A Bayesian Account of Uncertainty For Discrete-Event Dynamic Simulation:

Selection of Input Distributions

Stephen E. Chick
Department of Industrial & Operations Engineering
University of Michigan
Ann Arbor, MI 48109

Technical Report 96-14

October 1996

UM IOE TR-96-14

A Bayesian Account of Uncertainty For Discrete-Event Dynamic Simulation:

Selection of Input Distributions

Stephen E. Chick*

October 23, 1996

Abstract

The main thrust of this paper is to address problematic issues in the area of selecting statistical inputs to discrete-event dynamic simulation. A Bayesian framework is proposed which formalizes a link between the subjectivist and parametric frequentist techniques currently used in practice, and resolves known difficulties with each of those techniques. This paper discusses the determination of the probability that a distribution best represents the data; an algorithm to select input distributions and parameters for simulation replications which fully accounts for both parameter and distribution uncertainty; a mechanism that accounts for mixture distributions; and extensions of Latin hypercube sampling for variance reduction and sensitivity analysis when distributions and parameters are unknown. An extensive list of references to related literature is given, and approximation methods to aid implementation are presented.

Keywords: Discrete-event dynamic simulation, simulation input modeling, Bayesian methods, Bayesian statistics, mixture models, decision analysis, sensitivity analysis, variance reduction, Monte Carlo simulation, Latin hypercube sampling.

Contents

*Stephen.E.Chick@umich.edu, Department of Industrial and Operations Engineering, The University of Michigan, 1205 Beal Avenue, Ann Arbor, MI 48109-2117.

1 Introduction

The simulation of discrete-event dynamic systems is a widely-used operations research tool with applications to manufacturing, service, and production systems [2,34,49]. A major attraction of simulation is its ability to provide insights into system operation in the face of complex stochastic behavior.

A central problem in the design of stochastic dynamic simulations is the selection of appropriate input distributions to characterize the stochastic behavior of the modeled system [58,59]. Failure to select appropriate input distributions can lead to misleading simulation output, and therefore to poor system design decisions.

This paper addresses the issue of selecting input distributions and parameters for input to dynamic simulations. There are two widely-used techniques found in the literature and industrial practice: subjective specification of an input distribution, and classical statistical techniques based on parameter estimation and goodness-of-fit. Sec. 1.2 describes these techniques and argues that they can lead to unjustified conclusions. Another approach based on Bayesian statistics is presented in Sec. 2 which addresses known problems with

the existing techniques.

The Bayesian formulation leads directly to an algorithm for selecting input distributions and parameters for simulation replications, as described in Sec. 3. A key insight is that input distributions and parameters should be selected randomly (with an appropriate distribution) from replication to replication in order to account for both distribution and parameter uncertainty. For simulation, then, input distributions and parameters are chosen by sampling, rather than by the more common approach of selecting according to a point estimation rule (such as maximum likelihood, or Bayesian scoring rules). One benefit of this sampling is that simulation output for multiple replications will quantitatively incorporate the uncertainty in input distributions and parameters. At present, this uncertainty is typically ignored or qualitatively incorporated at best.

Modeling issues for input distributions are addressed in Sec. 5, with particular attention to the problem of mixture distributions. Implementation issues are addressed in Sec. 4, where it is noted that although there are no widely available user-friendly software packages which automate Bayesian distribution and parameter selection for simulation practitioners, there are a number of techniques and approximations available for custom applications. New results for variance reduction and sensitivity analysis issues are found in Sec. 7. The focus is on extending Latin hypercube sampling to account for distribution uncertainty and on sensitivity to initial (prior) probability assumptions. A comparison of Bayesian technique proposed here with existing approaches to input distribution selection is presented in Sec. 8. Sec. 9 summarizes the conclusions and indicates directions for further research.

1.1 Related literature for Bayesian techniques

Bayesian techniques have been used in a number of disciplines for statistical distribution selection, including statistics, econometrics, artificial intelligence, sociology and medicine (see e.g. [21, 26, 27, 29, 38, 44, 57] and references therein). Emphasis is often placed on improving the efficiencies of Monte-Carlo (or static) simulation for calculating the expected values of functions of the unknown parameters, as well as the efficient computer generation of random variables with a given posterior distribution [10,19,22,53]. Those techniques will prove useful for inplementing the discrete-event simulation replication algorithm presented in Sec. 3.

Bayesian techniques have not yet been formally employed nor developed in the field of discrete-event dynamic simulations, with the notable exception of [1], where the emphasis was on the Bayesian analysis of simulation output, rather than on modeling input distributions for simulation.

1.2 Critique of existing methodology

There are two widely espoused methodologies for selecting statistical input distributions in the discrete-event dynamic simulation community. They are (1) classical and (2) subjectivist statistics (see, e.g. [2,34,49]).

When data is available, classical techniques are generally employed, including parameter

estimation (MLE, Method of Moments, least-squares,...) in combination with goodness-of-fit tests (χ^2, Kolmogorov-Smirnov, Anderson-Darling,...). Indeed, many widely used simulation support software packages implement classical techniques [25,41,56]. In spite of this widely accepted literature and broadly implemented technology, statistical distribution selection remains a contested topic. Some problematic issues include:

1. Goodness-of-fit and P-value criteria are difficult to interpret and inconclusive at best, and misleading at worst [6,7].
2. The use of a single distribution and single parameter value under-estimates the uncertainty in the both parameter values and the distribution model [21].
3. In practice, both human and automated processes are too complex to be described by a single likelihood distribution and a fixed parameter value.
4. The lack of engineering motivation for most specific distributions.

Among the idiosyncrasies of classical goodness of fit and/or P-value type hypothesis tests are: with few data points, few distributions are rejected; with many data points, all distributions are rejected [44]; there is no coherent method for selecting among non-rejected distributions; and classical techniques can actually reject the (a posteriori) most probable distribution [6,36].

A subjectivist approach is often advocated in the discrete-event simulation literature, particularly (but not exclusively) when little or no data is available [2,34,49,58]. Equipment specifications, expert opinion, physical limitations, and the nature of the process are used to specify means, medians, modes, quantiles, or other relevant quantities, and a single distribution is postulated, such as the triangular, exponential, truncated normal, or Bézier. Alternatively, histograms are 'smoothed' to resemble a data set, with the resulting non-parametric density used as input to the simulation. One problem with this particular subjective approach is that a single, specific statistical distribution is chosen. As with the classical approach, this tends to underestimate the uncertainty in the random process generating the data. Extensive sensitivity analysis has been suggested as a means of compensating for this problem [2]. At the same time, there has been little research describing the best way to do this sensitivity analysis.

Two other approaches for selecting inputs to a discrete-event dynamic simulation are (1) the use of raw historical data and (2) empirical distributions. Law and Kelton [34] argue that for characterizing the stochastic behavior of a system, those approaches have several disadvantages. They are therefore not considered in this paper.

2 Bayesian Problem Formulation

This section presents a Bayesian formulation for evaluating statistical distributions which explicitly addresses the first three of the above criticisms of classical techniques. The formulation is also sufficiently general for selecting from specific distributions that are suggested

by engineering knowledge, although the current work does not discuss how to determine engineering-based statistical distributions.[1] The formulation generalizes existing approaches for specifying distributions subjectively by permitting distribution and parameter uncertainty and allowing an automatic 'updating' mechanism, should additional data become available.

Suppose that a statistical distribution and parameter for a sequence of random quantities $X_1, X_2, \ldots$ is needed for input to a discrete-event simulation of a dynamic system. Also suppose that Y is a relevant figure of merit (e.g. throughput, cycle time, work-in-process) for the decision-making process. Since Y depends the random inputs X_i, Y is a random variable which is a function of the X_i as well as the distribution and parameters that describe the X_i.

One Bayesian approach to the problem of distribution selection supposes that random variables X_i are exchangeable, or conditionally independent, given the distribution and parameter. A collection of M distributions are chosen as candidates for describing the data, where distribution j has parameter θ_j. M is assumed finite and θ_j either continuous or a combination of discrete and continuous (although these assumptions can be weakened).

Under these conditions, the marginal probability distribution for $\vec{x}_n = (x_1, \ldots, x_n)$ can be written using a deFinetti-type theorem (see e.g. [18]).

$$\begin{aligned} p(\vec{x}_n) &= \sum_{j=1}^{M} \int_{\Theta_j} p(\vec{x}_n \mid j, \theta_j)\pi(\theta_j \mid j)\pi(j)d\theta_j \\ &= \sum_{j=1}^{M} \pi(j) \int_{\Theta_j} \pi(\theta_j \mid j) \prod_{i=1}^{n} p(x_i \mid j, \theta_j)d\theta_j \end{aligned} \tag{1}$$

where $\pi(j)$ is the probability assigned to the j-th distribution, $\pi(\theta_j \mid j)$ is the probability assigned to θ_j given distribution j, and $p(x_i \mid j, \theta_j)$ is the probability distribution of x_i given j, θ_j.

In accordance with Bayesian tradition, we name $\pi(j)$ the *prior* probability that distribution j is the correct distribution, and use $\pi(\theta_j \mid j)$ to model the prior probability that the parameter $\Theta_j = \theta_j$, given that j is the correct distribution, $j = 1, \ldots, M$. A discussion of the specification of these prior probabilities is found in Sec. 4.1.

It is useful to define some conditional distributions: the probability $p(\vec{x}_n \mid j)$ of $\vec{x}_n$ given the distribution j; the probability $p(j \mid \vec{x}_n)$ of a model j given $\vec{x}_n$; the posterior probability $p(\theta_j \mid j, \vec{x}_n)$ of parameters θ_j, given j and $\vec{x}_n$; and the 'predictive' distribution $p(Y \mid \vec{x}_n)$ of a figure of merit Y, given $\vec{x}_n$.

$$p(\vec{x}_n \mid j) = \int_{\Theta_j} p(\vec{x}_n \mid j, \theta_j)\pi(\theta_j \mid j)d\theta_j \tag{2}$$

$$p(j \mid \vec{x}_n) = \frac{p(\vec{x}_n \mid j)\pi(j)}{\sum_{k=1}^{M} p(\vec{x}_n \mid k)\pi(k)} \tag{3}$$

[1] For the case of component time-to-failure, the interested reader is referred to the growing operational Bayes and probabilistic physics-of-failure literature [3, 14, 15, 40, 50, 55].

$$p(\theta_j \mid j, \vec{x}_n) = \frac{p(\vec{x}_n \mid j, \theta_j)\pi(\theta_j \mid j)}{p(\vec{x}_n \mid j)} \quad (4)$$

$$p(Y \mid \vec{x}_n) = \sum_{j=1}^{M} p(j \mid \vec{x}_n) \int_{\Theta_j} p(Y \mid j, \theta_j) p(\theta_j \mid j, \vec{x}_n) d\theta_j \quad (5)$$

The final equation is known as a Bayesian model average.

Suppose that $\vec{x}_n$ represents observed data from the sequence $X_1, \ldots$. Eq. 2-Eq. 5 can then be used to update prior beliefs about uncertainty in distributions and parameters. Further, the concerns raised in Sec. 1.2 are addressed in the following ways.

1. The belief that a given distribution and parameter is correct is directly interpretable in terms of the probabilities in Eq. 3 and Eq. 4 [6,7,36,44].
2. The effect of both distribution and parameter uncertainty on the system output are described by the formalism of Eq. 5 [21].
3. There is a capability to determine whether 'mixtures' of distributions and/or parameters best describe the data. (See Sec. 5).
4. Explicit use of judgment from an engineer or factory floor worker is available in the form specifying which distributions are possible, as well as $\pi(j)$ and $\pi(\theta_j \mid j)$.

3 Simulation Replication Algorithm

Turn now to the problem of determining a statistical distribution and parameters for a input to a simulation used for estimating a figure of merit Y. For simplicity, focus on the distribution and parameter selection process for a single random effect (say, time to failure of polishing machine 5). Suppose that historical data $D = \vec{x}_n$ has been collected, and that R simulation replications are run, with each simulation replication requiring S samples. We are therefore required to generate random variables $x_{k,i}$, representing the i-th sample for the k-th simulation replication, $k = 1, \ldots, R; i = 1, \ldots, S$.

Discrete-event simulation accounts for randomness by sampling from distributions, and the case of selecting an input distribution and parameters should be no different. Eq. 5 quantifies the uncertainty in input distribution and parameter, and therefore suggests that the distribution j and parameter θ_j for replication j should be independently sampled from Eq. 3 and Eq. 4.

Fig. 1 summarizes an algorithm which implements this methodology for selecting input distributions and parameters. The net effect of the algorithm is to account for the full extent of the uncertainty in the input parameters, while fully accounting for the data D. Although this idea of sampling distributions and models has recently appeared in the Monte-Carlo Markov chain literature, the idea seems to be completely new in the discrete-event simulation literature.

The output Y_k of simulation run k is an exchangeable sample from the marginal distribution $p(Y \mid D)$, conditional exclusively on the data. In other words, the algorithm accounts for

for $k = 1, \ldots, R$ replications
    sample the k-th distribution j_k from $p(j \mid D)$
    sample the k-th parameter θ_{j_k} from $p(\theta_j \mid j_k, D)$
    for $i = 1, \ldots, S$ service times
        sample service times $x_{k,i}$ from $p(x \mid j_k, \theta_{j_k})$
    end loop
    report figure of merit Y_k from replication k
end loop
generate output (e.g., estimate $E_{Y|D}[Y \mid D]$ as $\hat{Y} = \sum_{k=1}^{R} Y_k/R$)

Figure 1: Algorithm for running simulation replications by randomly sampling the input distribution and parameters.

the information provided by the historical data D, and averages over the uncertainty in the distribution and parameter for input X.

This differs significantly from common practice in discrete-event simulation. First, common practice uses the same input distribution and parameter for every replication (except for sensitivity analysis), rather than sampling those inputs. Second, the output has distribution $p\left(Y \mid j^*, \theta_j^*\right)$, where the distribution j^* and parameter θ_j^* in the condition are determined by parameter estimation and goodness-of-fit tests, and the probability that the right distribution and parameter was selected, $p\left(j^*, \theta_j^* \mid D\right)$ is unknown.

The algorithm in Fig. 1 is a naive approach to selecting parameters for replications to estimate functions of Y, and does not provide for variance reduction. Sec. 7.1 discusses an alternate way to select parameters which reduces variance. Further differences between the Bayesian and classical approaches are explored in Sec. 8.1.

4 Implementation of Bayesian Approach

Two difficulties with the Bayesian framework are the specification of suitable prior distributions, and numerical integration. This section is devoted to discussing these issues.

4.1 Evaluation of Prior Distributions

One of the primary areas of contention with Bayesian problem formulations in general is in the specification of prior distributions.

Subjectivism [48] is currently the dominant philosophical foundation for Bayesian statistics (in writing if not in practice). That is, a decision-maker's knowledge is used to specify the M structural models for input distributions and corresponding prior distributions. At least two problems arise in practice with this formulation. One is the potential non-interpretability of the distribution's parameters. (How can one operationally measure the shape parameter of a tool with Weibull-distributed time-to-failure?)

Even if this first problem is solved, a second problem arises. The introspective assessment of the prior probabilities may be extremely difficult. Methods of simplifying this difficulty have been addressed extensively, with focus on the search for a 'default', 'objective', or 'reference' prior. Some of these methods include Jeffrey's invariant prior, maximum entropy techniques proposed due to Jaynes, and optimization techniques for information maximization due to Zellner. Kass and Wasserman [32] present a comprehensive review of these and other techniques for automating the selection of a prior. One conclusion is that all approaches to automating prior selection have at least some pitfalls. For instance, many techniques give rise to an improper prior or require an arbitrary selection of compact subset of the parameter space in order to obtain a proper prior. (A proper prior integrates to 1.) While an improper prior can give rise to point estimates for parameters for a given distribution, a proper prior is required to take advantage of the full inferential power of Eq. 2-Eq. 5.

This paper uses the subjective approach, and compares results to the analagous classical results whenever possible. (Some results are not obtainable with classical techniques.)

For the case of mixture distributions (see Sec. 5), the paper follows the current practice [5,20] of evaluating priors for each of the components and mixing proportions separately, and multiplying as if they were statistically independent.

4.2 Practicalities of Calculation

The calculation and approximation of the marginal distributions $p(j \mid D)$ and $p(\theta j \mid j, D)$ has been analyzed extensively in the literature. A recent survey paper of Evans and Swartz [22] indicates that significant progress has been made using five general techniques: asymptotic methods, importance sampling, adaptive importance sampling, multiple quadrature, and Markov chain methods [12]. At present, however, there does not appear to be a widely available software package to automate the calculation of Eq. 2-Eq. 5 that also meets the level of user-friendliness of packages currently used in the discrete-event simulation community (see, e.g. [25,41,56]).

Still, some software is available as add-ons to existing statistical packages to support Bayesian inference. Raftery [45] has developed S software code for a number of applications (e.g. linear regression) which is available on the Internet (at `http://lib.stat.cmu.edu/`). Clyde [16] describes a computer tool for Bayesian design of experiments, which has been described as powerful but has a non-trivial learning curve [11]. A public-domain tool described by Wagner and Wilson [58] aids users in subjectively specifying bivariate input distributions, as well as finding classically best-fit Bézier distributions. The tool does not support Bayesian inference.

Markov chain Monte Carlo (MCMC) methods [10,22,23,53] have been used not only for approximating the required integrals, but also for generating random variables $p(j, \theta_j \mid D)$. Thus, samples from MCMC simulations can generate statistical distributions and parameters for use in driving discrete-event simulations. Until a widely-available user-friendly software package is available for assisting the discrete-event simulation modeler, custom programming will generally be required.

The following approximation techniques may simplify calculations. A comparison of various

techniques for approximating Eq. 3 and Eq. 4 are presented in [19]

4.2.1 Laplace Approximation and Schwartz Criterion

Assuming the posterior distribution $p(\theta j \mid j, D)$ is proper and unimodal, the Laplace approximation [33] can be employed:

$$p(j \mid D) \approx \frac{(2\pi)^{d_j/2} \cdot \left|\tilde{\Sigma}_j\right|^{1/2} \cdot p\left(D \mid j, \tilde{\theta}_j\right) \cdot \pi\left(\tilde{\theta}_j \mid j\right)}{\sum_{i=1}^{M} p(i \mid D)} + O(n^{-1}) \tag{6}$$

where the dimension of θ_j is d_j, $\tilde{\Sigma}_j = \left(-\mathbf{D^2} \log p\left(\tilde{\theta}_j \mid j, D\right)\right)^{-1}$ is the inverse of the Hessian of the log-posterior evaluated at the maximum a posteriori (MAP) estimate $\tilde{\theta}_j$ of θ_j given j, and n is the number of data points. The Laplace approximation is based on a normality assumption [28], and improves when the third derivatives of the log-posterior approach 0 [30].

A variation of this approximation is to use the MLE $\hat{\theta}_j$ and observed information matrix $\hat{\Sigma}_j = \left(-\mathbf{D^2} \ell_j(\theta_j; D)\right)^{-1} \mid_{\hat{\theta}_j}$ in place of the Bayesian analog $\tilde{\theta}_j, \tilde{\Sigma}_j$ and ignore the prior distributions. A further variation is to use the expected information matrix. This use of $\hat{\Sigma}_j$ differs in character from its typical application in simulation—that of estimating confidence intervals for parameters [35]. Here, $\hat{\theta}_j$ and $\hat{\Sigma} j$ are used to estimate the probability that the distribution is correct, as well as estimating the probability that a particular parameter is the correct parameter, via approximation of Eq. 4 by a normal distribution.

The Schwartz criterion (or BIC within a factor of 2) [33] is a simpler but less accurate approximation to the relative likelihood of two distributions.

$$\log\left(\frac{p(D \mid j)}{p(D \mid k)}\right) \approx \log p\left(D \mid \hat{\theta}_j, j\right) - \log p\left(D \mid \hat{\theta}_k, k\right) - \frac{d_j - d_k}{2} \log n + O(n^{-1/2}) \tag{7}$$

where large positive values provide relative support for distribution j over distribution k [44]. The Schwartz criterion ignores prior information. The factor $(d_j - d_k) 2 \log n$ can be thought of as a penalty for models with a large number of parameters.

Another potential problem with calculating Eq. 3 is that M may be extremely large for some applications. Although this is not the case in the current paper, Madigan and Raftery [37] propose a technique called Occam's window to include only the most likely and simple distributions in the analysis.

A primary advantage of the Laplace and Schwartz criteria approximations is that the posterior distributions of the θ_j need not be computed. This significantly reduces the computational burden imposed by high-dimensional numerical integration. The calculation of Hessians and point estimates (MAP) is still required.

4.2.2 Selection for a single replication

Suppose that only a single replication may be selected due to cost considerations. The replication algorithm in Fig. 1 may select an 'unlikely' distribution and parameter. For a

single replication, it may be advantageous to select a 'typical' distribution and parameter rather than to randomly sample.

The use of scoring rules which choose (j^*, θ_j^*) to maximize the posterior density $p(j, \theta_j \mid D)$ is inconclusive, because the densities are coordinate-dependent (by changing the parameterization of the θ_j, one could make *any* distribution with continuous parameters have the maximal posterior density, assuming the prior density was not 0). This problem is avoided in part by breaking the problem into two parts: selecting the distribution first, then the parameter.

Using the log scoring rule (see, e.g. [8]) for the *distribution*, the best choice j^* maximizes

$$\log p(j, D) = \log \pi(j) + \log p(D \mid j) \tag{8}$$

This formulation does not depend on the coordinate systems used to describe the parameters θ_j, since the parameter is integrated out of $p(D \mid j)$ via Eq. 2. Heckerman and Chickering [26] called this procedure for distribution selection the *scientific criterion.*

For a given distribution j^*, Berger [4] suggests two potential choices for θ_j^*. One is the MAP estimate $\tilde{\theta}_j$, when the maximum is unique. The other is the posterior mean $E[\theta_j \mid j, D]$, when the expectation exists. The latter has the advantage of minimizing posterior variance, but suffers in that parameter averaging may be meaningless for a given problem. Both estimators have been used in practice. Either choice of estimator gives asymptotically correct results as the number of data points in D becomes infinite, conditioned on the assumptions that the correct distribution is within the hypothesized set, that the prior probability density of the correct parameter is non-zero, and regularity conditions for the MLE to be strongly consistent are satisfied.

5 Mixtures Models and Nested Models

One common statistical assumption is that there is a single 'true' likelihood distribution $\hat{j}$ and parameter $\hat{\theta}_j$, as in interpretation I.1.

I.1. Data are described by a unique likelihood model and unique parameter.

For many applications, this assumption is not appropriate. For instance, service times, inter-arrival times, equipment time-to-failure and repair times may have several influencing factors, each of which has a differing underlying statistical distributions. For this reason, the following alternate interpretations are given.

I.2. Data are described by a mixture of likelihood models, but each likelihood model has a unique parameter.

I.3. Data are described by a unique likelihood model, with a mixture of parameters.

I.4. Data are described by a mixture of likelihood models, each with a mixture of parameters.

Interpretations I.2.-I.4. can be incorporated into the framework developed above by modifying what is meant by choosing the M distributions to represent not only likelihood models but also mixtures of likelihood models [21]. A distribution is not restricted to being a single likelihood model, but can be a mixture of likelihood models to reflect a theory of how the data were generated. Each candidate structure for a mixture distribution is counted as one of the M competing alternatives. Titterington et al. [54] present a review of the field of mixtures. Richardson and Green [46] present a recent Bayesian treatment of mixtures of normal distributions.

Although mixture distributions can provide better 'fit' to data because they have additional parameters, it is not the case that a mixture will always be more likely than one of its components, given the data. The asymptotic approximations in Sec. 4.2.1 give a rough idea of how more complex distributions are penalized (in some sense) relative to simple distributions. Model parsimony is therefore preserved.

A common formulation for a finite mixture distribution of c components is to assume that

$$p(x_i \mid j, \theta_j) = \sum_{l=1}^{c} p_{jl} f_{jl}(x_i \mid \theta_{jl}) dx_i \tag{9}$$

where p_{jl} is the proportion of samples from the l-th component of the mixture, with $\sum_{l=1}^{c} p_{jl} = 1$, and the densities f_{jl} have a specified parametric form with parameter θ_{jl}. The parameter for the entire mixture distribution is therefore $\theta_j = (p_{jl}, \theta_{jl})_{l=1}^{c}$.

Mixture distributions with a differing or unknown number of components have been represented either as competing distribution alternatives, or as a single model structure (with the number of components represented as an additional discrete parameter). Models with $c-1$ components are nested in models with c components (set $p_{jl} = 0$ for some l). If a $(c-1)$-component model is believed to be a possible theory for explaining the randomness in the data, it must be given non-zero probability (not just a non-zero density) to enable the selection of the $(c-1)$-component model by the Bayesian formalism described above.

Similarly, non-mixture distributions which are nested should be considered separate distributions (e.g. the exponential distribution nested inside the Weibull). The exclusion of the nested distribution (e.g. not including the exponential because the Weibull is included, even if memorylessness is considered possible) leads to the exclusion of the nested distribution as a possibility, as well as strange theories about how things work. Take an example from physics to clarify: 'Is the potential energy in a spring kx^c (fit two parameters, as in Weibull), or is it kx^2 (fit one parameter, as in exponential)?' The one-parameter model is nested in the two-parameter model. Excluding the one-parameter model from consideration would lead to laws of physics such as $E = kx^{1.9993}$, with updates for the exponent each time new data is collected. Including both models for comparison would enable the determination of whether a one-parameter or two-parameter model is best supported by the data.

6 A Mixture-Model Example

The time-to-failure for a manufacturing process is to be modeled. For the purposes of this paper, the time to failure is initially anticipated to be on the order of four hours. There are

$M = 3$ distributions under consideration, one of which has a mixture of parameters. There is nesting.

Model 1 The X_i are distributed exponentially. We parameterize with the mean, $\Theta_1 = m$, so

$$f_1(x \mid m) = \frac{1}{m} e^{-x/m}.$$

Model 2 The X_i have a gamma distribution. The parameter $\Theta_2 = (\alpha, m)$ is two dimensional, with

$$f_2(x \mid \alpha, m) = \frac{x^{\alpha-1}}{m^\alpha \Gamma(\alpha)} e^{-x/m}.$$

Model 3 The X_i are a mixture of two gamma distributions with an unknown mixing ratio. The parameter Θ_3 is five dimensional, and can be parametrized by $(p, \alpha_1, m_1, \alpha_2, m_2)$, where p is the unknown mixing proportion.

$$f_3(x \mid p, \alpha_1, m_1, \alpha_2, m_2) = p f_2(x \mid \alpha_1, m_1) + (1-p) f_2(x \mid \alpha_2, m_2).$$

We assume that $0 < m, \alpha; p \in (0,1)$. The selection of coordinates for the gamma was made so that when $\alpha = 1$, the parameterization is consistent with the parametrization for the exponential.

The priors probabilities chosen to reflect initial beliefs are:

$$\pi(m \mid 1) = \frac{500 dm}{m^2 e^{500/m}}, \tag{10}$$

$$\pi(\alpha, m \mid 2) = \frac{500 dm}{m^2 e^{500/m}} \frac{d\alpha}{20}, \tag{11}$$

$$\pi(p, \alpha_1, m_1, \alpha_1, m_2 \mid 3) = \frac{8 p^{1/2}(1-p)^{1/2} dp}{\pi} \frac{55 dm_1}{m_1^2 e^{55/m_1}} \frac{d\alpha_1}{20} \frac{945 dm_2}{m_2^2 e^{945/m_2}} \frac{d\alpha_2}{20}, \tag{12}$$

$$\pi(1) = \pi(2) = \pi(3) = 1/3 \tag{13}$$

where the range of the parameters is again assumed to satisfy $0 < \alpha_i < 20, 0 < m_i$.

The priors have been chosen to be relatively diffuse. For the exponential, the prior takes a maximum at $\theta_1 = 250$ (using this value gives rise to an exponential with mean 250). The expected value of the parameter, does not exist due to the diffuseness of the prior. For the gamma, the scale parameter was assigned the same distribution, and the shape parameter is chosen to be an independent uniform[0,20]. For the mixture of gammas, the mixing proportion was chosen to have a (mildly) non-uniform density to favor mixtures ($p \in (0,1)$) over non-mixtures ($p = 0$ or $p = 1$). The first of the gammas in the mixture was chosen to represent a failure mechanism which is believed to give short times-to-failure (m_1 maximized at 55). The second gamma represents another mechanism which is believed to permit longer operation times (m_2 maximized at 945).

The data D of Fig. 2, which is typical for machine times-to-failure in a manufacturing environment, was then used to update the distributions for these competing distributions. Results are summarized in Fig. 3. Calculations were implemented with Mathematica [60] on

Downtime Data D (rounded, in minutes)					
0.25	0.55	0.67	1.75	2.20	3.00
3.17	3.23	4.40	5.15	8.33	9.23
10.37	10.55	11.25	13.83	23.03	27.20
33.22	35.28	36.02	50.33	69.08	78.15
81.78	89.92	96.95	97.83	103.25	111.80
135.85	187.90	198.91	224.70	262.23	271.18
381.70	962.38	3435.23			

Figure 2: Sorted data D resemble the time-to-failure for a manufacturing process.

Time-To-Failure Inference			
Model	$j=1$	$j=2$	$j=3$
θ_j	m	(α, m)	$(p, \alpha_1, m_1, \alpha_2, m_2)$
$\tilde{\theta}_j$ (MAP)	185	$(.367, 464)$	$(.776, .576, 92.3, .478, 1039)$
$E[\theta_j \mid j, D]$	194	$(.364, 551)$	$(.876, .557, 138, 1.72, 1637)$
$p(j \mid D)$ (quad./MCMC)	1.41×10^{-9}	.164	.836
$p(j \mid D)$ (Laplace)	2.39×10^{-9}	0.267	.733
$p(j \mid D)$ (Schwartz)	8.87×10^{-10}	.605	.395
$\hat{\theta}_j$ (MLE)	182	$(.357, 509)$	$(.946, .524, 140, 1.96, 1065)$
χ^2-statistic (8 bins)	37.1	7.15	5.51
α for $\chi^2_{8-d_j-1}$	1.68×10^{-6}	.209	.064
α for χ^2_{8-1}	4.49×10^{-6}	.413	.598

Figure 3: Model comparisons for exponential ($j = 1$), gamma ($j = 2$), and mixture of two gammas ($j = 3$), given the downtime data D of Fig. 2.

a Pentium-based PC-clone. MAP and MLE estimates are based on numerical maximization techniques, and Laplace and Schwartz estimates were by direct calculation. For $E[\theta_j \mid j, D]$ and $p(j \mid D)$ (quad./MCMC), quadrature was used for the exponential and gamma distributions, and Monte Carlo simulation was used for the mixture of two gammas. The estimate of 0.836 for $p(3 \mid D)$ has a 95% confidence interval of (.813, .853). The χ^2-test statistic was calculated using 8 equiprobable bins, as well as the probability that χ^2_{8-1} and $\chi^2_{8-d_j-1}$ random variables exceed the test statistic.

The posterior distribution indicates that the mixture is the more likely distribution, given the data. The Laplace approximation gives a reasonably good estimate. The Schwartz approximation, while right within an order of magnitude, incorrectly concludes that the gamma is more likely than the mixture of gammas.

The χ^2-test is inconclusive. When parameters are fit with MLE, and the 'null hypothesis' is assumed, it is known [13] that the distribution of χ^2 satistfies $\chi^2_{8-d_j-1} \leq \chi^2 \leq \chi^2_{8-1}$, where $\leq$ is a stochastic ordering. One bound seems to suggest the gamma better represents the data, the other bound seems to suggest the mixture is better. A further inconclusivity of

this χ^2-test is that no discussion is given as to what the χ^2-statistic should be when the null hypothesis does not hold. It can therefore not determine the probability that a given model is correct, given the data D.

All tests and approximations ruled out the exponential distribution.

Model validation is an important step in simulation development, and a cursory validation of these conclusions was undertaken. The suppliers of the data in Fig. 2 were asked if there were two distinct failure processes were known to have accounted for the data, as suggested by the higher probability for the mixture model. The answer was affirmative, in that short times-to-failure associated with a simple equipment malfunction (jamming) caused some downtimes, and long times-to-failure were associated with an absense of jamming, but a presence of other mechanical failures.

7 Sensitivity Analysis and Variance Reduction

In manufacturing practice, the calculation of specific quantities such as the expected performance given the data in Eq. 5 is secondary to the insights gained from observing the sensitivity of system behavior to changes in simulation inputs [2]. A high sensitivity may indicate that alternate system designs should be considered, that additional data collection may be necessary to better characterize input distributions, or that the system behavior may have a wide range of uncertainty.

Because of the cost of running multiple simulation replications, it is desirable to reduce the number of simulation runs required for sensitivity analysis by selecting input distributions strategically to reduce the variance of the output.

This section makes several practical and theoretical comments regarding sensitivity analysis and variance reduction as they relate to input distribution selection. Topics treated are extensions for Latin hypercube, antithetic variates, and sensitivity to prior probability distributions. For reasons of scope, this paper does not consider aspects of sensitivity analysis requiring significant output analysis and modeling or continuous time analysis of the simulated state (e.g. perturbation analysis [47], response-surface methods, and Bayesian design of experiments [11]). Common random number variance reduction techniques are completely compatible with the current presentation in a straightforward way, and are therefore not discussed.

7.1 Extensions For Latin Hypercube Sampling

Latin hypercube sampling [39] for selecting parameters of a given input distribution has been used for both variance reduction [52] and sensitivity analysis [17]. In this section, two Latin hypercube techniques are extended to account for input distribution uncertainty. The first is simpler to implement, and provides variance reduction asymptotically. The second is more difficult to implement, but provides variance reduction even for a finite number of replications.

First, we describe the procedure for generating a Latin hypercube sample of size R as

originally presented by McKay et al. [39]. Each of the R samples is for use in one of R simulation replications. Suppose that there are K statistically independent real-valued parameters ϕ_k to be input to a simulation. Let $F_k(\phi_k)$ be the cumulative distribution function of ϕ_k, and let ϕ_{ik} be the $k-$th input parameter for sample i. Define $M = (m_{ik})$ to be an $R \times K$ matrix, where each column of M is an independent random permutation of $\{1, \ldots, R\}$. Let ζ_{ik} be a collection of uniform$[0,1]$ random variables which are independent of each other and of M. Then set ϕ_{ik} to be

$$\phi_{ik} = F_k^{-1}\left(\frac{m_{ik} - 1 + \zeta_{ik}}{R}\right) \tag{14}$$

The general effect of this sampling scheme is to 'prevent bunching' of samples by insuring that the marginal empirical distributions of the ϕ_k roughly match the F_k.

Direct application of this Latin hypercube technique is not straightforward for the Bayesian framework presented here. This is because one doesn't know what the parameters are, or even how many there are, because of distribution uncertainty.

A simple variation of the Latin hypercube sampling scheme to account for distribution uncertainty is to first sample the distribution, and then sample the parameter given the distribution. In other words, for simulation replication i, select distribution j_i with discrete Latin hypercube sampling according to $p(j \mid D)$. That is, choose j_i with $K = 1$ so that

$$j_i = \min_j p(\{1, \ldots, j\} \mid D) \geq \frac{m_{i1} - 1 + \zeta_{i1}}{R} \tag{15}$$

where ζ_{i1} are uniform$[0,1]$. This determines the number of times R_j distribution j is chosen for replications, $R_j = \sum_{k=1}^{R} \mathbb{I}_{\{j_k\}}\{j\}$, where the indicator function $\mathbb{I}_A\{b\}$ is 1 if $b \in A$, and 0 otherwise. For the R_j replications with distribution j, use Latin hypercube again (now with $K = d_j$) to sample parameters θ_j using the distribution $p(\theta_j \mid j, D)$. In general, the parameters θ_j will not be statistically independent. Stein [52] provides the necessary extension for Latin hypercube to approximate sampling of dependent variables. Further, by analogy with the proof in the appendix of Stein, this sampling scheme provides variance reduction as $R \to \infty$.

It is desirable to guarantee variance reduction for finite R.

A second extension to Latin hypercube sampling is therefore proposed which provides this variance reduction on the condition that the performance measure Y is a monotonic function of the expected value of the random input X. The idea is to sample an expected value for $E[X \mid j, \theta_j]$, then to select a distribution j and parameter θ_j with that mean. More specifically, let $g(j, \theta_j) = E[X \mid j, \theta_j]$ be the conditional mean of X, given (j, θ_j). Eq. 5 determines the posterior marginal density $p(g \mid D)$ of the random mean G given the data D. Set $K = 1$ in the Latin hypercube scheme, define $F_G(g)$ to be the cumulative distribution for G given D, and sample g_i, j_i, θ_{j_i} as follows:

$$g_i = F_G^{-1}(\zeta_{i1}) \tag{16}$$

$$j_i \sim p(j \mid D, E[X \mid D] = g_i) \tag{17}$$

$$\theta_{j_i} \sim p(\theta_{j_i} \mid D, j_i, E[X \mid D] = g_i) \tag{18}$$

where $\sim$ indicates that the left hand side is to be drawn from a sample whose distribution is on the right hand side. Again, j_i, θ_{j_i} can be sampled by MCMC methods or approximation techniques.

Variance reduction for $E[G]$ using this second variation on Latin hypercube, even for finite R, can be proved by first noting that

$$\mathrm{Var}(\bar{G}) = \frac{\mathrm{Var}(G)}{R} + \frac{(R-1)\mathrm{Cov}(g_1, g_2)}{R},$$

where $\bar{G} = \sum_{i=1}^{R} g_i/R$ is an estimate of $E[G] = E[E[X \mid j, \theta_j]]$, the outer expectation taken with respect to $p(j, \theta_j \mid D)$. As long as g is monotonic in its arguments, $\mathrm{Cov}(g_1, g_2)$ is negative [39]. Although g is not monotonic in j, θ_j, it is monotonic in ζ_{i1}, which in turn is a function of j, θ_j. The original argument of McKay et al. [39] is easily modified to show variance reduction for $\bar{G}$ based on the monotonicity of g in ζ_{i1}. Similarly, if a figure of merit Y for the system is a monotonic function of g, variance reduction for Y will be achieved as well. (This covers a wide variety of interesting figures of merit, such as mean work-in-process, mean utilization, and mean throughput capacity.)

The key is that variance reduction for finite R does not require monotonicity of all arguments of g, only that g be monotonic in some parameter $\zeta = F_G$, which is a function of the simulation inputs.

7.2 Antithetic Variates

This section discusses the use of antithetic variates (AV) for the selection of distributions and parameters for use as simulation inputs to simulation replications. AV is a variance reduction technique which pairs simulation runs, where one run takes random variables which have the inverse cumulative distribution of the other. That is, if X_1 is a service time in the one simulation, then the corresponding service time X_1' in the paired simulation satisfies $X_1' = F_X^{-1}(1 - F(X_1))$, where F_X is the cumulative distribution of X. Although variance reduction is not guaranteed with AV, good results are often obtained when there is a monotonic system response to the input random variable [34].

If a figure of merit is monotonic in the expected value of random inputs, $g(j, \theta_j) = E[X \mid j, \theta_j)$, AV can be implemented with some hope of variance reduction even if the distribution is uncertain. (This monotonicity was already used in the second extension of Latin hypercube in Sec. 7.1). Suppose that (j_i, θ_{j_i}) is selected as the input distribution/parameter input using Eq. 16-Eq. 18. The AV input (j_i', θ_{j_i}') for the paired replication can then be sampled as

$$g_i' = F_G^{-1}(1 - F_G(E[X \mid j_i, \theta_{j_i}])) \quad (19)$$

$$j_i' \sim p(j \mid D, E[X \mid D] = g_i') \quad (20)$$

$$\theta_{j_i}' \sim p(\theta_{j_i} \mid D, j_i', E[X \mid D] = g_i') \quad (21)$$

Note that the use of AV for selecting distributions and parameters for simulation can be used with or without AV for random samples for paired discrete-event simulations for a

fixed input distribution and parameter. That is, two levels of pairing may take place: one for selecting input distributions and parameters for a paired replication (Eq. 20-Eq. 21) and one for corresponding samples from a particular distribution (e.g. AV for the individual service times, given a specific input distribution and parameter).

7.3 Sensitivity To Prior Distribution

The simulation output distribution $p(Y \mid D)$ has a distribution which depends on the prior distributions. A number of authors have explored the effect of modifying the prior distribution on the posterior distribution or on functions of the posterior distribution (such as $E[Y \mid D]$) [4,9,27,31,42,51]. This section discusses conditions when sensitivity analysis is particularly easy to implement.

Let $\pi(j), \pi(\theta_j \mid j)$ be the initially chosen priors, and $\pi'_k(j), \pi'_k(\theta_j \mid j)$ be K competing priors for testing sensitivity, $k = 1, \ldots K$. Also, let $y_1, \ldots, y_R$ be the output from the R using $\pi(j), \pi(\theta_j \mid j)$. Let $\bar{Y}'_k$ be the output measure estimate, assuming the use of $\pi'_k(j)$. A formula used in importance sampling [24] gives

$$\bar{Y}'_k = \sum_{i=1}^{R} y_i \frac{p'_k(y_i \mid D)}{p(y_i \mid D)} \tag{22}$$

where $p(y_i \mid D), p'_k(y_i \mid D)$ are the posterior densities using the initial and k-th competing priors, respectively, and the dependence on the data D is noted explicitly.

The calculation of $p'_k(y_i \mid D)$ potentially requires a complete recalculation with Eq. 5 for each k. Each recalculation requires two numerical integrations (in Eq. 2 and Eq. 5), both possibly high-dimensional. This section explores two special cases of sensitivity analysis which permits each of the integrations to be calculated once, rather than $K + 1$ times.

First, suppose that a sensitivity analysis with respect to the prior $\pi(j)$ is desired, but the prior for the parameters remains the same ($\pi(j) \neq \pi'_k(j)$ for some j, k; but $\pi(\theta_j \mid j) = \pi'_k(\theta_j \mid j)$ for all j, k). Recalculation of $p'_k(y_i \mid D)$ in terms of $p(y_i \mid D)$ in this case only requires a few extra additions and multiplications, and does not incur additional expense in numerical integration—the integral of Eq. 5 does not change, only the $p'_k(j \mid D)$ need recalculation. Similarly, the calculation of $p'_k(j \mid D)$ can be calculated by substituting $\pi'_k(j)$ for $\pi(j)$ in Eq. 3. The calculation of the integral for $p(D \mid j)$ was already completed in order to run the initial batch of R simulations.

Alternately, suppose that the posterior distributions $p(Y \mid j, D), p'_k(Y \mid j, D)$ are heavily data dominated, and that $\pi(\theta_j \mid j), \pi'_k(\theta_j \mid j)$ are relatively constant near $\tilde{\theta}_j$. In this case, the integral for $p'_k(Y \mid j, D)$ in Eq. 5 can be approximated as

$$p'_k(Y \mid j, D) \approx \frac{p(Y \mid j, D)\pi'_k\left(\tilde{\theta}_j \mid j\right)}{\pi\left(\tilde{\theta}_j \mid j\right)} \tag{23}$$

The approximation should not be used if the priors are not roughly constant in the chosen coordinate system for θ_j, or if the posteriors are not highly unimodal.

Sensitivity analysis for priors has a computational cost which is a linear function of the number of alternatives K, whether either of the two special cases holds or not. The relative costs between simple multiplication and high-dimensional integration, however, proves to be of practical advantage for situations where either or both of these special cases hold.

8 Discussion

8.1 Comparison to Estimator/Goodness-of-fit Approach

We now compare the above Bayesian formulation with the standard classical approach [2,34, 49] to selecting an input probability distribution. The standard approach was summarized in Sec. 1.2 as (1) determine a set of eligible distributions, (2) establish point estimates for parameters for those distributions, and (3) evaluate the goodness-of-fit for each distribution.

The first step of the classical approach is similar to the present approach–a set of M distributions are selected as possible candidates. Although the classical approach, in theory, allows for mixture distributions, in practice it does not (subjective smoothing of histograms is generally employed). The inclusion of mixture distributions here is included as an initial assumption in response to an explicit need to account for multiple sources of randomness in industrial and other applications.

The second and third steps differ drastically from the present approach. The posterior marginal probability $p(j \mid D)$ for distribution j gives distinct 'weights' for comparing the descriptive value of each distribution. These weights have a probability interpretation. Although the classical goodness-of-fit tests give 'scores', these scores do not indicate the probability that a given distribution is correct (conditional on the correct distribution being among the M distributions proposed). The best one might hope for is to 'reject' as many distributions as possible and leave only a few good 'candidates'. The Bayesian approach does not suffer from either of these extremes.

Another difference between the approach presented here and the use of a point estimate of a parameter (such as MLE) for all simulation runs is that the *uncertainty in the parameter* is not accounted for until the sensitivity analysis stage. Similarly, the χ^2-test is often used to determine a unique 'best' distribution, without describing how much uncertainty there is as to which distribution is 'best'. One would be much more likely to believe simulation results when distributions were updated using 10,000 data points versus 100 points, if the standard approach were used. The point-estimate/goodness-of-fit approach to simulation input and output distributions do not formally capture this effect. The algorithm in Fig. 1 generates simulation output, however, which fully accounts for distribution and parameter uncertainty.

Some authors suggest evaluating the statistical independence of the data, in addition to the three modeling steps above [34]. This evaluation frequently takes the form of visual inspection of scatter diagrams or nonparametric tests. Independence versus correlation can equally well be analyzed as two competing model assumptions, and a Bayesian analysis can be carried through to test the relative probability of independence and correlation. (Technically, the exchangeability assumption may need to be dropped, but the use of Bayes

rule for probability calculation remains.)

8.2 System Stability

Because of the uncertainty in system input distributions, it is quite possible that the system is unstable with probability $p > 0$. By unstable, it is meant that the mean service time is at least as large as the mean inter-arrival time.

This effect should be considered an advantage, rather than a defect, of the current approach. Admittedly, an unstable system causes problems for steady state statistical analysis because queue lengths become unbounded. On the other hand, unstable systems do not pose such analytical difficulties in a practical production sense, as measured by money (or utility). It merely requires that the production process be modified (e.g. purchase additional equipment or schedule some overtime). Changes in production processes have costs, but they are not infinite. The calculation of these costs generally involves terminating simulations rather than steady-state simulations.

9 Conclusions

'Let the data speak for themselves' is an often heard call. This presentation attempts to formalize this statement in the Bayesian framework within the domain of determining input probability distributions for discrete-event simulation of dynamic systems.

Contributions of this paper to achieve that goal include:

- A presentation of a Bayesian formula for statistical distribution and parameter selection for application to the field of dynamic simulations.
- An algorithm for selecting statistical distributions and parameters as inputs to simulation models which fully accounts for both distribution and parameter uncertainty.
- Argumentation that mixture distributions arise as a matter of course in real systems, and that mixture distributions can naturally be included in the Bayesian framework.
- An extension to Latin hypercube sampling for input distribution and parameter selection for use when there is uncertainty in both distribution and parameter. Variance reduction is guaranteed for applications where the figure of merit is a monotonic function of the expected value of the input.
- Approximations for efficiently evaluating the sensitivity of the output to prior distributions.
- An application of MCMC simulation results to the field of discrete-event dynamic simulation.

The Bayesian framework has also been argued to generalize the current use of subjective 'density-drawing' techniques, and extends the classical framework to account for input distribution uncertainty (among other things). In this sense, the framework presented here generalizes the two seemingly different techniques used in current simulation practice.

An additional benefit of mixture modeling as presented here can be reaped in sensitivity analysis. Suppose that downtimes (or arrival times) are represented by mixtures, and that the cost of changing the mixing proportions are known. Additional replications can be made which determine the cost-benefit of changing the mix proportions and its effect on system performance.

Some potential criticisms of this Bayesian formulation for input distribution selection, together with responses, include:

- A Bayesian framework is computationally intensive. Response: A number of approximations have been given which make the approach more tractable. Computing speed has sufficiently increased to implement many calculations that were recently too expensive for a desktop computer.

- Classical techniques can handle mixture distributions, too. Response: Agreed. Their treatment, however, with goodness-of-fit and other standard tools suffers the same problems outlined in Sec. 1.2 for the non-mixture treatment.

- Classical techniques provide similar answers to Bayesian analysis when the posterior is data dominated. Response: Classical techniques can and have been used to provide useful approximations for some Bayesian calculations [43]. They have not provided, however, provided the formalism required for Eq. 5 and the replication algorithm in Fig. 1.

- The single distribution j^* and parameter θ_j^* selected in Sec. 4.2.2 will be different in general than that selected by the Latin hypercube technique of Sec. 7.1, when the number of replications R is 1. Response: The motivation for these two selections is vastly different. Sec. 4.2.2 is motivated by a distribution and parameter choice that 'best' gives rise to typical system behavior. It does so by maximizing probability or minimizing a certain loss. Sec. 7.1 is motivated by sensitivity analysis and variance reduction, neither of which is well served when $R = 1$.

- The prior distributions used are coordinate dependent and ad hoc. Response: *All* prior distributions are subjective, and coordinate dependency of a density function is a less important criterion than the easy specification of a prior for a practicing engineer, at least in the present context.

Much work remains to be done in a number of areas. Theoretical and computational results for general and specific mixture distributions needs further exploration. A comprehensive Bayesian accounting of output data is needed to extend the ideas initially proposed by Andrews et al. [1]. Additional sensitivity analysis techniques, also, remain to be framed in a Bayesian modeling framework. The implementation of user-friendly software tools is also an important step for the widespread use of these techniques. Added work is needed

on methodologies for simple specification of generic prior distributions with emphasis on embedding engineering meaning, rather than abstract statistical meaning. Finally, even more work is required to determine which classes of statistical likelihood models are compatible with the engineering processes they purport to model. It is surprising that more work has been focused on prior distributions than on likelihood distributions in this respect, particularly given the fact that data dominates priors for a wide variety of applications.

Acknowledgments. The author would like to thank Andy Andrews, Shane Henderson, Stephen Pollock, and Bruce Schmeiser for their helpful feedback. The interest of Edward Williams, John Harder, and Hwa-Sung Na, all of Ford Motor Company, is also appreciated.

References

[1] Richard W. Andrews, Thomas J. Schriber, Stephen Roberts, Jerry Banks, and Bruce Schmeiser. A Bayesian batch means methodology for analysis of simulation output. In *Proceedings of the Winter Simulation Conference.* Institute of Electrical and Electronics Engineers, Inc., 1983.

[2] Jerry Banks, John S. Carson, II, and Barry L. Nelson. *Discrete-Event System Simulation.* Prentice-Hall, Inc, Upper Saddle River, NJ, USA, 2nd edition, 1996.

[3] R. E. Barlow and M. B. Mendel. The operational Bayesian approach. In P. R. Freeman and A. F. M. Smith, editors, *Aspects of Uncertainty: A Tribute to D.V. Lindley.* John Wiley and Sons, Chichester, UK, 1994.

[4] James O. Berger. *Statistical Decision Theory and Bayesian Analysis.* Springer-Verlag, New York, 2nd edition, 1985.

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

[6] James O. Berger and M. Delampady. Testing precise hypothesis (with discussion). *Statistical Science*, 2:317–352, 1987.

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

[8] Jose Bernardo. Expected information as expected utility. *Annals of Statistics*, 7:686–690, 1979.

[9] G. C. Canavos. Bayesian estimation: A sensitivity analysis. *Naval Research Logistics Quarterly*, 22:543–552, 1975.

[10] Bradley P. Carlin and Siddhartha Chib. Bayesian model choice via Markov chain monte carlo. *Journal of the Royal Statistical Society, Ser. B*, 57:473–484, 1995.

[11] Kathryn Chaloner and Isabella Verdinelli. Bayesian experimental design: A review. *Statistical Science*, 10(3):273–304, 1995.

[12] Ming-Hui Chen and Bruce Schmeiser. Performance of the Gibbs, hit-and-run, and Metropolis samplers. *Journal of Computational and Graphical Statistics*, 2:251–272, 1993.

[13] H. Chernoff and E. L. Lehmann. The use of maximum likelihood estimates in χ^2 tests for goodness of fit. *Annals of Mathematical Statistics*, 25:579–586, 1954.

[14] Stephen E. Chick and Max B. Mendel. An engineering basis for statistical lifetime models with an application to tribology. *IEEE Transactions on Reliability*, 45(2):208–215, June 1996.

[15] Stephen Eric Chick. *Product and Process Design With Engineering-Based Statistical Models.* PhD thesis, UC Berkeley, Berkeley, CA., May 1995. Dept. of Industrial Engineering and Operations Research.

[16] Merlise A. Clyde. An object-oriented system for Bayesian nonlinear design using xlisp-stat. Technical Report 587, University of Minnesota, 1993. Statistics Department.

[17] Ramesh Dandekar and Nancy Kirkendall. Latin hypercube sampling for sensitivity and uncertainty analysis. In *Proceedings of the Section on Physical and Engineering Sciences.* American Statistical Association, 1993.

[18] Bruno de Finetti. *Theory of Probability, v. 2.* John Wiley & Sons, Inc., New York, 1990.

[19] Thomas J. DiCiccio, Robert E. Kass, Adrian Raftery, and Larry Wasserman. Computing Bayes factors by combining simulation and asymptotic approximations. Technical Report TR-630, Carnegie Mellon University, 1995. Statistics Department.

[20] J. Diebolt and C. Robert. Bayesian estimation of finite mixture distributions, part I: theoretical aspects. Technical Report Rapport tech. #110, Université Paris VI, LSTA, 1990.

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

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

[23] W. R. Gilks, D. J. Spiegelhalter, and S. Richardson. *Practical Markov Chain Monte Carlo.* Chapman and Hall, London, 1996.

[24] Peter W. Glynn and D.L. Iglehart. Importance sampling for stochastic simulations. *Management Science*, 35(11):1367–1392, 1989.

[25] Charles R. Harrell, Robert E. Bateman, Thomas J. Gogg, and Jack R. A. Mott. *System Improvement Using Simulation.* JMI Consulting Group and ProModel Corporation, Orem, Utah, 1993.

[26] David Heckerman and David Maxwell Chickering. A comparison of scientific and engineering criteria for Bayesian model selection. Technical Report Advanced Technology Division, MSR-TR-96-12, Microsoft Corporation, 1996.

[27] Stacy D. Hill and James C. Spall. Sensitivity of a Bayesian analysis to the prior distribution. *IEEE Trans. Systems Man Cybernet.*, 24(2):216–221, 1994.

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

[29] Robert E. Kass and Adrian E. Raftery. Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795, 1995.

[30] Robert E. Kass and Elizabeth H. Slate. Some diagnostics of maximum likelihood and posterior nonnormality. *The Annals of Statistics*, 22(2):668–695, 1994.

[31] Robert E. Kass, Luke Tierney, and Joseph B. Kadane. Approximate methods for assessing influence and sensitivity in Bayesian analysis. *Biometrika*, 76:663–674, 1989.

[32] Robert E. Kass and Larry Wasserman. Formal rules for selecting prior distributions: A review and annotated bibliography. Technical Report TR-583, Carnegie Mellon University, 1993. Statistics Department.

[33] Robert E. Kass and Larry Wasserman. A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *Journal of the American Statistical Association*, 90(431):928–934, 1995.

[34] Averill M. Law and W. David Kelton. *Simulation Modeling & Analysis.* McGraw-Hill, Inc, New York, 2nd edition, 1991.

[35] Lawrence M. Leemis. Input modeling for discrete-event simulation. In C. Alexopoulos, K. Kang, W. R. Lilegdon, and D. Goldsman, editors, *Proceedings of the Winter Simulation Conference*, pages 16–23. Institute of Electrical and Electronics Engineers, Inc., 1995.

[36] David D. Lindley. A statistical paradox. *Biometrika*, 44:187–192, 1957.

[37] David Madigan and Adrian Raftery. Model selection and accounting for model uncertainty in graphical models using Occam's window. *Journal of the American Statistical Association*, 89:1535–1546, 1994.

[38] David Madigan and Jeremy York. Bayesian graphical models for discrete data. *International Statistical Review*, 63(2):215–232, 1995.

[39] M. D. McKay, W. J. Conover, and R. J. Beckman. A comparison of three methods for selecting values of input variables in the analysis of output from computer code. *Technometrics*, 21:239–245, 1979.

[40] William Q. Meeker, Luis A. Escobar, and C. Joseph Lu. Accelerated degradation tests: Modeling and analysis. Technical report, Iowa State University, 1996. Statistics Department.

[41] Palisade Corporation. *BestFit User's Manual.* Newfield, NY, 1995.

[42] Luis Raul Pericchi and Peter Walley. Robust Bayesian credible intervals and prior ignorance. *International Statistical Review*, 59:1– 23, 1991.

[43] Adrian E. Raftery. Approximate Bayes factors and accounting for model uncertainty in generalized linear models. Technical Report 255, University of Washington, 1993. Statistics Department.

[44] Adrian E. Raftery. Bayesian model selection in social research (with discussion by Andrew Gelman & Donald B. Rubin, and Robert M. Hauser, and a rejoinder). In Peter V. Marsden, editor, *Sociological Methodology 1995.* Blackwells, Cambridge, Mass., 1995.

[45] Adrian E. Raftery. Hypothesis testing and model selection via posterior simulation. In W. R. Gilks, D. J. Spiegelhalter, and S. Richardson, editors, *Practical Markov Chain Monte Carlo.* Chapman and Hall, London, 1995.

[46] S. Richardson and P.J. Green. On Bayesian analysis of mixtures with an unknown number of components. Technical report, INSERM, France and University of Bristol, February 1996.

[47] Reuven Y. Rubinstein. *Monte Carlo Optimization, Simulation and Sensitivity of Queueing Networks.* John Wiley & Sons, New York, 1986.

[48] Leonard J. Savage. *The Foundations of Statistics.* Dover Publications, Inc., New York, 1972.

[49] Thomas J. Schriber. *An Introduction to Simulation using GPSS/H.* John Wiley, New York, 1991.

[50] Nozer D. Singpurwalla, Simon P. Wilson, Jr. Edwin R. Fuller, and Keith R. Eberhardt. Statistical aspects of failure processes in ceramics. In Jose Bernardo, A. P. David, and A. F. M. Smith, editors, *Bayesian Statistics 5.* Clarendon Press, Oxford, 1994.

[51] S. Sivaganesan. Sensitivity of some standard Bayesian estimates to prior uncertainty: A comparison. *Journal of Statistical Planning Inference*, 27:85–103, 1991.

[52] Michael Stein. Large sample properties of simulations using latin hypercube samples. *Technometrics*, 29(2):143–151, 1989.

[53] Luke Tierney. Markov chains for exploring posterior distributions. *The Annals of Statistics*, 22(4):1701–1762, 1994.

[54] D.M. Titterington, A.F.M. Smith, and U.E. Makov. *Statistical Analysis of Finite Mixture Distributions.* Wiley, Chichester, 1985.

[55] Jan M. van Noortwijk. *Optimal Maintenance Decisions for Hydraulic Structures Under Isotropic Deterioration.* PhD thesis, Delft University of Technology, Holland, May 1996. Dept. of Statistics, Stochastics, and Operations Research.

[56] Stephen Vincent and Averill M. Law. ExpertFit: Total support for simulation input modeling. In C. Alexopoulos, K. Kang, W. R. Lilegdon, and D. Goldsman, editors, *Proceedings of the Winter Simulation Conference*, pages 395–400. Institute of Electrical and Electronics Engineers, Inc., 1995.

[57] Chris T. Volinsky, David Madigan, Adrian E. Raftery, and Richard A. Kronmal. Bayesian model averaging in proportional hazard models: Assessing stroke risk. Technical Report no. 302, University of Washington, Seattle, WA, 1996. Department of Statistics.

[58] Mary Ann Flanigan Wagner and James R. Wilson. Graphical interactive simulation input modeling with bivariate Bézier distributions. *ACM Transactions on Modeling and Computer Simulation*, 5(3):163–189, July 1995.

[59] Edward J. Williams. Downtime data—its collection, analysis, and importance. In J. D. Tew, S. Manivannan, D. A. Sadowski, and A. F. Seila, editors, *Proceedings of the Winter Simulation Conference*, pages 1040–1043. Institute of Electrical and Electronics Engineers, Inc., 1994.

[60] Stephen Wolfram. *Mathematica: A System for Doing Mathematics by Computer.* Addison-Wesley Pub. Co., Inc., Redwood City, CA, 2nd edition, 1991.