present.* MCMC constructs a Markov chain whose stationary distribution is the desired posterior distribution $p(\theta \mid \mathcal{E})$. The state θ_t of the Markov chain can then be sampled to obtain (somewhat correlated) samples from $p(\theta \mid \mathcal{E})$.

A generic MCMC algorithm is the Metropolis-Hastings algorithm in Fig. 3. Potential state transitions to θ_t from θ_{t-1} are proposed by sampling from a candidate distribution $q(\cdot \mid \theta_{t-1})$. Candidates are accepted with a probability α chosen to insure the transition kernel results in $p(\theta \mid \mathcal{E})$ being

Initialize $t = 0$, θ_0
for $t = 1, 2, \ldots$
  sample a candidate $\theta \sim q(\cdot \mid \theta_{t-1})$
  compute acceptance probability
    $\alpha(\theta_{t-1}, \theta) = \min\left\{1, \frac{p(\theta \mid \mathcal{E}) \cdot q(\theta_{t-1} \mid \theta)}{p(\theta_{t-1} \mid \mathcal{E}) \cdot q(\theta \mid \theta_{t-1})}\right\}$.
  generate an independent $u \sim$ `uniform`$[0, 1]$
  if $u \leq \alpha(\theta_{t-1}, \theta)$ then set $\theta_t \leftarrow \theta$
    otherwise set $\theta_t \leftarrow \theta_{t-1}$
  set $t \leftarrow t + 1$
end loop

Figure 3: Metropolis-Hastings: An MCMC Algorithm

a stationary distribution. The choice of $q(\cdot \mid \cdot)$ can make Θ_t more or less correlated with Θ_{t-1}, or can make the algorithm more or less computationally intensive. The Gibbs sampler updates one randomly selected coordinate of a multidimensional $\theta = (\vartheta_1, \ldots, \vartheta_d)$ at a time, using a 'full conditional' distribution: updating ϑ_i with $q(\cdot \mid \theta) = p(\cdot \mid \theta_{.-i}, \mathcal{E})$, where $\theta_{.-i} = (\vartheta_1, \ldots, \vartheta_{i-1}, \vartheta_{i+1}, \ldots, \vartheta_d)$ is all of θ except the i-th coordinate. Such a sampler has an acceptance probability of $\alpha = 1$ (Gilks et al. 1996). Sampling intermittently from the chain (e.g. $t = 1, 11, 21, \ldots$) is one way to generate random inputs for the BMA in Fig. 2. More sophisticated methods exist. Samples and estimators based on MCMC need evaluation to assure reasonable convergence for estimators and faithfulness to the posterior distribution. This is a rich area of study, and calls for results from steady-state simulation.

A number of tools implement Bayesian inference. The `BUGS` and `WinBUGS` packages implement Gibbs sampling and some Metropolis sampling, and is available on the WWW (Spiegelhalter et al. 1996). BOA, for Bayesian output analysis (Smith 2004), is a set of MCMC diagnostic tools for convergence and data analysis that functions with the `R` or `S-PLUS` statistical packages. Gauss and Matlab are also used to implement MCMC methods.

At present, it is possible to input randomized input parameters to some commercial discrete-event simulation packages to implement the BMA algorithm of Fig. 2, but interfaces are not yet fully user friendly. A user-friendly tool to implement the BMA and other uncertainty analysis needs in commercial discrete-event simulation packages would be helpful. Uncertainty analysis for other Monte Carlo applications have been available as a spreadsheet tool for some time (e.g. Winston 2000).

5 CONCLUSIONS

A variety of applications of Bayesian methods apply to simulation experiments, including uncertainty analysis, ranking and selection, input distribution modeling, response surface modeling, and experimental designs. One main theme is

to represent all uncertainty with probability distributions, to update probability using Bayes' rule, and to use the expected value of information as a technique to make sampling decisions (e.g., the opportunity cost and 0-1 loss functions for selection procedures, or the Kullback-Leibler divergence for parameter estimation for linear response models). The other main theme is to use simulation to efficiently estimate quantities of interest for a Bayesian analysis. Asymptotic approximations are often helpful when exact optimal solutions are difficult to obtain. Research opportunities include:

- Input modeling and uncertainty analysis (kernel estimation of outputs as a result of input uncertainty; the effect of different candidate distributions on uncertainty; tools to help elicit prior distributions for simulation inputs, e.g. queueing parameters that do not lead *a priori* to the absence of posterior moments metrics like the average queue length, even conditioning on stability).
- Response modeling (further extending the Gaussian random field work in the world of stochastic simulation; sampling plans for input parameter inference to optimally reduce output uncertainty, including nonasymptotic results, to help understand what data is most important to collect to infer the value of inputs for simulations; theory and improved implementation for the calibration/inverse problem, i.e. how to 'invert' probability distributions for output results backwards to obtain a reasonable probability distribution on unknown 'inputs').
- Bayesian methods for experimental designs (estimating quantiles or other non-expected value goals; CRN for unknown input parameters for ranking and selection; non-Gaussian output for ranking and selection and GRFs).
- Improved computational tools (in MCMC and other sampling methods for posterior distributions).

REFERENCES

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

Andradóttir, S., and P. W. Glynn. 2004. Computing Bayesian means using simulation. ISYE Technical Report, Georgia Tech.

Barlow, R. E., and M. B. Mendel. 1992. De Finetti-type representations for life distributions. *Journal of the American Statistical Association* 87:1116–1122.

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

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

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

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

Berk, R. 1966. Limiting behaviour of posterior distributions when the model is incorrect. *Annals of Mathematical Statistics* 37 (1): 51–58.

Bernardo, J. 1979. Expected information as expected utility. *Annals of Statistics* 7:686–690.

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

Billingsley, P. 1986. *Probability and Measure*. 2nd ed. New York: John Wiley & Sons, Inc.

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

Cheng, R. C. H. 1999. Regression metamodelling in simulation using Bayesian methods. In *Proceedings of the 1998 Winter Simulation Conference*, ed. P. A. Farrington, H. B. Nembhard, D. Sturrock, and G. Evans, 330–335. Piscataway, NJ: IEEE, Inc.

Cheng, R. C., and C. S. Currie. 2003. Prior and candidate models in the Bayesian analysis of finite mixtures. In *Proceedings of the 2003 Winter Simulation Conference*, ed. S. Chick, P. Sánchez, D. Ferrin, and D. Morrice, 392–398. Piscataway, NJ: IEEE, Inc.

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

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

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

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

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

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

Chick, S. E., and K. Inoue. 2001b. New procedures for identifying the best simulated system using common

random numbers. *Management Science* 47 (8): 1133–1149.

Chick, S. E., and K. Inoue. 2002. Corrigendum: New selection procedures. *Operations Research* 50 (3): 566.

Chick, S. E., and M. B. Mendel. 1998. New characterizations of the no-aging property and the ℓ_1-isotropic models. *Journal of Applied Probability* 35 (4): 903–910.

Chick, S. E., and S.-H. Ng. 2004. Reducing parameter uncertainty for stochastic systems. Technical report, INSEAD, Technology Management Area, Fontainebleau, France.

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

Craig, P., M. Goldstein, J. Rougier, and A. H. Seheult. 2001. Bayesian forecasting for complex systems using computer simulators. *Journal of the American Statistical Association* 96 (454): 717–729.

Cressie, N. A. 1993. *Statistics for spatial data*. New York: John Wiley & Sons, Inc.

Devroye, L. 2004. Nonuniform random variate generation. In *Handbook in Operations Research and Management Science: Simulation*, ed. S. G. Henderson and B. L. Nelson. Elsevier. (to appear).

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

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

Dmochowski, J. 1999. Intrinsic priors via Kullback-Leibler geometry. In *Bayesian Statistics 5*, 543–549. Oxford University Press.

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

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

Evans, M., and T. Swartz. 1995. Methods for approximating integrals in statistics with special emphasis on Bayesian integration problems. *Statistical Science* 10 (3): 254–272.

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

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

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

Herzog, T. N. 2002. Bayesian statistics and the Monte Carlo method. In *Proceedings of the 2002 Winter Simulation Conference*, ed. E. Yücesan, C.-H. Chen, J. Snowdon, and J. Charnes, 136–146. Piscataway, NJ: IEEE, Inc.

Inoue, K., S. E. Chick, and C.-H. Chen. 1999. An empirical evaluation of several methods to select the best system. *ACM Transactions on Modeling and Computer Simulation* 9 (4): 381–407.

Jaynes, E. T. 1983. *E. T. Jaynes: Papers on probability, statistics and statistical physics*. Dordrecht, Holland: D. Reidel Publishing Co.

Jeffreys, H. 1946. An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society London. Series A*. 186:453–461.

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

Kass, R. E. 1989. The geometry of asymptotic inference. *Statistical Science* 4 (3): 188–234.

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

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

Kurowicka, D., and R. M. Cooke. 2002. The vine copula method for representing high dimensional dependent distributions: Application to continuous belief nets. In *Proceedings of the 2002 Winter Simulation Conference*, ed. E. Yücesan, C.-H. Chen, J. Snowdon, and J. Charnes, 270–278. Piscataway, NJ: IEEE, Inc.

Laplace, P. S. 1812. *Theorie analytique des probabilites*. Paris: Courcier.

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

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

Lindley, D. 1957. A statistical paradox. *Biometrika* 44:187–192.

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

McCabe, B. 2003. Monte Carlo simulation for schedule risks. In *Proceedings of the 2003 Winter Simulation Conference*, ed. S. Chick, P. Sánchez, D. Ferrin, and D. Morrice, 1561–1565. Piscataway, NJ: IEEE, Inc.

Merrick, J. R., V. Dinesh, A. Singh, J. R. van Dorp, and T. A. Mazzuchi. 2003. Propogation of uncertainty in simulation-based maritime risk assessment model utilizing Bayesian simulation techniques. In *Proceedings of the 2003 Winter Simulation Conference*, ed. S. Chick, P. Sánchez, D. Ferrin, and D. Morrice, 449–455. Piscataway, NJ: IEEE, Inc.

Muñoz, D. F. 2003. A Bayesian framework for modeling demand in supply chain simulation experiments. In *Proceedings of the 2003 Winter Simulation Conference*, ed. S. Chick, P. Sánchez, D. Ferrin, and D. Morrice, 1319–1325. Piscataway, NJ: IEEE, Inc.

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

Ng, S.-H., and S. E. Chick. 2001. Reducing input distribution uncertainty for simulations. In *Proceedings of the 2001 Winter Simulation Conference*, ed. B. Peters, J. Smith, M. Rohrer, and D. Madeiros, 364–371. Piscataway, NJ: IEEE, Inc.

Ng, S.-H., and S. Chick. 2002. Joint criterion for factor identification and parameter estimation. In *Proceedings of the 2002 Winter Simulation Conference*, ed. E. Yücesan, C.-H. Chen, J. Snowdon, and J. Charnes, 400–406. Piscataway, NJ: IEEE, Inc.

O'Hagan, A., M. Kennedy, and J. Oakley. 1999. Uncertainty analysis and other inference tools for complex computer codes (with discussion). In *Bayesian Statistics 6*, 503–524. Oxford University Press.

Popova, E., and D. Morton. 1998. Adaptive stochastic manpower scheduling. In *Proceedings of the 1998 Winter Simulation Conference*, ed. D. J. Madeiros, E. F. Watson, J. S. Carson, and M. S. Manivannan, 661–668. Piscataway, NJ: IEEE, Inc.

Porteus, E. L. 2002. *Stochastic inventory theory*. Stanford Univ. Press.

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

Sacks, J., W. Welch, T. Mitchell, and H. Wynn. 1989. Design and analysis of computer experiments. *Statistical Science* 4:409–435.

Santner, T. J., B. J. Williams, and W. I. Notz. 2003. *The design and analysis of computer experiments*. New York: Springer.

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

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

Smith, A. F. M., and I. Verdinelli. 1980. A note on Bayesian design for inference using hierarchical linear model. *Biometrika* 67 (3): 613–619.

Smith, B. J. 2004. Bayesian output analysis program (BOA) version 1.1 user's manual. Technical report, University of Iowa, Available online via `http://www.public-health.uiowa.edu/boa/` (accessed May 12, 2004).

Spiegelhalter, D. J., N. G. Best, W. R. Gilks, and H. Inskip. 1996. Hepatitis B: a case study in MCMC methods. In *Markov Chain Monte Carlo in Practice*, ed. W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. London: Chapman and Hall.

Steckley, S. G., and S. G. Henderson. 2003. A kernel approach to estimating the density of a conditional expectation. In *Proceedings of the 2003 Winter Simulation Conference*, ed. S. Chick, P. Sánchez, D. Ferrin, and D. Morrice, 383–391. Piscataway, NJ: IEEE, Inc.

Tanner, M. A. 1996. *Tools for statistical inference: Methods for exploration of posterior distributions and likelihood functions*. 3rd ed. Springer.

van Beers, W., and J. P. C. Kleijnen. 2003. Kriging for interpolation in random simulation. *Journal of the Operational Research Society* 54:255–262.

van Beers, W. C., and J. P. C. Kleijnen. 2004. Kriging interpolation in simulation: A survey. In *Proceedings of the 2004 Winter Simulation Conference*, ed. R. Ingalls, M. Rossetti, J. Smith, and B. Peters. Piscataway, NJ: IEEE, Inc.

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

Winston, W. 2000. *Financial models using simulation and optimization*. 2nd ed. Palisade Corp.

Zouaoui, F., and J. R. Wilson. 2001. Accounting for input model and parameter uncertainty in simulation input modeling. In *Proceedings of the 2001 Winter Simulation Conference*, ed. B. Peters, J. Smith, M. Rohrer, and D. Madeiros, 290–299. Piscataway, NJ: IEEE, Inc.

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

AUTHOR BIOGRAPHY

STEPHEN E. CHICK is Associate Professor of Technology Management at INSEAD in Fontainebleau, France. In addition to stochastic simulation, his research focuses on Bayesian statistics, operations management, public health, decision analysis and computational methods in statistics. He is VP of the INFORMS College on Simulation, and is part of the editorial teams of *Management Science*, *Naval Research Logistics*, and *Health Care Management Science*. His e-mail is `<stephen.chick@insead.edu>`.