Proceedings of the 1999 Winter Simulation Conference
P. A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. W. Evans, eds.

STEPS TO IMPLEMENT BAYESIAN INPUT DISTRIBUTION SELECTION

Stephen E. Chick

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

ABSTRACT

There are known pragmatic and theoretical difficulties associated with some standard approaches for input distribution selection for discrete-event simulations. One difficulty is a systematic underestimate of the variance of the expected simulation output that comes from not knowing the 'true' parameter values. Another is a lack of quantification of the probability that a given distribution is best. Bayesian methods have been proposed as an alternative, but acceptance has not yet been achieved, in part because of increased computational demands, as well as challenges posed by the specification of prior distributions. In this paper, we show that responses to questions like those already asked and answered in practice can be used to develop prior distributions for a wide class of models. Further, we illustrate techniques for addressing some computational difficulties thought to be associated with the implementation of Bayesian methodology.

1 INTRODUCTION

A central problem in the design of simulations is the selection of appropriate input distributions to characterize the stochastic behavior of the modeled system (Law and Kelton 1991; Wagner and Wilson 1995). Failure to select appropriate input distributions can lead to misleading simulation output, and therefore to poor system design decisions.

Still, there is controversy about both classical and subjective techniques that are commonly used for input distribution selection. Critiques of classical techniques include: use of a single distribution and parameter underestimates the uncertainty in the distribution's functional form and parameter (Draper 1995); goodness-of-fit and P-value criteria don't quantify the probability that an input distribution is best (Berger and Delampady 1987); with few data points, few distributions are rejected, and with many data points, all distributions are rejected (Raftery 1995); and there is no coherent method for selecting among non-rejected distributions.

A critique of several approaches to subjective specification of an input distribution is that a single probability distribution (such as the triangular, truncated normal, Bézier, or 'smoothed' histograms) with a specific parameter value is chosen. As with the classical approach, this tends to underestimate the uncertainty about the underlying nature of the random process generating the data. Further, it is difficult to analyze how additional relevant data should affect the selected input distribution.

The Bayesian model average (BMA) approach described in Sec. 3 has been proposed as a mechanism for overcoming these difficulties (Draper 1995). The BMA has been considered in a simulation context (Cooke 1994, Scott 1996, Chick 1997), as well as applications in econometrics, artificial intelligence, sociology and medicine (e.g., see Draper 1995, Madigan and York 1995, Raftery 1995 and references therein). The BMA approach (like classical maximum likelihood techniques) uses likelihood functions to infer parameter values. Unlike classical techniques, it also quantifies uncertainty about which input distribution is most appropriate. Both the BMA approach and subjectivist techniques incorporate prior information. Unlike some subjectivist techniques that fit a specific histogram and provide little formalism for incorporating additional historical data, the BMA explicitly uses Bayes' rule and historical data for inference.

Still, there have been difficulties with implementing the BMA approach in practice. First, the BMA approach requires the assessment of prior probability distributions, a task perceived by many as difficult. Second, there is a computational price to pay for implementing the BMA approach. Two sources of this price are the (a) computation of posterior probability that a given model is correct, given available historical data, and (b) implementation of 'model averaging', which requires that input parameters be sampled from an appropriate posterior distribution for input into each replication. This latter point of the BMA approach allows for an evaluation of how input distribution uncertainty induces uncertainty in the expected value of the simulation output.

This paper describes a general approach for converting responses to questions similar to those *already asked in simulation practice* into prior distributions for parameters of distributions in the regular exponential family (including the exponential, gamma, normal, and Bernoulli distributions) or for parameters of shifted versions of distributions in the regular exponential family (such as the three-parameter gamma distribution).

The general approach is presented in Sec. 5 and illustrated by numerical example. For some cases where regularity conditions are lacking, such as for the shifted gamma distribution, we indicate that it is possible to apply Markov Chain Monte Carlo techniques, including the adaptive rejection Metropolis sampler (ARMS) of Gilks, Best, and Tan (1995), to generate random variates from appropriate posterior distributions. A Bayesian analysis, therefore, is a feasible and implementable approach for selecting input distributions for stochastic simulations in a way that coherently represents the effects of input distribution uncertainty on output uncertainty.

2 INPUT SELECTION PROBLEM

The input selection problem has many nuances and variations (e.g., see Cheng 1994, Cario and Nelson 1997, and Leemis 1995). Here, we restrict attention input distribution selection for a single source of randomness in a simulation. Multiple sources of randomness are treated similarly.

Suppose that a sequence of real-valued random quantities $X_1, X_2, \ldots$ are needed as input to a simulation, where the X_i are believed to be independent given the distribution and parameter, and historical data $\boldsymbol{x}_N = (x_1, \ldots, x_N)$ are available to help selection an input distribution.

The general problem is how to select distributions and their parameters for input into stochastic simulations. To describe the data, we choose $q < \infty$ candidate distributions, where distribution m has continuous parameter $\boldsymbol{\Lambda}_m = (\Lambda_{m,1}, \ldots, \Lambda_{m,d_m})$, and d_m is the dimension of $\boldsymbol{\Lambda}_m$, for $m = 1, \ldots, q$. The parameter takes on values $\boldsymbol{\lambda}_m = (\lambda_{m,1}, \ldots, \lambda_{m,d_m})$ in the space $\mathcal{Q}_m$. Denote by $f_{X|m,\boldsymbol{\lambda}_m}(x)$ the probability density function (pdf) for X, given m and $\boldsymbol{\lambda}_m$.

By means of example, suppose that a distribution for the times to failure of a machine is required for input to a stochastic simulation, and that 3 distributions are considered.

1. Exponential. $\boldsymbol{\lambda}_1 = (\theta)$, with rate $\theta > 0$,
$$f_{X|m=1,\boldsymbol{\lambda}_1}(x) = \theta e^{-x\theta}.$$
2. Normal. $\boldsymbol{\lambda}_2 = (\mu, \tau)$, with $\tau = 1/\sigma^2 > 0$,
$$f_{X|m=2,\boldsymbol{\lambda}_2}(x) = \frac{\tau^{1/2}}{(2\pi)^{1/2}} e^{-(x-\mu)^2\tau/2}.$$
3. Shifted gamma. $\boldsymbol{\lambda}_3 = (\xi, \alpha, \beta)$, with $\xi, \alpha, \beta > 0$.
$$f_{X|m=3,\boldsymbol{\lambda}_3}(x) = \frac{\beta^\alpha}{\Gamma(\alpha)}(x-\xi)^{(\alpha-1)} e^{-(x-\xi)\beta}.$$

This choice of parametrization simplifies the analysis below. Although the support of the normal distribution is the entire real line, and failure times are non-negative, in practice the parameters of the normal distribution give exceedingly small probabilities to negative values. If a negative value is generated during the simulation, it might be rejected (giving a truncated normal distribution). We similarly 'reject' negative values, but discuss the non-truncated normal distribution to simplify the exposition.

3 A BAYESIAN FORMULATION

Since the input distribution is unknown, let M be a random variable so that $M = m$ is the event that input distribution m is correct, for $m = 1, \ldots, q$. A Bayesian approach requires prior probability distributions to describe initial uncertainty regarding the unknown input distribution and parameter. Let $p_M(m)$ be the probability mass function (pmf) that describes prior belief that distribution m is the correct distribution. Let $f_{\boldsymbol{\Lambda}_m|m}(\boldsymbol{\lambda}_m)$ be the prior pdf that $\boldsymbol{\lambda}_m$ is the true parameter, given that m is correct. Sec. 5 describes how prior distributions can be specified.

Given the above assumptions, a Bayesian version input distribution selection problem is focused on the several posterior probability distributions, given that historical data $\boldsymbol{x}_N$ is available. The relevant distributions are: the marginal pdf $f_{\boldsymbol{X}_N|m}(\boldsymbol{x}_N)$ of the data, given that distribution m is correct,

$$f_{\boldsymbol{X}_N|m}(\boldsymbol{x}_N) = \int_{\mathcal{Q}_m} f_{\boldsymbol{X}_N|m,\boldsymbol{\lambda}_m}(\boldsymbol{x}_N) f_{\boldsymbol{\Lambda}_m|m}(\boldsymbol{\lambda}_m) d\boldsymbol{\lambda}_m \quad (1)$$

the posterior pmf $p_{M|\boldsymbol{x}_N}(m)$ that distribution M is correct, given $\boldsymbol{x}_N$,

$$p_{M|\boldsymbol{x}_N}(m) = \frac{f_{\boldsymbol{X}_N|m}(\boldsymbol{x}_N)\, p_M(m)}{\sum_{k=1}^{q} f_{\boldsymbol{X}_N|k}(\boldsymbol{x}_N)\, p_M(k)} \quad (2)$$

the posterior pdf $f_{\boldsymbol{\Lambda}_m|m,\boldsymbol{x}_N}(\boldsymbol{\lambda}_m)$ that a parameter $\boldsymbol{\Lambda}_m$, given m and $\boldsymbol{x}_N$

$$f_{\boldsymbol{\Lambda}_m|m,\boldsymbol{x}_N}(\boldsymbol{\lambda}_m) = \frac{f_{\boldsymbol{X}_N|m,\boldsymbol{\lambda}_m}(\boldsymbol{x}_N)\, f_{\boldsymbol{\Lambda}_m|m}(\boldsymbol{\lambda}_m)}{f_{\boldsymbol{X}_N|m}(\boldsymbol{x}_N)} \quad (3)$$

and the predictive pdf $f_{O|\boldsymbol{x}_N}(o)$ of the simulation output O, given $\boldsymbol{x}_N$:

$$f_{O|\boldsymbol{x}_N}(o) = \sum_{m=1}^{q} p_{M|\boldsymbol{x}_N}(m) \cdot \int_{Q_m} f_{O|m,\boldsymbol{\lambda}_m}(o)\, f_{\boldsymbol{\Lambda}_m|m,\boldsymbol{x}_N}(\boldsymbol{\lambda}_m)\, d\boldsymbol{\lambda}_m \tag{4}$$

Eq. 4 is called a Bayesian model average, and it describes uncertainty in O due to both input uncertainty and stochastic effects. To estimate $E[O \mid \boldsymbol{x}_N]$ with simulation, one samples input distributions and parameters for each replication from the posterior distributions in Eq. 2 and Eq. 3, and holds them fixed during a replication used to generate an observation $O = o$. This allows for uncertainty in the input parameters to propogate through to the output.

Eq. 4 assumes that $f_{O|m,\boldsymbol{\lambda}_m,\boldsymbol{x}_N}(o) = f_{O|m,\boldsymbol{\lambda}_m}(o)$, which reflects the reasonable modeling assumption that historical data for X and simulation output O are conditionally independent, given the distribution m and parameter $\boldsymbol{\lambda}_m$. This assumption states that if $(m, \boldsymbol{\lambda}_m)$ is the known true model, neither additional historical data nor additional random variates from the $(m, \boldsymbol{\lambda}_m)$ distribution will change our mind that $(m, \boldsymbol{\lambda}_m)$ is the true distribution.

4 INPUT SELECTION COMPUTATIONS

One practical issue for implementing the BMA approach is the determination of $p_{M|\boldsymbol{x}_N}(m)$ and $f_{\boldsymbol{\Lambda}_m|m,\boldsymbol{x}_N}(\boldsymbol{\lambda}_m)$. In some situations, closed-form results are available, as discussed in Sec. 4.1. In many cases, however, $f_{\boldsymbol{\Lambda}_m|m,\boldsymbol{x}_N}(\boldsymbol{\lambda}_m)$ may be known only up to a constant of proportionality. To determine this constant, one might apply one of five general techniques (Evans and Swartz 1995): asymptotic methods, Markov chain methods, importance sampling, adaptive importance sampling, and multiple quadrature. Here, we look at closed-form expressions, and the use of the ARMS sampler. For further discussion, see Chick (1999).

A second practical issue is the generation of variates from the posterior distribution of $\boldsymbol{\Lambda}_m$, given $m, \boldsymbol{x}_N$. We indicate that even for many distributions that do not satisfy certain regularity conditions, such as the shifted gamma distributions, it is possible to generate samples from the posterior distributions of their parameters, given historical data.

4.1 Closed-Form Expressions

The posterior distribution $f_{\boldsymbol{\Lambda}_m|m,\boldsymbol{x}_N}(\boldsymbol{\lambda}_m)$ can be determined exactly for certain special cases. Of particular importance is the regular exponential family of distributions, when *conjugate* prior distributions are employed (e.g., see Bernardo and Smith 1994). Members of this family include the exponential, normal, gamma, and Bernoulli, among others.

More formally, let the likelihood function be

$$f_{X|\boldsymbol{\lambda}}(x) = a(x) g(\boldsymbol{\lambda}) e^{\sum_{j=1}^{b} c_j \phi_j(\boldsymbol{\lambda}) h_j(x)} \tag{5}$$

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

$$f_{\boldsymbol{\Lambda}|\boldsymbol{t}}(\boldsymbol{\lambda}) = \frac{[g(\boldsymbol{\lambda})]^{t_0}}{K(\boldsymbol{t})} e^{\sum_{j=1}^{b} c_j \phi_j(\boldsymbol{\lambda}) t_j} \tag{6}$$

where the prior distribution parameter $\boldsymbol{t} = (t_0, \dots, t_b)$ is chosen so that $f_{\boldsymbol{\Lambda}|\boldsymbol{t},\boldsymbol{x}_N}(\boldsymbol{\lambda})$ is proper (i.e., $K(\boldsymbol{t}) = [g(\boldsymbol{\lambda})]^{t_0} \exp \left[\sum_{j=1}^{b} c_j \phi_j(\boldsymbol{\lambda}) t_j\right] d\boldsymbol{\lambda} < \infty$). Then the posterior distribution of $\boldsymbol{\lambda}$ is

$$f_{\boldsymbol{\Lambda}|\boldsymbol{t},\boldsymbol{x}_N}(\boldsymbol{\lambda}) = f_{\boldsymbol{\Lambda}|\boldsymbol{t}+\boldsymbol{t}'(\boldsymbol{x}_N)}(\boldsymbol{\lambda}), \tag{7}$$

where $\boldsymbol{t}'(\boldsymbol{x}_N) = (N, \sum_{i=1}^{N} h_1(x_i), \dots, \sum_{i=1}^{N} h_b(x_i))$ are sufficient statistics for $\boldsymbol{x}_N$, and $f_{\boldsymbol{\Lambda}|\boldsymbol{t}+\boldsymbol{t}'(\boldsymbol{x}_N)}(\boldsymbol{\lambda})$ indicates that the posterior results from inserting the coordinatewise sum $\boldsymbol{t} + \boldsymbol{t}'(\boldsymbol{x}_N)$ into Eq. 6.

For the regular exponential family with conjugate prior, then, the integral in Eq. 1 simplifies to:

$$f_{X_N|m}(\boldsymbol{x}_N) = \frac{K(\boldsymbol{t} + \boldsymbol{t}'(\boldsymbol{x}_N)) \prod_{i=1}^{N} a(x_i)}{K(\boldsymbol{t})} \tag{8}$$

4.1.1 Examples of Interest for Discrete-Event Stochastic Simulation

We now provide examples of conjugate analysis for distributions that arise often in discrete-event stochastic simulation practice, and indicate how the parameters of the conjugate prior distributions relate to statements about the mean and variance of the unknown input parameters themselves. Most of the results are known in Bayesian circles, but are not always presented in the form given here, and are not all well-known in non-Bayesian circles. The results are used in Sec. 5.1 to help assess prior distributions. The analysis for the gamma distribution has apparently not yet appeared in the literature.

Exponential. For the exponential distribution of Sec. 2, $b = 1$, $a(x) = 1$, $g(\theta) = \theta$, $c_1 = 1$, $\phi_1(\theta) = -\theta$, $h_1(x) = x$. This leads to the conjugate prior

$$f_{\Theta|t_0,t_1}(\theta) \propto \theta^{t_0} e^{-\theta t_1}, \tag{9}$$

the gamma$(t_0 + 1, t_1)$ distribution, so $K(t_0, t_1) = \Gamma(t_0 + 1)/(t_1)^{t_0+1}$. It follows that the mean time to failure $1/\theta$ has inverted gamma distribution with mean t_1/t_0 and variance ${t_1}^2/({t_0}^2(t_0 - 1))$.

Normal. For the normal distribution with unknown mean μ and precision τ, $b = 2$, $a(x) = 1$, $g(\mu, \tau) = (\tau/2\pi)^{1/2} \exp[-\tau\mu^2/2]$, $c_1 = 1$, $\phi_1(\mu, \tau) = \mu\tau$, $h_1(x) = x$, $c_2 = 1$, $\phi_2(\mu, \tau) = -\tau/2$, $h_2(x) = x^2$. The conjugate prior is then

$$f_{\mu,\tau|\boldsymbol{t}}(\mu, \tau) = \frac{1}{K(\boldsymbol{t})} \left[\left(\frac{\tau}{2\pi} \right)^{1/2} e^{-\frac{\tau\mu^2}{2}} \right]^{t_0} e^{\mu\tau t_1 - \frac{\tau t_2}{2}}, \quad (10)$$

which can be manipulated to deduce that τ has gamma($(t_0 + 1)/2, \tilde{t}_2/2$) distribution, where $\tilde{t}_2 = t_2 - {t_1}^2/t_0$, and the conditional distribution of μ, given τ, is normal($t_1/t_0, \tau t_0$). Thus $\sigma^2 = 1/\tau$ has inverted gamma distribution with mean $\tilde{t}_2/(t_0 - 1)$, and μ has Student-t marginal distribution with mean t_1/t_0, precision $t_0(t_0 + 1)/\tilde{t}_2$, and $t_0 + 1$ degrees of freedom. The marginal variance of μ is then $\tilde{t}_2/(t_0(t_0 - 1))$. By integrating Eq. 10 with respect to μ, then τ, and noting the distributions of μ, τ, one obtains $K(\boldsymbol{t}) = \Gamma((t_0 + 1)/2) / \left((t_0)^{\frac{1}{2}} (2\pi)^{\frac{t_0 - 1}{2}} (\tilde{t}_2/2)^{(t_0+1)/2} \right)$.

Two-parameter Gamma. Set $b = 2$, $a(x) = 1$, $g(\alpha, \beta) = (\beta)^\alpha / \Gamma(\alpha)$, $c_1 = 1$, $\phi_1(\alpha, \beta) = \alpha - 1$, $h_1(x) = \log x$, $c_2 = 1$, $\phi_2(\alpha, \beta) = -\beta$, $h_2(x) = x$. The conjugate prior is then

$$\begin{aligned} f_{\alpha,\beta|t_0,t_1,t_2}(\alpha, \beta) &\propto \left(\frac{\beta^\alpha}{\Gamma(\alpha)} \right)^{t_0} e^{(\alpha-1)t_1 - \beta t_2} \\ &\propto \frac{e^{t_1(\alpha-1)}}{(\Gamma(\alpha))^{t_0}} \beta^{\alpha t_0} e^{-\beta t_2} \qquad (11) \end{aligned}$$

Thus, the conditional distribution of β, given α, is gamma($\alpha t_0 + 1, t_2$). The expected value of the unknown mean α/β is $E[\alpha/\beta] = E[\alpha E[1/\beta \mid \alpha]] = E[\alpha(t_2/(\alpha t_0))] = t_2/t_0$. The variance α/β^2 does not exist when $\alpha t_0 \leq 1$, and the support of α allows for this event. However, conditional on $\alpha t_0 > 1$ (e.g., to reflect prior belief that a failure rate is increasing), $E[\alpha/\beta^2 \mid \alpha t_0 > 1] = E[E[\alpha/\beta^2 \mid \alpha, \alpha t_0 > 1]] = (E[1/(\alpha t_0 - 1) \mid \alpha t_0 > 1])\, {t_2}^2/t_0$. Similarly, uncertainty about the unknown mean can be expressed $\mathrm{Var}[\alpha/\beta \mid \alpha t_0 > 1] = (E[1/(\alpha t_0 - 1) \mid \alpha t_0 > 1])\, {t_2}^2/{t_0}^2$.

Bernoulli. The Bernoulli distribution with parameter $\nu \in [0, 1]$ and outcomes $x \in \{0, 1\}$ can be written $f(x \mid \nu) = (1 - \nu) \exp\left[\log\left(\frac{\nu}{1-\nu} \right) x \right]$. Thus $b = 1$, $a(x) = 1$, $g(\nu) = (1 - \nu)$, $c_1 = 1$, $\phi_1(\nu) = \log\left(\frac{\nu}{1-\nu} \right)$, $h_1(x) = x$. The conjugate prior is then

$$f_{\nu|t_0,t_1}(\nu) = [K(t_0, t_1)]^{-1} \nu^{t_1} (1 - \nu)^{t_0 - t_1} \qquad (12)$$

This recovers the beta distribution as the conjugate prior distribution (often written with parameters $\alpha = t_1 + 1, \beta = t_0 - t_1 + 1$). It follows that $K(t_0, t_1) = \Gamma(t_1 + 1)\Gamma(t_0 - t_1 + 1)/\Gamma(t_0 + 2)$.

4.2 Non-Regular Distribution?

There are many distributions of interest in simulation that are not members of the regular exponential family. This includes many shifted versions of exponential distributions, such as the shifted gamma and shifted log-normal, the triangular, and the Weibull distribution. The analysis of Sec. 4.1 does not apply to these distributions for which regularity conditions do not hold.

Fortunately, imporantace sampling, Markov chain Monte Carlo(MCMC), and other techniques are often useful for determining the posterior distributions of Eq. 2 and Eq. 3, and even generating samples from those distributions.

A published Bayesian work that deals with the shifted gamma distribution in particular is unknown to the author, however, we were able to modify the generic adaptive rejection Metropolis sampler (ARMS) of Gilks, Best, and Tan (1995) to the specific problem of the shifted distribution. The ARMS was adapted for calculations in Sec. 6 that involve the shifted gamma distribution.

5 PRIOR DISTRIBUTION SELECTION

The BMA approach requires the selection of a prior pmf $p_M(m)$ and a proper prior pdf $f_{\boldsymbol{\Lambda}_m|m}(\boldsymbol{\lambda}_m)$ (e.g., see Savage 1972 for an early discussion of subjective specification of prior distributions, and Bernardo and Smith 1994 for a more recent treatment). However, prior distribution assessment is typically viewed as an onerous task, requiring significant involvement on the part of the decision-maker. There is therefore significant interest in tools that can simplify the process of selecting a prior distribution over the space of unknown input distributions and parameters. Sec. 5.1 describes a moment-matching method that uses responses to questions similar in nature to those already asked and answered in simulation practice.

Note that there are several alternative approaches for automating the selection of prior distributions (e.g., Kass and Wasserman 1996). Berger and Pericchi (1996) discuss an intrinsic Bayes factor, present examples for the exponential, log-normal, and Weibull distributions, and more generally consider all location-scale distribution, i.e. such that $f_{\boldsymbol{X}_N|\mu,\sigma}(\boldsymbol{x}_N) = \prod_{j=1}^N g((x_j - \mu)/\sigma)/\sigma$. An alternate approach is the fractional Bayes factor of O'Hagan (1995). Each approach has strengths and weaknesses.

5.1 Moment-Matching Method

First, a common (but not obligatory) choice for $p_M(m)$ is the discrete uniform distribution, $p_M(m) = 1/q$. For continuous parameters $\boldsymbol{\lambda}_m$, we describe a moment matching approach that more fully explores a general idea for selecting prior distributions described by Berger (1985). The idea is to select prior distributions by insuring that certain moments of functions of the unknown parameters reflect beliefs of a decision-maker. Here, we treat separately the specific case of parameters of input distributions in the regular exponential family, as well as the case of shifted versions of those distributions.

In broad strokes, we use the conjugate prior distribution for regular exponential distributions to model prior beliefs, and to choose a parameter $\boldsymbol{t}$ so that the prior distribution (i) is proper (integrates to 1), and (ii) covers a 'reasonable range' of values, as judged by the decision-maker. Parameters for the conjugate prior distribution are determined by the responses to a few general questions about the decision-makers beliefs about likely values of low-order moments of the random quantity being modeled. Sample questions are:

- What is a likely value of the unknown mean of the random quantity being modeled (e.g., what is a likely value for the mean time to failure)?
- What is a likely range for the unknown mean of the random quantity being modeled (e.g., the unknown mean time to failure is most likely to be found in what range)?
- What is a likely value for the unknown variance of the random quantity being modeled (e.g., what is a likely value for the variance of times to failure)?

These questions are *similar to those already asked in common practice* when subjective assessment of input distributions is required for a simulation. The number of questions required to assess the parameters equals the number of parameters of the conjugate distribution (one more than the dimensionality of the parameter $\boldsymbol{\lambda}$).

Examples. Suppose that a prior distribution for the unknown rate θ of the exponential distribution is desired, and that a decision-maker indicates that the unknown mean of the exponential distribution is likely to be 240 ± 100 minutes (in response to the first two questions above). This is reasonably be translated into the statements $E[1/\theta] = 240$ and $\mathrm{Var}[1/\theta] = 100^2$. Using Sec. 4.1.1, this implies that the parameters for the conjugate prior satisfy $t_1/t_0 = 240$ and ${t_1}^2/({t_0}^2(t_0 - 1)) = 100^2$, or $(t_0, t_1) = (6.76, 1622.4)$.

There may be argument as to whether or not this is the 'best' translation of the responses to mathematical statements, as some might argue that modes are better than means, or that a multiplicative factor should be included with the variance. However, for each variations, the basic idea can be used without change: constrain the parameters of the prior distribution with responses to a few simple questions.

Turn now to the assessment of the prior distribution for unknown mean μ and precision τ of the normal distribution. The conjugate prior has 3 parameters, so we use the three questions listed at the beginning of this section. In addition to the response that the unknown mean is likely to be 240 ± 100, we use a response of 120^2 min^2 as an estimate for the unknown variance in times to failure. Following Sec. 4.1.1, these natural language statements are translated into $t_1/t_0 = 240$; $\tilde{t}_2/(t_0(t_0-1)) = 100^2$; $\tilde{t}_2/(t_0-1) = 120^2$. Algebra indicates that $(t_0, t_1, \tilde{t}_2) = (1.44, 345.6, 6336)$, so that $t_2 = 89280$.

A modeling issue might arise in general for the normal distribution, although the issue did not arise for this specific example. To obtain a proper prior distribution, one must have $t_0 > 1$. This requires that the estimate for the unknown variance (120^2 in the example) exceed the estimate for the variance in the unknown mean (100^2 in the sexample). If the decision maker's responses violate this constraint, one might modify the response to the third question to a value slightly larger than the response to the second question. Then t_0 will exceed 1, and historical data will typically outweigh the prior distribution.

Shifted distributions. The shifted gamma distribution, like many shifted distributions, does not itself lie within the regular exponential family. There is therefore no finite-dimensional conjugate distribution for the shifted gamma distribution. Here we propose a method for assessing prior distributions for shifted versions of members of the regular exponential family when there is no finite-dimensional conjugate prior, and apply the methodology to the specific case of the shifted gamma distribution.

The idea is (i) to assess a prior distribution for the shift parameter, then (ii) to assess a prior distribution for the remaining parameters *independent* of the value of the shift parameter. While this assumption might not accurately reflect all beliefs about input distributions, one can assass the parameters in a way that produces a sufficiently diffuse prior distribution so as not to dominate the posterior distribution. For a given value of the shift parameter ξ, then, the data $x_i - \xi$ have an unshifted regular exponential family distribution. Typically, ξ and the other parameters are correlated, conditional on $\boldsymbol{x}_N$, even though the prior distribution initially has them independent.

We illustrate this process for the specific case of the shifted gamma distribution. First, we assess a prior distribution for the shift parameter ξ. Since times-to-failure must be positive, we have $\xi > 0$. Suppose that the decision-maker has a hard time believing that the minimum time to failure is more than 10 minutes. Any prior distribution on (0,10) might be selected for ξ. For sake of argument, we select uniform(0,10).

By analogy for with the conjugate prior for the two-parameter gamma distribution in Sec. 4.1.1, for the three-parameter gamma distribution we select a prior distribution for α, β as

$$f_{\alpha,\beta|3}(\alpha, \beta) \propto \left(\frac{\beta^{\alpha}}{\Gamma(\alpha)}\right)^{t_0} e^{(\alpha-1)t_1-\beta t_2}, \qquad (13)$$

so the conditional posterior pdf, given ξ, is

$$f_{\cdot|3,\xi,\boldsymbol{x}_N}(\alpha, \beta) \propto \frac{\beta^{\alpha(t_0+N)}}{\Gamma(\alpha)^{t_0+N}} \frac{e^{(\alpha-1)(t_1+\sum_{i=1}^{N}\log(x_i-\xi))}}{e^{\beta\left(t_2+\sum_{i=1}^{N}(x_i-\xi)\right)}}. \quad (14)$$

Next, we match moments as above. Clearly,

$$E[\xi+\alpha/\beta] = E[\xi] + E[\alpha/\beta] = 5 + t_2/t_0.$$

Similarly, it is straightforward to determine

$$\begin{aligned} \mathrm{Var}[\xi+\alpha/\beta \mid \alpha t_0 > 1] &= \mathrm{Var}[\xi] + \varpi\left(\frac{t_2}{t_0}\right)^2 \\ E[\alpha/\beta^2 \mid \alpha t_0 > 1] &= \varpi\frac{{t_2}^2}{t_0} \end{aligned}$$

where $\varpi = E[1/(\alpha t_0 - 1) \mid \alpha t_0 > 1]$, by analogy with the arguments in Sec. 4.1.1.

These equations provide constraints to help solve for (t_0, t_1, t_2), using the same responses to the questions used to determine a prior for the unknown parameters of the normal distribution. In particular, $t_2/t_0 = 235$; $100/12 + \varpi(t_2/t_0)^2 = 100^2$; and $\varpi {t_2}^2/t_0 = 120^2$. The constraint $t_0 > 1$ is implicitly required, as with the normal distribution. Thus $t_0 = 1.441$; $t_2 = 338.7$; and t_1 is chosen to satisfy some additional requirements that bear a complicated relationship with the other parameters (a larger t_1 corresponds to a smaller sample variance for α/β in MCMC experiments).

A ‘reasonable’ value of t_1 can be determined heuristically, since one might not wish to condition on the event $\alpha t_0 > 1$ (this event does not have probability 1 under the conjugate prior). Intuitively, since t_1 is associated with the sum of logarithms of the data in terms of data sufficiency, and t_2 is associated with the sum of the roughly t_0 data elements, the choice $t_1 = \log(t_2/t_0) \approx 5.46$ may be reasonable. Since t_0 is small, the posterior distribution will not be too sensitive to small changes in t_1 if any appreciable amount of historical data is available.

Comments on Moment Matching. The questions asked for specifying moments are similar to, but not the same as, questions already asked in practice for subjectively specifing an input distribution. Take the normal distribution, for example. In practice, a decision-maker is asked to specify a mean and variance (response to first and third question above) that is input deterministically into each replication. With the above moment matching approach, a *range* for the unknown mean and an estimated variance are obtained from the decision maker, and that prior is used with the data to provide appropriate samples of parameters for input into each replication.

Although the above treatment of the shifted gamma distribution requires some *ad hoc* tricks, it does not appear more (or less) *ad hoc* than ‘tricks’ used in widely-available software to handle related difficulties for maximum likelihood approaches for non-regular distributions.

In the above examples, the same responses to the questions at the beginning of this section were used to determine the prior distributions for the unknown parameters of each candidate input distribution. It is also possible to allow the responses to be conditional on the input distribution. This allows a decision-maker to tailor responses to address peculiarities associated with each input distribution.

6 COMPUTATIONAL EXPERIMENTS

Because of space limitations, charts and graphs that illustrate the results of computational experiments are not presented. They will be displayed during the conference presentation.

6.1 Inferring a Downtime Distribution

The 3 input distributions from Sec. 2 and the prior distributions from Sec. 5 were used together with time-to-failure data (N=37 observations) taken from a factory floor. The posterior distributions for the input distributions and parameters in Eq. 2 and Eq. 3 were then determined. Calculations for the exponential and normal distributions used the closed-form calculations of Sec. 4.1.1. Calculations for the shifted gamma distributions were implemented by (a) running a customized ARMS algorithm (1000 iterations) to explore the shape of the posterior distribution of ξ, α, β, (b) using importance sampling (2000 samples) to estimate the integral in Eq. 1 (the importance sampling measure resembled the histogram of the marginals of the ξ, α, β).

The exponential and normal distributions were effectively eliminated from consideration (posterior probabilities of less than 10^{-8}), and the shifted gamma was the clear favorite. Results took about 2 seconds on a vanilla SPARC-station, but less time is actually required, as less than 2000 IS samples would be required to identify the shifted gamma as the most likely distribution in this case.

Discrete-event simulations of system indicate that maximum likelihood/goodness of fit techniques indeed underestimate the variance in the mean output, as a result of ignoring the structural uncertainty about the values of the input parameters.

6.2 Detecting a Known Distribution

Several experiments were run to test the ability of identifying a known, true distribution. Artificially generated sets of data with (i) exponential, (ii) normal, and (iii) shifted gamma distributed random variates were used in conjunction with the same prior distributions that were used above. The Bayesian formulation identified the true distribution as the 'favorite' for experiments whre the true distribution was exponential or normal. For the shifted gamma with $\alpha = 1/2$, the exponential was sometimes initially favored with few data points, but the shifted gamma is favored as the number of data points increases. When $\alpha = 3$, the normal distribution is also sometimes heavily favored. For those data sets, however, classical goodness-of-fit tests also favored the normal distribution, rather than the true shifted gamma.

6.3 Diagnostics

One Bayesian diagnostic is to test the sensitivity of the posterior probabilities to small changes in the parameters of the prior distributions. The results of the above analysis are relatively stable with respect to the parameters of the prior distribution.

Another diagnostic is to nest the most likely distribution into a larger distribution. Since the most likely input distribution in Sec. 6.1 was the shifted gamma distribution, the analysis was re-run by testing if the data is better described by a shifted gamma or a shifted mixture of two gamma distributions. The shifted mixture of two gamma distributions turned out to be more likely, a result that supports the field observation that two failure mechanisms were at work (one for short, the other for long times-to-failure). The χ^2-test of goodness of fit, on the other hand, was ambiguous as to whether the shifted gamma or shifted mixture of two gamma distributions was better supported by the data.

7 DISCUSSION AND CONCLUSIONS

A number of authors (e.g., Draper 1995, Chick 1997) argue that Bayesian techniques are required in order to faithfully account for the effect of parameter uncertainty on the output of a model.

This paper illustrates the practical use of tools for implementing Bayesian input distribution selection. Particular attention is given to addressing two of the reasons that have been presented for avoiding Bayesian methods for input distribution selection. First, the paper presents practical examples for specifying prior probability distributions for the unknown input parameters. Second, the paper indicates that the ARMS algorithm can be used to handle shifted distributions, distributions that typically do not satisfy a number of regularity conditions.

While the techniques outlined in this paper are applicable to a wide variety of distributions used in simulation practice, they do not apply to all distributions. Some additional work is needed to handle distributions that are not in the regular exponential family, and there may be some demand from practitioners to have data-driven prior distribution selection, so that no questions need be asked.

REFERENCES

Berger, J. O. 1985. *Statistical Decision Theory and Bayesian Analysis* (2nd ed.). New York: Springer-Verlag.

Berger, J. O. and M. Delampady. 1987. Testing precise hypothesis (with discussion). *Statistical Science 2,* 317–352.

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

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

Cario, M. C. and B. L. Nelson. 1997. Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix. *Northwestern University, IEMS Technical Report*.

Cheng, R. C. H. 1994. Selecting input models. In J. D. Tew, S. Manivannan, D. A. Sadowski, and A. F. Seila (Eds.), *Proceedings of the Winter Simulation Conference,* pp. 184–191. Institute of Electrical and Electronics Engineers, Inc.

Chick, S. E. 1997. Selecting the best system: A decision-theoretic approach. In S. Andradóttir, K. Healy, D. Withers, and B. Nelson (Eds.), *Proceedings of the Winter Simulation Conference,* pp. 326–333. Institute of Electrical and Electronics Engineers, Inc.

Chick, S. E. 1999. Input distribution selection for simulation experiments: Accounting for input uncertainty. *in submission*.

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

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

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.

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

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

Law, A. M. and W. D. Kelton. 1991. *Simulation Modeling & Analysis* (2nd ed.). New York: McGraw-Hill, Inc.

Leemis, L. M. 1995. Input modeling for discrete-event simulation. In C. Alexopoulos, K. Kang, W. R. Lilegdon, and D. Goldsman (Eds.), *Proceedings of the Winter*

Simulation Conference, pp. 16–23. Institute of Electrical and Electronics Engineers, Inc.

Madigan, D. and J. York. 1995. Bayesian graphical models for discrete data. *International Statistical Review 63*(2), 215–232.

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

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

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 J. M. Charnes, D. J. Morrice, D. T. Brunner, and J. J. Swain (Eds.), *Proceedings of the Winter Simulation Conference*, pp. 255–259. Institute of Electrical and Electronics Engineers, Inc.

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

AUTHOR BIOGRAPHY

STEPHEN E. CHICK is an assistant professor of Industrial and Operations Engineering at the University of Michigan, Ann Arbor. In addition to stochastic simulation, his research interests include Bayesian statistics, reliability, decision analysis, and computational methods in statistics. His research is motivated by projects in manufacturing and health care. His work experience includes several years of material handling system design for the automotive industry using simulation analysis.