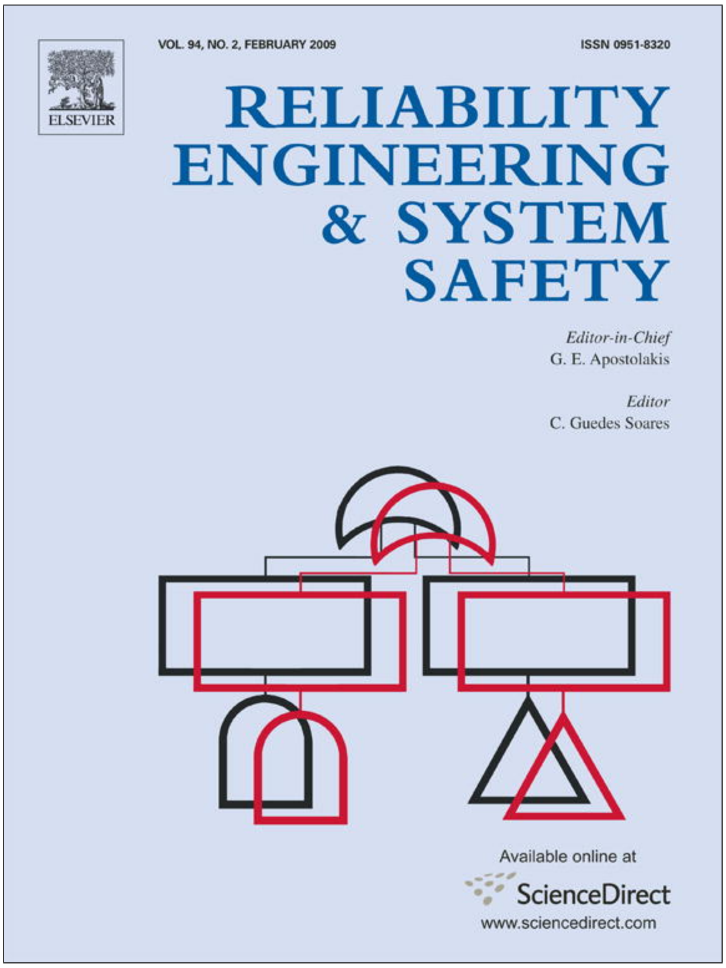
VOL. 94, NO. 2, FEBRUARY 2009
ISSN 0951-8320
ELSEVIER
RELIABILITY ENGINEERING & SYSTEM SAFETY
Editor-in-Chief
G. E. Apostolakis
Editor
C. Guedes Soares
Available online at
ScienceDirect
www.sciencedirect.com

Reliability Engineering and System Safety 94 (2009) 628–643

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

Bayesian inference in probabilistic risk assessment—The current state of the art

Dana L. Kelly *, Curtis L. Smith

Idaho National Laboratory, P.O. Box 3815, Idaho Falls, ID 83415, USA

ARTICLE INFO

Article history:
Received 31 March 2008
Received in revised form
1 July 2008
Accepted 2 July 2008
Available online 9 July 2008

Keywords:
Bayesian inference
Markov chain Monte Carlo
Parameter estimation
Model validation
Probabilistic risk analysis

ABSTRACT

Markov chain Monte Carlo (MCMC) approaches to sampling directly from the joint posterior distribution of aleatory model parameters have led to tremendous advances in Bayesian inference capability in a wide variety of fields, including probabilistic risk analysis. The advent of freely available software coupled with inexpensive computing power has catalyzed this advance. This paper examines where the risk assessment community is with respect to implementing modern computational-based Bayesian approaches to inference. Through a series of examples in different topical areas, it introduces salient concepts and illustrates the practical application of Bayesian inference via MCMC sampling to a variety of important problems.

1. Introduction

The advent of Markov chain Monte Carlo (MCMC) sampling has proliferated Bayesian inference throughout the world, across a wide array of disciplines. The freely available software package known as Bayesian inference using Gibbs sampling (BUGS) has been in the vanguard of this proliferation since the mid-1990s [1]. However, more recent advances in this software, leading first to WinBUGS and now to an open-source version (OpenBUGS), including interfaces to the open-source statistical package R [2], have brought MCMC to a wider audience. Problems which would have been intractable a decade ago can now be solved in short order with these software packages. We will use this software to illustrate several applications from the field of risk assessment. While our examples are from applications of risk assessment to technology, it should be noted that there have also been many applications of Bayesian inference in behavioral science [3], finance [4], human health [5], process control [6], and ecological risk assessment [7]. We mention these other applications to indicate the degree to which Bayesian inference is being used in the wider community, including other branches of risk assessment; however, our principal focus is on applications of risk assessment to engineered systems (including non-nuclear technology). Within this focus, we will explore specific Bayesian topics and examples in greater detail, including

- Hierarchical modeling of variability
- Modeling of time-dependent reliability (with and without repair)
- Modeling of random durations, such as the time to suppress a fire or recover power
- Treatment of uncertain and missing data
- Regression models
- Model selection and validation

Siu and Kelly's earlier work [8] is the starting point for this paper, as it presented a tutorial on Bayesian inference for probabilistic risk assessment, and the elementary portions of that paper remain vital today. For the reader needing a more basic introduction to Bayesian inference, that paper is a good starting point, and it also provides a good point of entry to the vast (and vastly expanding) literature on Bayesian inference. However, advances have been made since the publication of that paper in 1998, and it is to these advances that the current paper is devoted. Additional relevant information on the Bayesian approach may be found in [9] and [10] has become a popular practical reference for the PRA community.

2. Background

First, we provide a brief overview of key concepts relevant to later discussion. In general, we have a goal of performing

* Corresponding author. Tel.: +1 360 539 8388; fax: +1 360 539 8401.
E-mail address: Dana.Kelly@inl.gov (D.L. Kelly).

0951-8320/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ress.2008.07.002

inference calculations, starting with data. "Data" are the observed values of a physical process and may be subject to uncertainties, such as imprecision in measurement, censoring, and interpretation errors. Next, at a higher level from data, we have "information," which is the result of evaluating, manipulating, or organizing data and other information in a way that adds to knowledge. "Knowledge" is what is known from gathered information. Lastly, inference is the process of obtaining a conclusion based on what one knows.

One must analyze data in order to produce information, information that will ultimately be used for making inferences. Like data, information has organizational value. However, in order to evaluate or manipulate data, we must have a "model of the world" (or simply "model") that allows us to translate real-world observables into information [11,12]. Within this model of the world, there are two fundamental types of model abstraction, aleatory (also referred to as probabilistic or stochastic) and deterministic. The models that will be described herein are aleatory and parametric, and most of the model parameters are themselves imprecisely known, and therefore uncertain. Consequently, to describe this second layer of uncertainty, we employ the notion of epistemic uncertainty (also referred to as state-of-knowledge uncertainty). See [13] for a more extensive discussion.

All of the inference we will illustrate hinges upon Bayes' Theorem, which we restate here for convenience

$$\pi_1(\theta|x) = \frac{f(x|\theta)\pi_o(\theta)}{\int_\Theta f(x|\theta)\pi_o(\theta)\,d\theta} \tag{1}$$

In this equation, θ is the unknown parameter of interest (often a vector), $\pi_o(\theta)$ is the prior distribution of θ, $f(x|\theta)$ is the likelihood function (i.e., the aleatory model for x, conditional upon a value of θ), and $\pi_1(\theta|x)$ is the posterior distribution of θ. See [8] or [10] for more details.

3. Hierarchical Bayesian modeling of variability

As discussed by Siu and Kelly [8], treatment of variability that can exist among sources of data is important if total uncertainty, including population variability, is to be properly represented by the resulting posterior distribution. Siu and Kelly [8] discusses hierarchical Bayesian analysis briefly, but (due to software and computer limitations that existed at that time) focuses on an older approach of Kaplan [14] called "two-stage Bayes," and on (parametric) empirical Bayes. As pointed out by Siu and Kelly [8], the two-stage Bayes approach can be viewed as a particular example of the more general hierarchical Bayes, and empirical Bayes can be viewed as an approximation to hierarchical Bayes. Thus, with the advent of MCMC software, the hierarchical Bayesian approach represents an achievable, fully Bayesian state of the art and two-stage Bayes is no longer a recommended approach [10]. Empirical Bayes, which is not fully Bayesian, still has a role to play, however, as will be discussed below.

We first review the general framework of hierarchical Bayes. We will then illustrate this framework with numerical examples for hardware failure. Hierarchical Bayes is so-named because it utilizes hierarchical or multistage prior distributions. In the hierarchical Bayes framework, the prior distribution for the parameter of interest, denoted $\pi(\theta)$, is written as

$$\pi(\theta) = \int_\Phi \pi_1(\theta|\varphi)\pi_2(\varphi)\,d\varphi \tag{2}$$

In Eq. (2), $\pi_1(\theta|\varphi)$ is the first-stage prior, representing the population variability in θ for a given value of φ (note that φ is typically a vector), and $\pi_2(\varphi)$, called the hyperprior, is the distribution representing the uncertainty in φ, whose components are called hyperparameters. The first-stage prior, $\pi_1(\theta|\varphi)$, is usually assumed to be of a particular parametric form, such as a gamma or lognormal distribution when $\theta > 0$ or a normal distribution when $-\infty < \theta < \infty$. It is also typical, although not necessary, to use independent diffuse hyperpriors for the components of φ. Although nothing limits the analysis to two stages, the use of more than two stages has been rare in applications.

3.1. Illustrative example of hierarchical Bayes

Assume the data in Table 1 from 12 sources are available on the failure rate of a particular type of component.

The parameter of interest is λ, the intensity of the Poisson distribution that describes the number of failures, x, in exposure time t:

$$f(x|\lambda) = \frac{(\lambda t)^x e^{-\lambda t}}{x!}, \quad x = 0, 1, \ldots \tag{3}$$

Our prior belief is that there is source-to-source variability in λ. We will model this variability with a gamma-distribution. Thus, in the hierarchical Bayesian approach, the first-stage prior for λ, denoted $\pi_1(\theta|\varphi)$ above, will be

$$\pi_1(\lambda|\alpha, \beta) = \frac{\beta^\alpha \lambda^{\alpha-1} e^{-\beta\lambda}}{\Gamma(\alpha)} \tag{4}$$

Therefore $\varphi = (\alpha,\beta)^T$ is the vector of hyperparameters. Hierarchical problems such as this one are most simply represented via a Bayesian network, also known as a directed acyclic graph or DAG [15]. In fact, because the WinBUGS software package uses the underlying DAG model to express the joint posterior distribution as a product of conditional distributions, the network approach allows for almost "math-free" inference, freeing the analyst from the need to write out Bayes' Theorem, which can quickly become quite complex as the number of parameters in a problem increases. The DAG for our example problem is shown in Fig. 1. Note in this figure that α and β are independent, prior to the observation of the data. This obviates the need to develop a joint prior distribution for α and β that includes dependence. Once the data are observed, these nodes become dependent, and the joint posterior distribution will reflect this dependence.

We will assume independent diffuse hyperpriors on α and β. Note that in some cases the high degree of correlation between α and β can lead to very slow convergence to the joint posterior distribution. In such cases, it may be helpful to reparameterize the problem in terms of the mean and coefficient of variation, which are approximately independent in the joint posterior distribution. For more details on this and other issues that can arise, see [16].

The posterior predictive distribution of λ, representing source-to-source variability, will be given by an average of the posterior

Table 1
Component failure rate data for hierarchical Bayes example

Source	Failures	Exposure time (h)
1	0	87,600
2	7	525,600
3	1	394,200
4	0	87,600
5	8	4,555,200
6	0	306,600
7	0	394,200
8	0	569,400
9	5	1,664,400
10	1	3,766,800
11	4	3,241,200
12	2	1,051,200

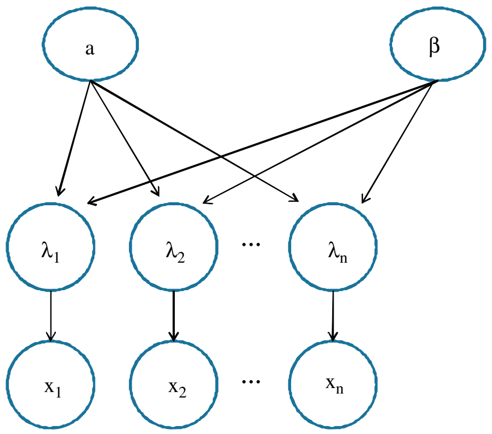

Fig. 1. Directed acyclic graph for hierarchical Bayes model of population variability.

Table 2
WinBUGS script for hierarchical Bayes analysis of population variability in λ

```
model {
for(i in 1:12) {
x[i]~dpois(mu[i]) #Poisson distribution for number of failures in
   each source
mu[i] <- lambda[i]*t[i] #Parameter of Poisson distribution
lambda[i]~dgamma(alpha, beta) # Distribution for selecting lambda
   in each source
}
#Posterior predictive distribution for lambda
lambda.star~dgamma(alpha, beta)
#Hyperpriors on alpha and beta
alpha~dgamma(0.0001, 0.0001)
beta~dgamma(0.0001, 0.0001)
}

Inits
list(alpha = 0.5, beta = 6.E+5)
list(alpha = 1.5, beta = 4.E+5)
```

distribution for λ, conditional upon α and β (a gamma-distribution), weighted by the posterior distribution for α and β. We can take advantage of the fact that, as Fig. 1 illustrates, the components λ_i are conditionally independent, given α and β, to write

$$\begin{aligned}\pi(\lambda_i|\tilde{x},\tilde{t}) &= \int_0^\infty \int_0^\infty \cdots \int_0^\infty \left\{ \int\int \left[\prod_{i=1}^{n} \pi_1(\lambda_i|\alpha,\beta)\right] \pi_2(\alpha,\beta|\tilde{x},\tilde{t})\,\mathrm{d}\alpha\,\mathrm{d}\beta \right\} \\ &\quad \times \mathrm{d}\lambda_1\,\mathrm{d}\lambda_2 \ldots \mathrm{d}\lambda_{i-1}\,\mathrm{d}\lambda_{i+1} \ldots \mathrm{d}\lambda_n \\ &= \int\int \pi_1(\lambda_i|\tilde{x},\tilde{t},\alpha,\beta)\pi_2(\alpha,\beta|\tilde{x},\tilde{t})\,\mathrm{d}\alpha\,\mathrm{d}\beta \end{aligned} \qquad (5)$$

The third line in Eq. (5) is obtained by interchanging the order of integration. Thus, the marginal posterior distribution for λ for any particular source is a continuous mixture of gamma-distributions, mixed over the posterior distribution of the hyperparameters, α and β. The distribution describing source-to-source variability in λ is the posterior predictive distribution, also referred to as the average population variability curve.

$$\begin{aligned}\pi(\lambda^*|\tilde{\lambda}) &= \int\int \pi_1(\lambda^*|\alpha,\beta,\tilde{\lambda},\tilde{x},\tilde{t})\pi_2(\alpha,\beta|\tilde{x},\tilde{t})\,\mathrm{d}\alpha\,\mathrm{d}\beta \\ &= \int\int \pi_1(\lambda^*|\alpha,\beta,\tilde{x},\tilde{t})\pi_2(\alpha,\beta|\tilde{x},\tilde{t})\,\mathrm{d}\alpha\,\mathrm{d}\beta \end{aligned} \qquad (6)$$

It is thus a similar mixture of gamma-distributions. It is generated in WinBUGS by sampling α and β from their joint posterior distribution, and then sampling λ^* from a gamma-distribution. The WinBUGS script used to analyze this problem is shown in Table 2.

Two Markov chains, each starting from a separate point in the parameter space, were used in the analysis. Running more than two chains may be needed in general, although two are sufficient for this example. Running more than one chain aids in checking convergence. Each of the chains must be given an initial value of α and β. As discussed by Gelman [17], the starting values should be over-dispersed around the mode of the posterior distribution of the hyperparameters to ensure adequate coverage of the posterior distribution. Empirical Bayes, discussed at length by Siu and Kelly [8], can provide a reasonable estimate for the location of the posterior mode in many cases, as it provides the maximum likelihood estimates of the marginal likelihood function. In this example, because the first-stage gamma prior is conjugate to the Poisson likelihood, the marginal likelihood can be written in

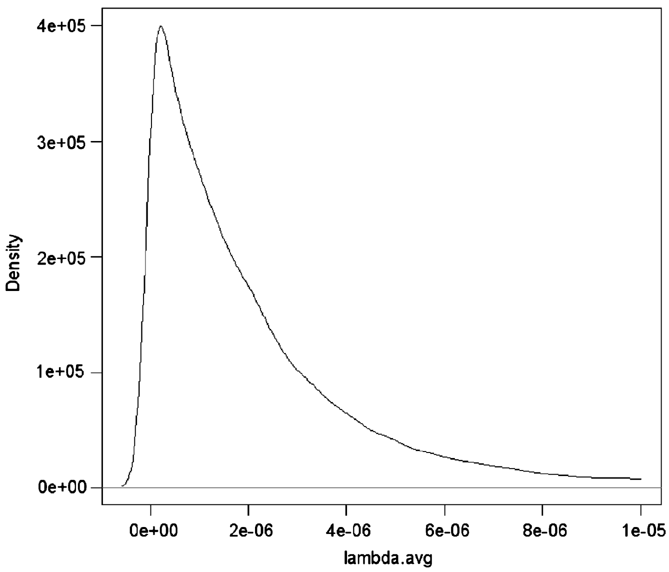

Fig. 2. Predictive distribution for λ in hierarchical Bayes example, representing population variability in λ.

closed form

$$\begin{aligned}L(\alpha,\beta) &= \prod_i \int_0^\infty \frac{(\lambda_i t_i)^{x_i} \mathrm{e}^{-\lambda_i t_i}}{x_i!} \frac{\beta^\alpha \lambda_i^{\alpha-1} \mathrm{e}^{-\beta\lambda_i}}{\Gamma(\alpha)} \\ &= \prod_i \frac{\Gamma(\alpha+x_i)}{x_i!\Gamma(\alpha)} \left(\frac{t_i}{\beta}\right)^x (1+t_i/\beta)^{-(\alpha+x_i)} \end{aligned} \qquad (7)$$

The natural logarithm of Eq. (7) is maximized numerically to find estimates of α and β, and these estimates are used to approximate the mode of the posterior distribution for α and β. Initial values for the Markov chains are then dispersed around this approximate mode. For this example, we find the empirical Bayes estimates of α and β to be about 1.1 and 500,000 h.

The model in Table 2 was run with 1000 burn-in iterations, followed by 100,000 iterations for each chain to estimate parameter values. The posterior predictive distribution for λ is shown in Fig. 2. The mean of this distribution is 2.5×10^{-6}/h, with a 90% credible interval of $(3.9 \times 10^{-8}, 8.0 \times 10^{-6})$. Note that this is a symmetric interval from the posterior distribution, as opposed to a highest posterior density (HPD) interval. Symmetric intervals are reported in this work because they are more typically used in risk assessment applications. HPD intervals are not available directly in WinBUGS, but can be easily obtained, if desired, with the R package [2]. Fig. 3 shows credible intervals for λ for each of the 12 sources.

As a further example, we revisit the emergency diesel generator (EDG) example from Section 6 of [8]. In that earlier work, the authors did not carry out hierarchical Bayes analysis. The problem is very similar in structure to the one above, with the exception that the aleatory model is now a binomial distribution. The first-stage prior is a beta-distribution with parameters α and β. For the two-stage Bayes calculation, Siu and Kelly [8] used a noninformative hyperprior proportional to $(\alpha+\beta)^{-2.5}$. For this analysis, we exploit the fact that α and β are independent until the data are observed, allowing us to write the joint prior as the product of independent diffuse gamma-hyperpriors priors. The WinBUGS script shown in Table 3 is used to carry out the hierarchical Bayes analysis. The marginal posterior distribution for p for EDG #1 is summarized in Table 4, along with the results from [8]. The results from the hierarchical Bayes analysis are generally comparable, but with somewhat wider uncertainty bounds. Also, as noted by Siu and Kelly [8], α and β are highly correlated in the posterior distribution: the rank correlation coefficient calculated by WinBUGS is 0.98. Note that this correlation is automatically accounted for in the MCMC sampling process in WinBUGS.

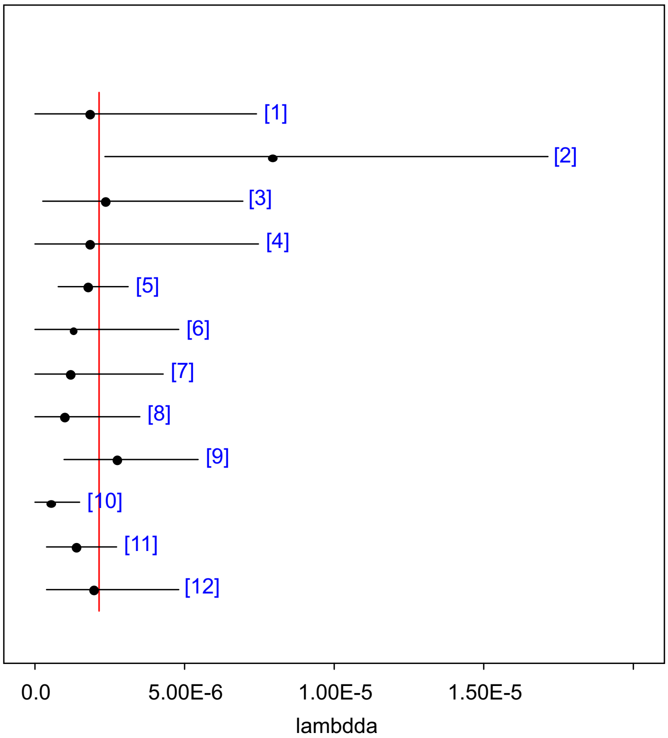

Fig. 3. 95% credible intervals for λ for each of the 12 sources in hierarchical Bayes example, obtained by updating Jeffreys prior for each source. Dots are posterior means for each interval, red line is average of posterior means.

Table 3
WinBUGS script for hierarchical Bayes analysis of EDG example from [8]

```
model {
for (i in 1 : N) {
x[i]~dbin(p.fts[i], n[i]) #Binomial dist. for EDG failures
p.fts[i]~dbeta(alpha, beta) #Beta prior for FTS probability
}
p.star~dbeta(alpha, beta) #Posterior predictive distribution for p
alpha~dgamma(0.0001, 0.0001) #Vague hyperprior for alpha
beta~dgamma(0.0001, 0.0001) #Vague hyperprior for beta
}
```

Table 4
Results for EDG example from [8]

	5th	50th	95th	Mean
Empirical Bayes	4.7E–04	4.4E–03	1.7E–02	5.9E–03
Two-stage Bayes	1.2E–04	3.3E–03	1.8E–02	5.2E–03
Hierarchical Bayes	5.9E–05	4.5E–03	1.9E–02	6.3E–03

4. Time-dependent reliability

4.1. Modeling time trends

It is sometimes the case that the usual Poisson and binomial models are rendered invalid because the parameter of interest (λ or p, respectively) is not constant over time. In the US for example, recent data analysis in [18] has suggested decreasing values of λ and p for several important components. As an example, let us examine valve leakage data from [19]. These data are shown in Table 5.

We first carry out a qualitative check to see if there appears to be any systematic time trend in p. To do this, we update the Jeffreys prior with the data for each year, and plot the interval estimates obtained side by side. The Jeffreys prior is used because we want the resulting intervals to be driven by the observed data. Note that these are 95% intervals, as this is the coverage produced by WinBUGS, and cannot be changed easily by the user. The result is shown in Fig. 4. The graph appears to indicate an increasing trend with time, but significant uncertainty in the individual estimates clouds this conclusion. Therefore, a quantitative approach is needed.

A generalized linear model (GLM) is often used to model a time trend in p or λ. Various link functions can be used in such a model, but a standard choice for p, as suggested by Atwood et al. [10], is the logit function. In this model, logit(p), which is the log of the odds ratio, is defined to be a linear function of time

$$\log\left(\frac{p}{1-p}\right) = a + bt \tag{8}$$

Note that $b = 0$ in this model corresponds to no trend. If p is increasing over time, then we will have $b>0$. The WinBUGS script for this model is shown in Table 6. We have used diffuse priors over the real axis for a and b; WinBUGS refers to this distribution as dflat(). This is an improper prior; for this type of problem, the posterior distribution will be proper and the estimates of a and b will be numerically similar to frequentist estimates, but in general one should be careful when using the dflat() prior, as it may lead to an improper posterior distribution. Such a problem may be

Table 5
Valve leakage data from [19]

Year	Number of failures	Demands
1	4	52
2	2	52
3	3	52
4	1	52
5	4	52
6	3	52
7	4	52
8	9	52
9	6	52

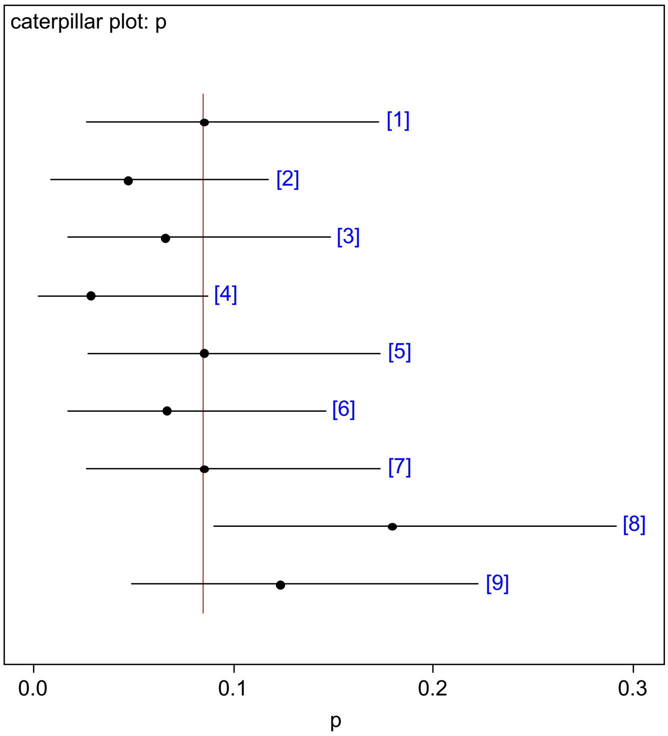

Fig. 4. 95% posterior credible intervals for valve leakage probability over time, obtained by updating Jeffreys prior in each year. Dots are posterior means for each interval, red line is average of posterior means.

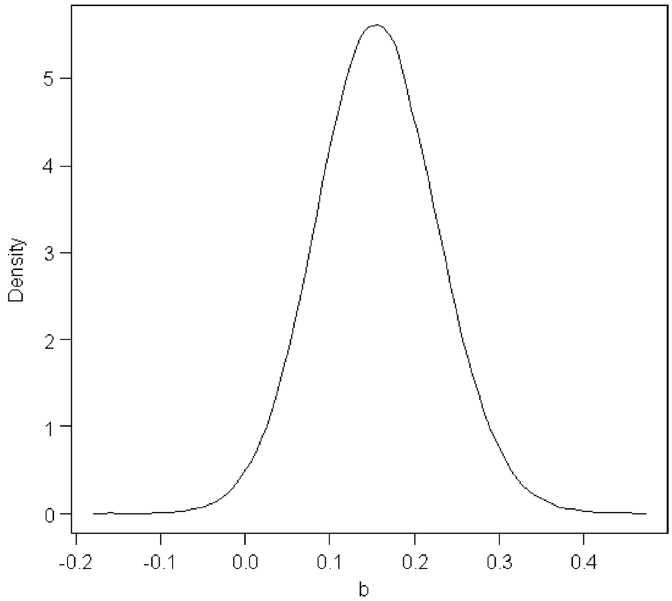

Fig. 5. Marginal posterior density for trend parameter in logit model for p. Density is highest for $b>0$, suggesting an increasing trend in p.

Table 6
WinBUGS script for modeling time trend in p of binomial distribution

```
model
{
for (i in 1:N) {
    x[i]~dbin(p[i], n[i]) #Binomial distribution for failures in
  each year
    logit(p[i]) <- a+b*i #Use of logit() link function for p[i]
    }
    a~dflat() #Diffuse prior for a
    b~dflat() #Diffuse prior for b
}
```

indicated by convergence problems and can often be ameliorated by switching to a normal distribution with a mean of zero and a very small precision, such as 10^{-6}.

The marginal posterior distribution for b is shown in Fig. 5, which illustrates that the probability that $b>0$ is very high, suggesting an *increasing* trend in time for p.

Fig. 6 shows the posterior means and credible intervals for p in each year, including a prediction for year 10, which could be used in a risk assessment. For comparison, treating p as constant over time and updating a Jeffreys prior with the pooled data from all 9 yr would give a posterior mean of 0.08, with a 90% credible interval of (0.06, 0.099).

A similar approach can be taken for Poisson failure rates, where the standard link function for a GLM is $\log(\lambda)$. The details of implementation are not described in this paper, but are presented in [10].

4.2. Modeling failure with repair

Much past work has been devoted to modeling failures with repair under the assumption that the repairs restore the component to "same as new" condition, that is, under the assumption that the stochastic point process being observed is a renewal process. Disproportionately less work has addressed the perhaps more reasonable assumption that repairs make the component "same as old." Furthermore, most of this work has been devoted to qualitative analysis and frequentist estimation. Under the assumption of a renewal process, the times between failures (inter-arrival times) are independently and identically distributed (iid), and this makes the statistical analysis

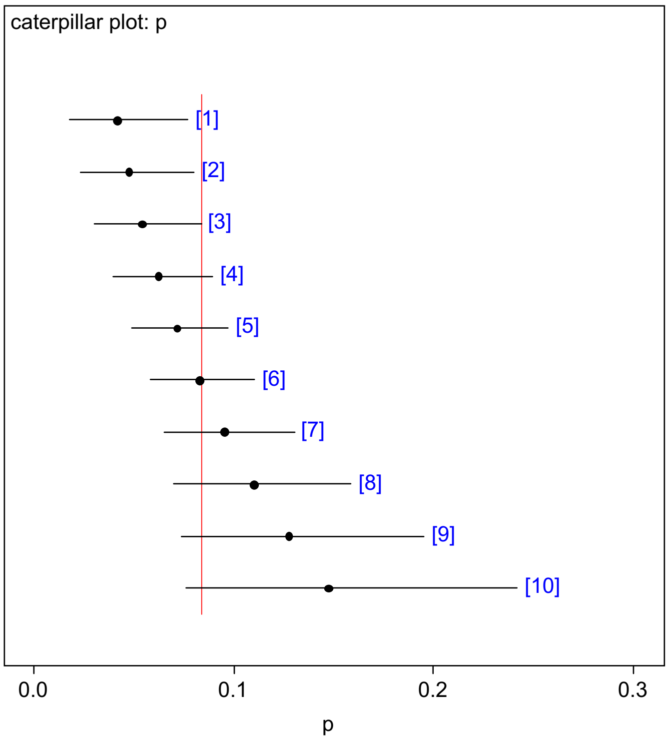

Fig. 6. Posterior mean and 95% credible interval for p in each year, from update of Jeffreys prior (year 10 is predicted values). Dots are posterior means for each interval, red line is average of posterior means.

straightforward. However, under the "same as old" assumption for repair, the inter-arrival times are not iid; the distribution for the ith time is dependent upon t_{i-1}. See [20] for further discussion.

Under a homogeneous Poisson process (HPP), the number of failures, x, in time t is described by a Poisson distribution

$$f(x) = \frac{(\lambda t)^x e^{-\lambda t}}{x!} \tag{9}$$

For the HPP, λ is independent of time, and the expected number of failures in time t is given by λt.

Relaxing the assumption of constant λ leads to the nonhomogeneous Poisson process (NHPP). The number of failures in time t is still Poisson-distributed, but the expected number of failures in any given time interval, $[t_1, t_2]$, is given by

$$\int_{t_1}^{t_2} \lambda(t)\,dt \tag{10}$$

If $\lambda(t)$ is increasing with time, the times between failures are decreasing with time; the component is aging or wearing out. Conversely, if $\lambda(t)$ is decreasing with time, the times between failures are increasing with time, and the component is experiencing reliability growth.

The functional form of $\lambda(t)$ must be specified in order for parametric analysis to proceed. Common forms for $\lambda(t)$ used for hardware include the power-law process

$$\lambda(t) = \frac{\alpha}{\beta}\left(\frac{t}{\beta}\right)^{\alpha-1}, \tag{11}$$

the loglinear model

$$\lambda(t) = \exp(a + bt), \tag{12}$$

and the linear model

$$\lambda(t) = a + bt \tag{13}$$

The NHPP process has also been used to model software reliability growth during development and testing. The literature on applications of the NHPP process to software reliability is vast; a good introduction is provided in [21] and a more recent reference is [22].

For this paper, we will focus on the power-law process for hardware reliability. The power-law process has been the subject of some past analysis and subsumes both the constant model ($\alpha = 1$) and the linear model ($\alpha = 2$). The time to first failure for the power-law process has a Weibull distribution with shape parameter α and scale parameter, β

$$f(t_1) = \frac{\alpha}{\beta}\left(\frac{t_1}{\beta}\right)^{\alpha-1} \exp[-(t_1/\beta)^{\alpha}] \tag{14}$$

For this reason, the power-law process is sometimes referred to as a Weibull process. This name is unfortunate in that it can lead to the mistaken notion that a sample of inter-arrival times (or in some cases the times themselves!) from a power-law process is an iid sample from a Weibull(α,β) distribution. As pointed out above, this assumption is only valid if one is observing a renewal process.

Relatively little work has been done on Bayesian analysis of a power-law process. Notable references are [23–25]. A reason for the dearth of work in this area may be the relative intractability of the Bayesian approach without MCMC. As noted by Guida et al. [23], "[Bayesian procedures] are computationally much more onerous than the corresponding maximum likelihood ones, since they in general require a numerical integration." This problem has been obviated by the advent of MCMC techniques and software for implementing such approaches.

We will analyze the case in which the observation process is failure-truncated. The alternative, in which the observations stop after a fixed time, is straightforward to analyze in a similar manner. As pointed out above, the time to first failure has a Weibull(α,β) distribution, given in Eq. (14). For $i = 2, \ldots, n$, we must use the condition that the failure times are ordered

$$f(t_i|t_{i-1}) = f(t_i|T_i > t_{i-1}) = \frac{f(t_i)}{Pr(T_i > t_{i-1})} \tag{15}$$

This is a truncated Weibull distribution. Thus, for $i = 2, \ldots, n$, we have

$$f(t_i|t_{i-1}) = \frac{\alpha}{\beta^{\alpha}}(t_i)^{\alpha-1} \exp\left[-\left(\frac{t_i}{\beta}\right)^{\alpha} + \left(\frac{t_{i-1}}{\beta}\right)^{\alpha}\right] \tag{16}$$

Therefore, the likelihood function becomes

$$\begin{aligned} f(t_1, t_2, \ldots, t_n|\alpha,\beta) &= f(t_1)\prod_{i=2}^{n} f(t_i|t_{i-1}) \\ &= \frac{\alpha^n}{\beta^{n\alpha}}\left(\prod_{i=1}^{n} t_i^{\alpha-1}\right)\exp\left[-\left(\frac{t_n}{\beta}\right)^{\alpha}\right] \end{aligned} \tag{17}$$

Strictly for comparison purposes it is worth noting that the maximum likelihood estimate (MLE) for α is given by

$$\hat{\alpha} = \frac{n}{\sum_{i=1}^{n}\ln(t_n/t_i)} \tag{18}$$

As pointed out by [26], the MLE for α is biased; an unbiased estimate is given (for the failure-truncated case) by

$$\tilde{\alpha} = \frac{n-2}{n}\hat{\alpha} \tag{19}$$

This bias becomes important for small sample sizes (small n).

WinBUGS was used to perform the MCMC sampling from the joint posterior distribution of α and β and to obtain marginal posterior distributions and summary statistics. However, the likelihood function given by Eq. (17) is not pre-programmed into WinBUGS, but the user can define a new aleatory model in WinBUGS by creating a vector of size n, with every component equal to zero. This vector is assigned a generic distribution with parameter, ϕ. Defining $\phi = \log(\text{likelihood})$ allows WinBUGS to update the parameters in the likelihood function.

Independent, diffuse priors were used for α and β. Diffuse priors were used strictly to allow the results to be compared with the MLEs; a strength of the Bayesian approach is that it allows information about parameters to be encoded into a joint prior distribution for α and β, if such information is available. For β, which is a scale parameter determined by the units of time in the problem, a gamma(10^{-4}, 10^{-4}) prior was chosen. For α, which is a shape parameter, one might think that a uniform distribution over a reasonable range (a range of 0.3–3 is suggested by Guida et al. [23]) would be appropriate. However, it is often the case that the marginal posterior distribution for α is very sensitive to the upper limit in the uniform prior. For this reason, a gamma(10^{-4}, 10^{-4}) prior was used for α, also. The WinBUGS script is shown in Table 7.

As data, we will use the failure times given for the "sad," "happy," and "noncommittal" systems given on pp. v–vi of [20]. We will present the marginal posterior distributions and summary statistics for α and then compare these results to the MLEs.

For the "sad" system the posterior mean of α is 2.92, with a 90% credible interval of (1.27, 5.13). For reference, the MLE of α is 3.4, larger than the posterior mean. However, the bias-corrected estimate of α is 2.4, less than the posterior mean. A 90% confidence interval for α is (1.6, 5.1), a bit narrower than the 90% credible interval from the Bayesian analysis. The marginal posterior distribution of α is shown in Fig. 7. The probability that $\alpha > 1$, implying that λ is increasing with time, is near unity, confirming the "sad" nature of the system.

For the "happy" system (which is the "sad" system with the inter-arrival times in reverse order), the posterior mean of α is

Table 7
WinBUGS script for analyzing failures with repair using power-law process

```
model {
for(i in 1:N) {
zeros[i] <- 0
   zeros[i] ~dgeneric(phi[i])
   #phi[i] = log(likelihood)
   }
#Power-law model (failure-truncated)
   for(j in 1:N) {
   phi[j] <- log(alpha)-alpha*log(beta)+(alpha-1)*log(t[j])-
   pow(t[N]/beta, alpha)/N }
alpha~dgamma(0.0001, 0.0001)
beta~dgamma(0.0001, 0.0001)
}

Inits
list(alpha = 1, beta = 100) #Inits for power-law model
list(alpha = 2, beta = 500)
list(alpha = 0.5, beta = 250)
```

Table 8
Times to recover offsite ac power for grid-related disturbances, taken from [29]

Site	Date	Potential recovery time (min)
Davis-Besse	14/8/2003	657
Fermi	14/8/2003	384
Fitzpatrick/nine mile point 1	14/8/2003	142
Ginna	14/8/2003	54
Indian point	16/6/1997	42
Indian point	14/8/2003	102
Nine mile point 2	14/8/2003	110
Palo verde	14/6/2004	37
Peach bottom	15/9/2003	16
Perry	14/8/2003	87
Summer	11/7/1989	100
Vermont yankee	17/8/1987	17

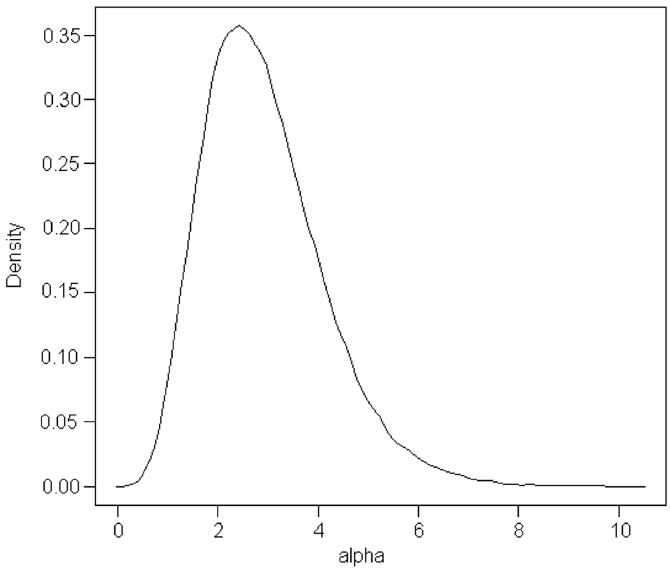

Fig. 7. Marginal posterior density of α for "sad" system.

0.61, with a 90% credible interval of (0.28, 1.06). The probability that $\alpha<1$, implying that λ is decreasing with time, is quite large. For the "noncommittal" system, the posterior mean of α is 1.09, with a 90% credible interval of (0.48, 1.94).

5. Modeling random durations

There are many instances in which time is the random variable of interest. For example, a facility may be threatened by fire, and the time of interest is the time needed to suppress the fire, which must be shorter than the time to damage vital equipment (which is also a random variable). Another important application for many modern facilities is inference on the time needed to restore ac power once it has been lost. And of course such models can be used for the renewal process described above.

The simplest stochastic model for applications where time is the random variable of interest is the exponential distribution, with density function and cumulative distribution function given by

$$f(t) = \lambda e^{-\lambda t}$$
$$F(t) = Pr(T \leqslant t) = 1 - e^{-\lambda t} \tag{20}$$

This is the model used for fire suppression time in the Nuclear Regulatory Commission's recent guidance for fire PRA [27]. However, there are numerous applications in which this model, with its assumption of time-independent rate (of suppression, recovery, etc.) is not realistic. For example, past work described in [28,29] has shown that the rate of recovering offsite ac power at a commercial nuclear plant is often a decreasing function of time after power is lost. Therefore, the exponential distribution is not usually an appropriate model for this case, and the analyst is led to models that allow for time-dependent recovery rates, such as the Weibull or lognormal distribution.

Bayesian inference is more complicated when the likelihood function is other than exponential, and much past work has been done on various approximate approaches for these cases. In the past, the difficulty of the Bayesian approach has led analysts to use frequentist methods, such as MLEs, the approach adopted for example in both [28,29]. Today, however, WinBUGS allows a fully Bayesian approach to the problem to be implemented.

5.1. Recovery of offsite ac power

To illustrate the use of WinBUGS for this type of problem, we will examine offsite power recovery data from a recently published analysis done for the USNRC [29]. As pointed out above, the analysis of power recovery in this work was not Bayesian, but relied on maximum likelihood estimation to obtain estimates of the parameters of the Weibull and lognormal distributions that were candidates for modeling recovery time. To illustrate the modern Bayesian approach to this analysis, we will reanalyze a portion of the data, shown in Table 8. These are the times to recover offsite power for losses of ac power caused by problems in the offsite grid.

From Ref. [29], we found the following MLEs for this set of recovery times:

Weibull: α (shape) = 0.929, β (scale) = 2.332 h
Lognormal: $\mu = 0.300$, $\sigma = 1.064$

The script for the Weibull analysis is shown in Table 9. Note that WinBUGS parameterizes the Weibull distribution in terms of a shape parameter, α, and a scale parameter, λ. The scale parameter used by WinBUGS is related to the "standard" scale parameter by the following equation:

$$\lambda = \beta^{-\alpha} \tag{21}$$

Table 9
WinBUGS script for estimating parameters of Weibull distribution for grid-related offsite power recovery times

```
model {
for(i in 1:N){
   duration.hr[i] <- duration[i]/60
   duration.hr[i] ~dweib(alpha, lambda) #Weibull model for
   duration
   }
beta <- pow(lambda, -1/alpha) #''Standard'' scale parameter
alpha ~dgamma(0.0001, 0.0001)
lambda ~dgamma(0.0001, 0.0001)
}

list(alpha = 1, lambda = 0.1)
list(alpha = 0.5, lambda = 0.5)
list(alpha = 2, lambda = 1)
```

Table 10
Posterior summaries for Weibull model of offsite power recovery

Parameter	Posterior mean	90% credible interval
α	0.914	(0.812, 1.254)
β	2.58	(1.343, 4.299)

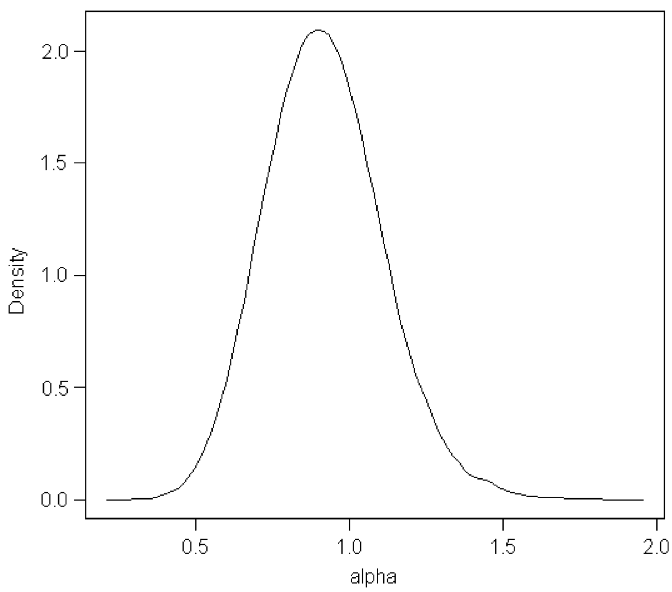

Fig. 8. Marginal posterior density for shape parameter of Weibull distribution used to analyze grid-related offsite power recovery times.

We used independent, diffuse gamma(10^{-4}, 10^{-4}) priors for both α and λ. Diffuse priors were chosen to allow the numerical results to be compared with the MLEs obtained by [29].

One thousand iterations were used to burn-in the three chains (three chains were used to accommodate the three possible general choices for α: $\alpha<1$, $\alpha=1$, and $\alpha>1$, and then another 10,000 iterations for each chain were used to obtain the joint posterior distribution of α and β. The posterior means and 90% credible intervals for α and β are shown in Table 10.

Note that the posterior distribution of α, which is shown in Fig. 8, has significant probability mass centered about 1.0, indicating that an exponential distribution (which the Weibull distribution reduces to when $\alpha = 1$) might be a reasonable model. We will return to this observation when we discuss Bayesian model validation.

Table 11
WinBUGS script for estimating parameters of lognormal distribution for grid-related offsite power recovery times

```
model {
   for(i in 1:N) {
   duration.hr[i] <- duration[i]/60
   duration.hr[i] ~dlnorm(mu, tau) #Lognormal model for duration
   }
mu ~dflat()
tau <- 1/pow(sigma, 2)
sigma ~dgamma(0.0001, 0.0001)

Inits
list(mu = 0, sigma = 2)
list(mu = 1, sigma = 1)
```

Table 12
Posterior summaries for lognormal model of offsite power recovery

Parameter	Posterior mean	90% interval
μ	0.299	(−0.269, 0.886)
σ	1.194	(0.833, 1.725)

The WinBUGS script used to analyze the lognormal model is shown in Table 11. Note that WinBUGS parameterizes the lognormal distribution in terms of the logarithmic mean, μ, and the logarithmic precision, τ. The precision is related to the logarithmic standard deviation, σ, by $\tau = \sigma^{-2}$.

One thousand iterations were used to burn-in the chains, and then another 20,000 iterations for each chain were used to obtain the joint posterior distribution of μ and σ. The posterior means and 90% credible intervals for μ and σ are shown in Table 12.

6. Treatment of missing and uncertain data

It is not uncommon in risk applications to encounter situations in which the observed data, which would normally enter into Bayes' theorem via the likelihood function, are either missing or the exact values are not known with certainty. For example, in the simple Poisson aleatory model used for initiating event frequencies and failure rates, one may not know the exposure time, t, accurately. Sometimes only an interval estimate is available. This is especially the case for failure rates, less often the case for initiating event frequencies. For example, in an example in which we are interested in plugging of service water strainers, we may have two parallel strainers in a continually operating system, but we do not know if both strainers are always in service. Thus, over an observation time of 24,140 h, our estimate of exposure time, t, could be as low as 24,140 and as high as 48,180 strainer-hours.

We can handle this type of uncertainty quite easily in the Bayesian framework. We simply assign t, a distribution that quantifies our available information, again reinforcing the idea that the Bayesian methodology encodes information via probability distributions. For example, we might assign t a uniform distribution between 24,140 and 48,180 h in our example, expressing indifference among values in this interval, and placing zero weight on values outside the interval. In finding the marginal distribution of λ, or calculating $\Pr(X = x)$, we have to average over this distribution. This makes the mathematics more complicated, since the posterior for λ is no longer a gamma-distribution and must be evaluated numerically, but with modern tools this is a straightforward calculation. This type of problem has been analyzed in [30,31].

Table 13
WinBUGS script for estimating strainer-plugging rate when exposure time is uncertain

```
model {
x~dpois(mu) #Poisson distribution for number of events, x
mu <- lambda*time.exp
lambda~dgamma(0.5, 0.0001) #Jeffreys prior for lambda, monitor
   this node
time.exp~dunif(24140, 48180) #Models uncertainty in exposure time
}
data
list(x = 4)
```

Table 14
WinBUGS script for estimating MOV failure probability when demands are uncertain

```
model {
x~dbin(p, N)
N~dunif(275, 440) #Models uncertainty in demands
p~dbeta(0.5, 0.5) #Jeffreys prior for p, monitor this node
}
data
list(x = 4)
```

The WinBUGS script used for this example is shown in Table 13. Assume we have observed four strainer-plugging events during an uncertain exposure time, which we model as being uniformly distributed between the lower limit of 24,140 and the upper limit of 48,180 strainer-hours. We use the Jeffreys prior distribution for λ. Running this script gives a posterior mean for λ of 1.3×10^{-4}/h and a 90% credible interval of $(4.5\times10^{-5}, 2.6\times10^{-4})$. Note that this is nearly what we would get by using the mean of the exposure time distribution (36,160 strainer-hours).

The same concept can be applied to the case of failure on demand, where the number of demands, n, may not be known accurately. Let us consider an example of motor-operated valves failing to open on demand. Let us assume that the number of demands is nominally 381, but could have been as high as about 440, and as low as 275. Again, let us assign n a uniform distribution over this range. We will use the Jeffreys prior for p. Assume that we have seen four failures to open.

With the uncertainty in the number of demands considered, our posterior for p is no longer a beta-distribution and must be described numerically with WinBUGS. Using the script shown in Table 14, the posterior mean of p is estimated to be 0.01, the same value to 2 decimal places we would have obtained with the demands treated as known (381). Again, because our central estimate of n is near the fixed value of 381, modeling the uncertainty in n does not affect the posterior distribution of p significantly. In many cases, we may find that modeling the uncertainty in n does not affect our estimate of p very much.

There are also occasions where the event or failure count is uncertain. For example, in examining plant records, one may not be able to tell if a particular event constituted failure of equipment with respect to its function in the PRA model. Within the Bayesian framework, one can assign a subjective distribution to the event or failure counts, and the posterior distribution for the parameter of interest is then averaged over this distribution, giving the marginal posterior distribution for the parameter of interest. This is called the posterior-averaging approach in [8] and has been illustrated for Poisson event counts by [30,31]. Alternative approaches to Bayesian inference with uncertain event or failure counts have been proposed, based on modifications to the likelihood function. We do not discuss these in detail here. See [32,33] for more details on these approaches.

Under the posterior-averaging approach, the marginal posterior distribution for λ becomes, considering both uncertainty in x and t

$$g_{\mathrm{avg}}(\lambda)=\sum_{i=1}^{N}\left[\int_{t_{\mathrm{lower}}}^{t_{\mathrm{upper}}}\frac{f(x_i|\lambda,t)g(\lambda)}{\int_0^\infty\int_{t_{\mathrm{lower}}}^{t_{\mathrm{upper}}}f(x_i|\lambda,t)g(\lambda)\pi(t)\,\mathrm{d}t\,\mathrm{d}\lambda}\pi(t)\,\mathrm{d}t\right]\Pr(x_i) \tag{22}$$

The equation for the binomial case is similar.

As a first example, consider strainer plugging at Plant Y again. Earlier, the observed number of failures was treated as known (four). Now assume that plant records described seven plugging events over the time period of interest, but it was unclear if three of these events would have been considered as plugging events from the perspective of the PRA model. Therefore, the actual number of plugging events is uncertain, and could be four, five, six, or seven. In the posterior-averaging approach, the analyst assigns a subjective (discrete) distribution to these values, representing his confidence in the correctness of each value. Assume the analyst, in poring over the plant records, has arrived at the following distribution for the observed data:

$\Pr(x=4)=0.75;$
$\Pr(x=5)=0.15;$
$\Pr(x=6)=0.075;$
$\Pr(x=7)=0.025.$

Note that these probabilities must sum to unity. WinBUGS is used to analyze this problem numerically, via the script shown in Table 15. For the case where the exposure time is known with certainty to be 48,180 h, the posterior mean is $1.0\times10^{-4}\,\mathrm{h}^{-1}$, with 90% credible interval $(3.7\times10^{-5}/\mathrm{h}, 1.9\times10^{-4}/\mathrm{h})$. If we include the uncertainty in the exposure time, we find the posterior mean to be 1.5×10^{-4}/h, with 90% credible interval $(5.2\times10^{-5}/\mathrm{h}, 3.0\times10^{-4}/\mathrm{h})$.

As a second example, consider MOV failures. Earlier, we took the observed number of failures as four. Consider now the case where this value is uncertain, and assume it could be 3, 4, 5, or 6, with $\Pr(3)=0.1$, $\Pr(4)=0.7$, $\Pr(5)=0.15$, and $\Pr(6)=0.05$. Using the WinBUGS script shown in Table 16, for the case where the number of demands is known with certainty to be 381, we obtain a posterior mean of 0.01, with 90% credible interval (0.004, 0.02). If the uncertainty in the number of demands is included as before, the posterior mean is still 0.01 to 2 decimal places, and the 90% credible interval has shifted slightly to (0.005, 0.03).

Table 15
WinBUGS script for estimating strainer plugging by averaging posterior distribution over uncertainty in event count and exposure time

```
model {
for(i in 1:N) {
   x[i]~dpois(mu[i])
   mu[i] <- lambda[i]*time.exp
   lambda[i]~dgamma(0.5, 0.0001) #Jeffreys prior for lambda
   }
lambda.avg <- lambda[r] #Overall composite lambda, monitor this
   node
r~dcat(p[])
time.exp ~dunif(24140, 48180) #Models uncertainty in exposure time
}
data
list(x = c(4,5,6,7), p = c(0.75, 0.15, 0.075, 0.025))
```

Table 16
WinBUGS script for estimating MOV failure probability by averaging posterior distribution over uncertainty in failure count and number of demands

```
model {
   for (i in 1:N){
   x[i]~dbin(p[i], D)
   p[i]~dbeta(0.5, 0.5) #Jeffreys prior
   }
p.avg <- p[r] #Composite posterior, monitor this node
r~dcat(q[])
D~dunif(275, 440) #Models uncertainty in demands
}
data
list(x = c(3, 4, 5, 6), q = c(0.1, 0.7, 0.15, 0.05))
```

7. Bayesian regression models

Regression models are commonplace in statistics, and have been applied in risk analysis, as well. An example discussed earlier is the use of a logit or loglinear model for trend in a binomial or Poisson parameter. In these models, time was the predictor variable. In human reliability analysis (HRA), regression models have been used to estimate human error probabilities. The original SLIM (see [34]) and its derivatives are examples. We will examine a simple example from the published literature. Another recent example for estimating early system reliability is [35].

Dalal et al. [36] performed a frequentist regression analysis of space shuttle field and nozzle O-ring data collected prior to the ill-fated launch of the Challenger in January 1986. The purpose of their analysis was to show how regression analysis could be used to provide information to decision-makers prior to a launch, information that could have been expected to lead to a decision to scrub the Challenger launch due to the low temperatures (~31 °F.) present at the launch pad on the morning of the scheduled launch. The analysis in [36] found that a logistic regression model provided a relatively good fit to the past data.

In the second portion of the analysis in [36], parameter uncertainties were propagated through the fitted logistic regression model in order to estimate the probability of shuttle failure due to O-ring failure at the estimated launch temperature of ~31 °F. Because the analysis was frequentist in nature, probability distributions representing epistemic uncertainty in the input parameters were not available, and the authors had to resort to an approximate approach based on bootstrap confidence intervals, an approach developed by Efron [37].

Ref. [38] repeats the analyses of [36] from a Bayesian perspective. MCMC sampling was used to sample from the joint posterior distribution of the model parameters, and to sample from the posterior predictive distributions at the estimated launch temperature, a temperature that had not been observed in prior launches of the space shuttle. Uncertainties, which are represented by probability distributions in the Bayesian approach, were propagated through the model via Monte Carlo sampling to obtain a probability distribution for O-ring failure, and subsequently for shuttle failure as a result of O-ring failure. No approximations are required in the Bayesian approach and the resulting distributions could be input to a decision analysis to calculate expected utility for the decision to launch.

7.1. Stochastic model for O-ring distress

There are six O-rings on the shuttle, so during each launch, the number of distress events, defined as erosion or blow-by of a primary field O-ring, is modeled as binomial with parameters p and 6: $X \sim \text{binomial}(p, 6)$. Data from launches prior to the Challenger (taken from [36]) are shown in Table 17.

Table 17
Space shuttle field O-ring thermal distress data, taken from [36]

Flight	Distress events	Temp (°F.)	Press (p sig)
1	0	66	50
2	1	70	50
3	0	69	50
5	0	68	50
6	0	67	50
7	0	72	50
8	0	73	100
9	0	70	100
41-B	1	57	200
41-C	1	63	200
41-D	1	70	200
41-G	0	78	200
51-A	0	67	200
51-C	2	53	200
51-D	0	67	200
51-B	0	75	200
51-G	0	70	200
51-F	0	81	200
51-I	0	76	200
51-J	0	79	200
61-A	2	75	200
61-B	0	76	200
61-C	1	58	200

In the model used in [36], p is a function of temperature and leak-test pressure. The standard link function is the logit function

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right) \tag{23}$$

Following the approach in [36,38] considered two potential explanatory regression models:

(1) $\text{logit}(p) = a + b \times \text{temp} + c \times \text{press}$
(2) $\text{logit}(p) = a + b \times \text{temp}$

The WinBUGS script for the first model, which includes both temperature and pressure as explanatory variables, is shown in Table 18. Diffuse priors were used for the coefficients in this model to allow the numerical results to be compared with the maximum likelihood estimates and confidence intervals obtained by [36].

One thousand burn-in iterations were required for convergence, followed by 500,000 iterations for each of two chains to estimate the parameters. Table 19 shows the posterior mean, standard deviation, and symmetric 95% credible interval for each of the parameters in the logistic regression model for p (95% intervals are used because (36) presented 95% confidence intervals).

Note the very large standard deviation on the intercept parameter. Also, when plotted, the marginal posterior distributions for a and b appear approximately normal with the listed posterior means and standard deviations. The marginal posterior distribution for the pressure coefficient, because it is centered about zero, indicates that pressure is not a significant explanatory variable.

The model that includes both temperature and pressure predicts about four distress events at 31 °F, the approximate temperature for the disastrous launch of the Challenger.

The model with only temperature predicts essentially the same number of events as the model with both temperature and pressure. In a later section, we will illustrate a method for

Table 18
WinBUGS script for logistic regression of primary O-ring distress on temperature and pressure

```
model {
for(i in 1:K){
   distress[i] ~dbin(p[i], 6)
logit(p[i]) <- a+b*temp[i]+c*press[i] #Model with temperature and
   pressure
   distress.rep[i] ~dbin(p[i], 6) #Replicate values for model
   validation
   diff.obs[i] <- pow(distress[i]-6*p[i], 2)/(6*p[i])
   diff.rep[i] <- pow(distress.rep[i]-6*p[i], 2)/(6*p[i])
   }
distress.31 ~dbin(p.31, 6) #Predicted number of distress events for
   launch 61-L
logit(p.31) <- a+b*31+c*200
#Prior distributions
a~dnorm(0, 0.000001)
b~dnorm(0, 0.000001)
c~dnorm(0, 0.000001)
}
```

Table 19
Summary posterior estimates of logistic regression parameters, temperature and pressure included as explanatory variables

Parameter	Mean	Standard dev.	95% credible interval
a (intercept)	2.24	3.74	(−4.71, 9.92)
b (temp. coeff.)	−0.105	0.05	(−0.20, −0.02)
c (press. coeff.)	0.01	0.009	(−0.004, 0.03)

choosing between a model with only temperature as an explanatory variable, versus a model with both temperature and pressure.

8. Bayesian model validation

The frequentist approach to model checking or validation typically involves comparing the observed value of a test statistic to percentiles of the sampling distribution for that statistic. Given that the null hypothesis is true, we would not expect to see "extreme" values of the test statistic. One Bayesian approach to model checking involves calculating the posterior probability of the various hypotheses and choosing the one that is most likely. One can also use summary statistics derived from the posterior predictive distribution, as described by Gelman et al. [39]. We will discuss two such statistics: Bayesian χ^2 and Cramer-von Mises, which lead to a Bayesian analog of p-value. We will also discuss the deviance information criterion (DIC), a Bayesian analog of a penalized likelihood measure employed by frequentist statisticians. Other Bayesian approaches to Bayesian model validation have been proposed. Some of these can be found in [40–42].

Both of the summary statistics we will use are based on the posterior predictive distribution. This is the predictive distribution for future values of an observed random variable, given past empirical data. It is defined as

$$\pi(\tilde{x}|x) = \int_{\Theta} f(\tilde{x}|\theta)\pi_1(\theta|x)\,\mathrm{d}\theta \tag{24}$$

Similarly, one can define a prior predictive distribution, in which the likelihood is averaged over the prior distribution for θ. The prior predictive distribution can help in selecting an appropriate prior distribution, via so-called preposterior analysis. In this approach, the probabilities of hypothetical sets of data are calculated. If the probability of an expected set of data is low, then the prior distribution leading to that probability is questioned.

The posterior predictive distribution can be used in a similar fashion after data have been observed. If the observed data are extreme values from a tail of the posterior predictive distribution, this indicates that the "model" (prior and likelihood) is not able to replicate the observed data very well, and suggests a potential problem with the prior distribution, the likelihood function, or both.

We illustrate this use of predictive distributions with an example related to frequency of a particular initiating event. The usual stochastic model for the occurrence of initiating events is that the number of such events, X, in a specified time period, t, has a Poisson distribution with parameter λt, where λ is the initiating event frequency, with λ being the unknown parameter. Thus, the probability of seeing a specific outcome, such as one event in 10 yr, or at least two events in 25 yr, is obtained from the Poisson distribution function

$$\Pr(X = x|\lambda) = \frac{(\lambda t)^x \mathrm{e}^{-\lambda t}}{x!}, \quad x = 0, 1, \ldots \tag{25}$$

Note that this probability is conditional upon a value of the unknown initiating event frequency, λ. Strictly as an aid in illustrating the calculations that follow, we will use a conjugate gamma-distribution to describe our uncertainty about λ. A gamma-distribution has probability density function

$$g(\lambda) = \frac{\beta^{\alpha}\lambda^{\alpha-1}\mathrm{e}^{-\beta\lambda}}{\Gamma(\alpha)}, \quad \alpha > 0, \beta > 0 \tag{26}$$

Because the gamma-distribution is conjugate to the Poisson likelihood, a Bayesian update of x observed events in time t produces a posterior distribution that is gamma with parameters $\alpha_{\text{post}} = \alpha+x$ and $\beta_{\text{post}} = \beta+t$. The posterior mean is then $(\alpha+x)/(\beta+t)$. See [8] for additional details.

Let us look at a specific example for loss of main feedwater. Assume the prior distribution for this initiating event is gamma(1.2, 12 yr). Assume that we anticipate two losses of main feedwater in the next 2.7 yr. The prior mean frequency is 1.2/12 yr = 0.1/yr. Based on the prior distribution, there is a 90% probability that the frequency is in the range (0.008/yr, 0.28/yr). The sparse plant-specific data are suggestive of a frequency of 2.7/2 yr = 0.74/yr, well outside the 90% interval from the prior distribution. However, because the data are sparse, there is considerable uncertainty in the estimate.

Is it appropriate to use this prior distribution for the frequency of loss of main feedwater if we expect two such events in the next 2.7 yr? To answer this question, we use the prior predictive distribution of X, which gives the probability of seeing x events in time t, unconditional upon λ. In equation form, this can be shown to be

$$\Pr(X = x) = \frac{\Gamma(\alpha + x)}{x!\Gamma(\alpha)}\left(\frac{t}{\beta}\right)^x (1 + t/\beta)^{-(\alpha+x)} \tag{27}$$

Eq. (27) can be implemented in a spreadsheet or can be simulated with WinBUGS. The probability of interest for our example is the probability of seeing two or more losses of main feedwater in 2.7 yr, as this tells us how far out in the upper tail of the prior predictive distribution the expected failure count is located. If this probability is small, then we have evidence that the expected data are inconsistent with the proposed prior distribution for loss-of-main-feedwater frequency. Using either a spreadsheet or WinBUGS this probability is found to be 0.04. An often-used guideline is that this probability is "small" if it is less than 0.05. We are just under this guideline, so the expected data appear to be somewhat inconsistent with the proposed gamma(1.2, 12 yr) prior distribution.

Turning now to the posterior predictive distribution, what if we had used this prior distribution, and had then observed three events in 2.9 yr? A naïve update of the prior distribution would give a posterior mean frequency of 0.28/yr. However, it is prudent to ask if the resulting model (prior+likelihood) is reasonable, in the sense where we question whether the model can replicate the observed data with reasonable probability. The posterior predictive distribution is given by Eq. (27), with α and β replaced by α+x and β+t, the parameter values from the gamma posterior distribution for λ. A spreadsheet or WinBUGS can be used to calculate the probability of seeing at least three events in the next 2.9 yr. This probability is found to be 0.06, suggesting that the model is on the edge of predictive validity.

We next illustrate a more sophisticated use of the posterior predictive distribution for offsite power recovery time in our earlier example. We will use WinBUGS to generate replicate times from each of the three models under consideration, and then compare a summary statistic, denoted T_{rep}, for these replicated times to the same statistic for the observed times, denoted T_{obs}. If our model has predictive validity, then $P(T_{\text{rep}}>T_{\text{obs}})$ should be near 0.5; values near 0 or 1 indicate a model with poor predictive validity. More details can be found in [39,43,44].

One possible choice of summary statistic is the sum of the observed times. Unfortunately, with a Jeffreys prior and an exponential likelihood, it can be shown that the sum of the observed times, because it is a sufficient statistic for λ, will always be at the median of the distribution of the sum of the replicated times, regardless of the aleatory process that generates the observed times. Therefore, a different summary statistic is needed. We will use a Bayesian analog of the Cramer-Von Mises statistic [26], discussed and illustrated below.

8.1. Bayesian posterior predictive statistics and Bayesian p-value

In frequentist statistics, a commonly encountered test statistic is

$$\chi^2 = \sum_i \frac{(x_i - \mu_i)^2}{\sigma_i^2} \tag{28}$$

In this equation, x_i is the ith observed value, μ_i is the ith expected, or mean value, and $\sigma_i{}^2$ is the ith variance. The distribution of χ^2 is often approximately chi-square, with degrees of freedom related to the sample size. We use this as motivation for the following summary statistics.

We use the observed values of x to form the statistic

$$\chi^2_{\text{obs}} = \sum_i \frac{(x_{\text{obs},i} - \mu_i)^2}{\sigma_i^2} \tag{29}$$

We then generate replicate values of x from its posterior predictive distribution, and construct an analogous statistic:

$$\chi^2_{\text{rep}} = \sum_i \frac{(x_{\text{rep},i} - \mu_i)^2}{\sigma_i^2} \tag{30}$$

Both of these statistics, defined analogously to the frequentist χ^2 statistic, have a posterior distribution. χ^2_{obs} plays the role of the theoretical distribution in the frequentist setting, and χ^2_{rep} plays that of the summary statistic based on the data; in this case the "data" are replicate values from the Bayesian model (prior plus likelihood). In the frequentist setting, if the summary statistic calculated from the data is in a tail of the theoretical distribution, we are led to reject our model. The p-value is sometimes used to measure the degree to which the data are at conflict with the model. We will adopt that term here, and define the Bayesian p-value to be $\Pr(\chi^2_{\text{rep}} \geqslant \chi^2_{\text{obs}})$. However, instead of choosing an arbitrary p-value (e.g., 0.05) and rejecting a model with a p-value below the cutoff, we will use the p-value to select the model that is best at replicating the observed data. This will be the model whose Bayesian p-value is closest to 0.5. Note that, as pointed out by various authors, for example [40], the Bayesian p-value does not possess all of the properties of the corresponding frequentist p-value. For example, its distribution is not asymptotically uniform for a simple null hypothesis. However, from a practical perspective, the Bayesian p-value is useful for identifying poorly performing models and has the attractive property of being easy to calculate in the MCMC framework.

We first illustrate the posterior predictive approach to model validation with an application to our earlier example of modeling source-to-source variability in component failure rate, λ. There, we had 12 sources, as shown in Table 1. We will use the Bayesian analog of p-value to compare the model with variable failure rate to a simpler model in which each of the 12 sources is assumed to have the same failure rate. The portion of the WinBUGS script used to calculate the Bayesian p-value is shown in Table 20.

The marginal posterior distributions for χ^2_{obs} and χ^2_{rep} are shown in Figs. 9 and 10 for each model. There is significantly more overlap between χ^2_{obs} and χ^2_{rep} in the model with variable λ, indicating superior predictive validity of the variable-λ model.

Table 21 shows the Bayesian p-values (mean of node p.value) for the hierarchical (variable-λ) model and the constant-λ model. Based on these results, significantly better predictive validity is provided by the hierarchical model, which includes source-to-source variability in λ.

Applying this approach to the valve leakage data given in Table 5, the model with constant p across the 9 yr (no trend) gives a Bayesian p-value of 0.18. In comparison, the model that includes a monotonic trend in p over the 9 yr period has a Bayesian p-value of 0.47, significantly nearer 0.5 than the value of 0.18 for the model with no trend, suggesting that the model with an increasing trend in p has better predictive validity.

As a final example, we illustrate the use of a Bayesian analog of the Cramer-von Mises statistic [26] with the offsite power recovery data from our earlier example. The script implementing this test is shown in Table 22.

Running this script gives the results shown in Table 23. All of the models are quite good at replicating the observed data, with the lognormal distribution being the best (i.e., closest to 0.5) by a slight amount. Ref. [29] chose a lognormal distribution because it gave the best fit to the observed data, based on several different frequentist test statistics.

Table 20
WinBUGS script for calculating Bayesian p-value for 12 sources of component failure rate data

```
model {
for(i in 1:12) {
   x[i] ~dpois(mu[i]) #Poisson distribution for number of failures
   in each source
   mu[i] <- lambda[i]*t[i] #Model for source-to-source
   variability
   lambda[i] ~dgamma(alpha, beta) #Distribution for selecting
   lambda in each #source
   #lambda[i] <- lambda.constant #Model ignoring variability
x.rep[i] ~dpois(mu[i]) #Replicate value from posterior predictive
   distribution
   #Inputs to calculate Bayesian p-value
   diff.obs[i] <- pow(x[i]-mu[i], 2)/mu[i]
   diff.rep[i] <- pow(x.rep[i]-mu[i], 2)/mu[i]
   }
#Calculate Bayesian p-value
chisq.obs <- sum(diff.obs[])
chisq.rep <- sum(diff.rep[])
p.value <- step(chisq.rep-chisq.obs)
```

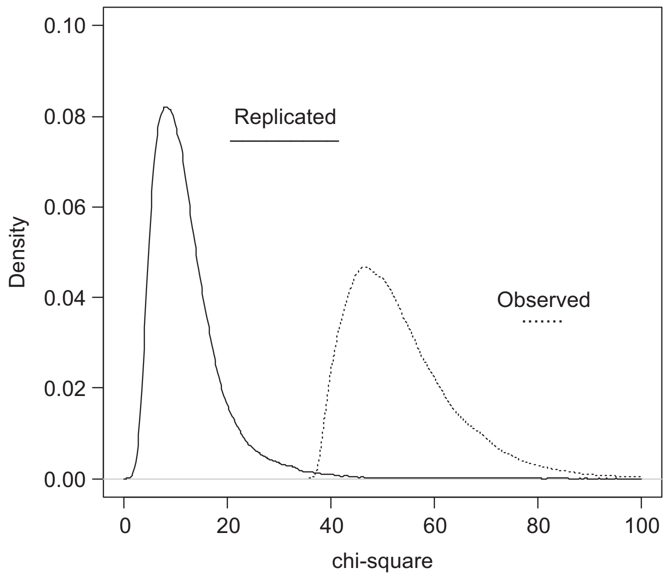

Fig. 9. Marginal posterior densities for observed and replicated Bayesian χ^2 statistics show little overlap in the constant-λ model, indicating model with constant λ is poor at replicating the observed data.

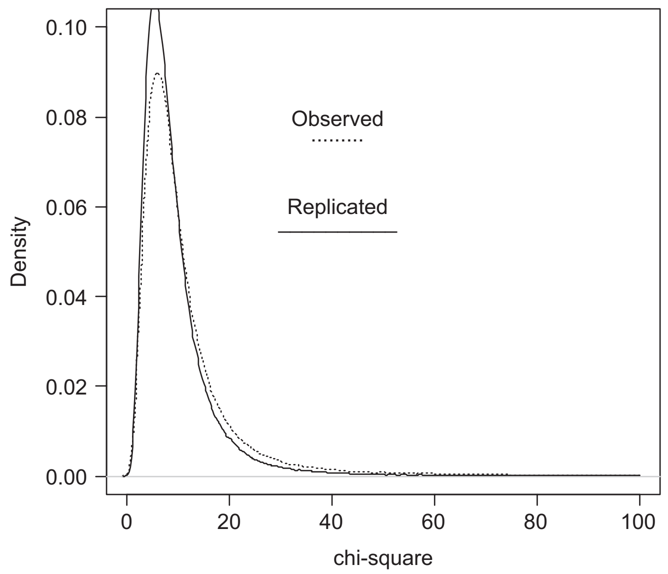

Fig. 10. Marginal posterior densities for observed and replicated Bayesian χ^2 statistics show significant overlap in the variable-λ model, indicating model with variable λ is good at replicating the observed data.

Table 21
Bayesian p-values for 12 sources, constant-λ model and hierarchical Bayes model for variable λ

Model	Bayesian p-value
Variable λ	0.46
Constant λ	0.002

8.2. Deviance information criterion

Frequentist statistics has long made us of the deviance, defined as deviance $= -2\log(\text{likelihood})$, as a measure of how well a model fits observed data. If the data are normally distributed with

Table 22
WinBUGS script for implementing Bayesian analog of Cramer-von Mises statistic

```
model
{
for (i in 1 : N) {
#Remove comments from model being used
   time[i] ~dexp(lambda) #Exponential distribution
   time.rep[i] ~dexp(lambda)
   #time[i] ~dweib(alpha, lambda) #Weibull distribution
   #time.rep[i] ~dweib(alpha, lambda)
   #time[i] ~dlnorm(mu, tau) #Lognormal distribution
   #time.rep[i] ~dlnorm(mu, tau)
   time.ranked[i] <- ranked(time[],i)
   time.rep.ranked[i] <- ranked(time.rep[], i)
   F.obs[i] <- 1-exp(-lambda*time.ranked[i]) #CVM for
   exponential #distribution
   F.rep[i] <- 1-exp(-lambda*time.rep.ranked[i])
   #F.obs[i] <- 1-exp(-lambda*pow(time.ranked[i], alpha)) #CVM
   for Weibull #distribution
   F.rep[i] <- 1-exp(-lambda*pow(time.rep.ranked[i], alpha))
   #F.obs[i] <- phi((log(time.ranked[i])-mu)*tau/2) #CVM for
   lognormal #distribution
   F.rep[i] <- phi((log(time.rep.ranked[i])-mu)*tau/2)
   diff.obs[i] <- pow(F.obs[i]-(2*i-1)/(2*N), 2)
   diff.rep[i] <- pow(F.rep[i]-(2*i-1)/(2*N), 2)
   }
CVM.obs <- sum(diff.obs[])
CVM.rep <- sum(diff.rep[])
p.value <- step(CVM.rep-CVM.obs) #Mean value should be near 0.5
```

Table 23
Results of Bayesian Cramer-von Mises test for offsite power recovery times

Distribution	Bayesian p-value
Exponential	0.488
Weibull	0.413
Lognormal	0.496

known variance, σ^2, then the deviance can be written as

$$D \propto \sum \frac{(x_i - \mu)^2}{\sigma^2} \quad (31)$$

Thus, the deviance generalizes the χ^2 statistic used earlier.

As discussed by Gelman [44] the expected deviance, computed by averaging Eq. (31) over the true sampling distribution, equals twice the Kullback–Leibler information (up to a fixed constant). As the sample size grows large, the model with the lowest Kullback–Leibler information, and therefore the lowest expected deviance, has the highest posterior probability.

Note that the discrepancy between the model and the observed data also depends on the unknown parameters (μ and σ^2 in Eq. (31)). Therefore, one can average the deviance over the posterior distribution of the unknown parameters, using draws from their simulated posterior distribution to implement this in a numerical framework.

The deviance averaged over the posterior distribution of the unknown parameters will be different than the deviance calculated for a single value, such as the posterior mean. The difference represents the effect of model fitting, and can be thought of as the effective number of parameters in a Bayesian model. This difference represents the decrease in deviance (i.e., expected improvement in model fit) expected from estimating the parameters in the model.

One can also estimate the expected deviance in applying the fitted model to replicated data from the posterior predictive distribution, with the average carried out over this distribution. This expected predictive deviance is estimated by the deviance information criterion (DIC)

$$\mathrm{DIC} = 2D_{\mathrm{avg}} - D_{\hat{\theta}} \tag{32}$$

DIC, which is computed by WinBUGS, is a Bayesian analog of the Akaike information criterion (AIC) used by frequentists [45]. For more details on DIC, see [46].

We first illustrate the use of DIC for the valve leakage data in Table 5. WinBUGS includes a menu option to calculate DIC, so no additional scripting is required. DIC should not be calculated until the chains have converged, however. The DIC for the model with a monotonic trend is 37.9, less than the value of 40.75 for the model with constant p. The lower DIC, coupled with the much better Bayesian p-value of 0.47, would likely lead us to adopt the logistic model for the valve leakage probability (p) over the simpler model with p constant in each year.

As a second example, we use DIC to choose among candidate regression models for the space shuttle O-ring data in Table 17. Recall from an earlier section that two candidate logistic regression models had been developed for the shuttle O-ring erosion probability (p): a model with temperature as the only explanatory variable, and a more complicated model with both temperature and leak-test pressure. The DIC is nearly the same for both models. Because the simpler model is essentially equivalent to the more complex ones, we would recommend it for any predictive analyses.

As a final illustration, we use DIC to help select among aleatory models for random durations. First, we will use DIC to select among the three models considered for offsite power recovery time in our earlier example from [29]. In that example we used Bayesian inference to obtain posterior distributions for the parameters of the putative Weibull and lognormal models for offsite power recovery time. We will use DIC to compare both of these models with the simpler exponential model, in which the recovery rate is constant in time.

Table 24 shows the results for the three models under consideration. The model with the smallest DIC is the Weibull model, although the difference from the other models is not very large. In fact, none of the three models differ significantly from the others based on DIC. From the standpoint of analytical simplicity, the exponential model would be preferable. Note that the lognormal distribution, with its heavy tail, will tend to give the most conservative results in terms of the probability of not recovering offsite power by a particular time. As noted earlier, the lognormal model was selected in [29], based on frequentist goodness-of-fit tests.

As a second example involving random durations, consider the following times to failure (in hours) for a cooling pump: 1258, 1388, 1022, 1989, 2024, 1638, 390, 4362, 2240, 1215, 2146, 655. Treating this as a renewal process, as described earlier, we are interested in inferring the renewal distribution. We will again consider three candidate aleatory models: exponential, Weibull, and lognormal. Using the scripts shown earlier, with diffuse priors on the distribution parameters, we find the results shown in Table 25. In this case, based on both Bayesian p-value and DIC, either the Weibull or lognormal model is clearly a better alternative than the simple exponential model, but there is no significant difference between the Weibull and lognormal models.

Table 24
DIC results for potential models for recovery time following grid-related loss of offsite power

Model	DIC
Exponential	145.6
Weibull	144.4
Lognormal	145.3

Table 25
Results of model checking for cooling pump failure times

Model	Bayesian p-value	DIC
Exponential	0.16	204.5
Weibull	0.49	200.7
Lognormal	0.50	200.3

9. Other models

A variety of other "models of the world" are amenable to inference via Bayes' Theorem, including Bayesian belief networks (BBN), influence diagrams, and fault trees. Graph models such as BBNs and influence diagrams are very similar to the directed acyclic graphs discussed earlier. In fact, an influence diagram is just a directed acyclic graph to which decision nodes have been added, as discussed in [47]. An example of these types of models is shown in Fig. 11, where we list six possible hypotheses relating how performance-shaping factors (PSFs) may (or may not) affect human performance. An application of Bayesian networks for estimating multilevel system reliability is given in [48].

Models such as fault trees, while not frequently used as the aleatory model for Bayesian inference, nonetheless provide a structure wherein both high-level data (e.g., system failures), intermediate-level data (e.g., subsystem failures), and low-level data (e.g., component failures) can all be used, simultaneously, to perform inference at any level, as discussed in Refs. [49–51]. For example, assume that we are interested in failure rates and probabilities for components that make up a system, but the observed data are at multiple "levels," including the subsystem and system level (rather than at the component level). We will use the system fault tree and the resulting expression of system failure probability in terms of constituent component failure probabilities to carry out the analysis in WinBUGS. Consider the simple fault tree shown in Figs. 6–9 of the NASA PRA Procedures Guide [52], reproduced in Fig. 12.

Assume we have seen three failures (in 20 demands) for Gate E (the only subsystem) and one failure (in 13 demands) for the Top Event (the system). Assume that basic events "A" and "B" represent a standby component, which must change state upon receipt of a demand. Assume that "A" represents failure to start of this standby component and "B" represents failure to run for the required time, which we will take to be 100 h. Assume basic events "C" and "D" represent normally operating components. We will assume the information about the basic event parameters is as shown in Table 26. Note that the error factor of the lognormal distribution given in Table 26 is related to the logarithmic precision, τ, which WinBUGS uses, by the following equation:

$$\tau = \frac{1}{[\ln(EF)/1.645]^2} \tag{33}$$

We are using the same distribution to represent epistemic uncertainty in the parameters for events C, and D, so, as discussed in [53], we should account for this state-of-knowledge dependence in the Bayesian inference by using a single sampled value for each of the λs in the problem. The WinBUGS script in Table 27 will be used to find posterior distributions for p and the λs in this model, using the available data at the subsystem and system level.

Running the script gives the results in Table 28, which are compared to the prior means. These results show that the observed data have significantly increased the shared failure rate

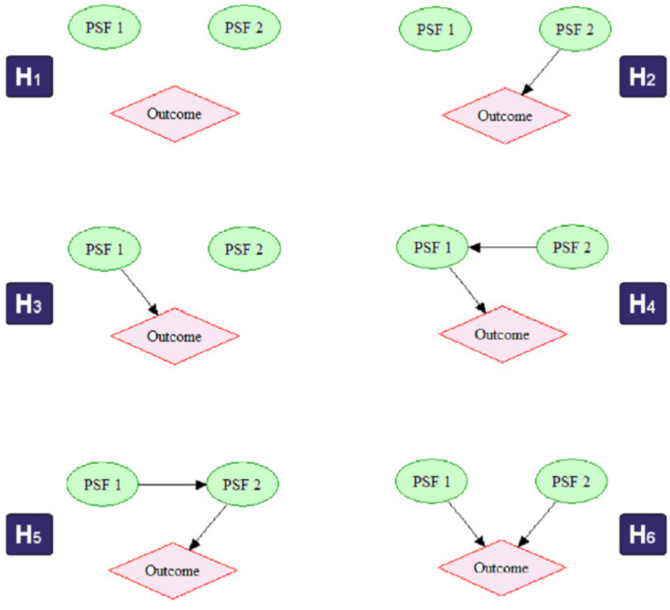

Fig. 11. Six different causal models related performance shaping factors to outcomes.

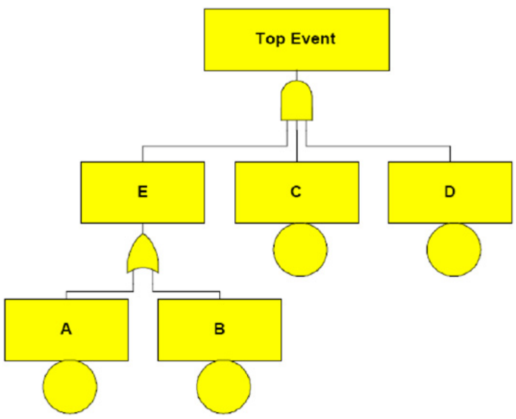

Fig. 12. Example fault tree from NASA PRA Procedures Guide [52].

Table 26
Basic event parameters for the example fault tree

Basic event	Parameters of interest	Prior distribution
A	*p*	Lognormal, mean = 0.001, EF = 5
B	Lambda	
	Mission time = 100 h	Lognormal, median = 0.005/h, EF = 5
C	Lambda	
	Mission time = 100 h	Lognormal, mean = 0.0005/h, EF = 10
D	Lambda	
	Mission time = 100 h	Lognormal, mean = 0.0005/h, EF = 10

(λ) used for basic events C and D. The estimate of the failure rate of component B has decreased from its prior mean value.

10. Conclusions

In the scientific and engineering communities, we rely on mathematical models of reality, both deterministic and aleatory. These models contain parameters—whose values are estimated from information—of which data are a subset. Uncertain parameters (in the epistemic sense) are inputs to the models used to infer the values of future observables, leading to an increase in scientific knowledge. Further, these parameters may be known to high precision and thus have little associated epistemic uncertainty (e.g., the speed of light), or they may be imprecisely known and therefore subject to large epistemic uncertainties (e.g., probability of failure of a component). The advent of MCMC-based sampling methods, coupled with easy-to-use software and powerful computers, allows us to encode information via Bayes' Theorem for a large variety of problems, domains, model types, data sets, and complications.

In this paper, data were defined as distinct observed outcomes of a physical process. It was noted that data may be factual or not and that data may be known with complete certainty or not. We described processes that are available to carry out inference in both situations.

We stated that we view data as an observable quantity. This statement implies a temporal constraint on the collection of data: data are collected prior to the present moment. While it is possible to speak about data being collected in the future, such data do not exist until that future time when they are observed.

Table 27
WinBUGS script for using higher-level information for fault tree shown in Error! Reference source not found

```
model {
x.TE ~ dbin(p.TE,          # This is system (fault tree top) observable #(x.TE
   n.TE)                   number of failures)
x.Gate.E ~ dbin(p.Gate.E, n.Gate.E)
p.TE <- (p.A               # Probability of Top Event in terms of basic #events
   +p.B)*p.C*p.D           (from fault tree)
p.Gate.E <- p.A+p.B-       # Probability of Gate E from fault tree
   p.A*p.B
p.C <- p.ftr               # Account for state-of-knowledge dependence
                           #between C and D
p.D <- p.ftr               # by setting both to the same event
p.B <- 1-exp(-lambda.B*time.miss)
p.ftr <- 1-exp(-lambda.ftr*time.miss)
# Priors on basic event parameters
p.A ~ dlnorm(mu.A, tau.A)
mu.A <- log(mean.A)-pow(log(EF.A)/1.645, 2)/2
tau.A <- pow(log(EF.A)/1.645, -2)
lambda.B ~ dlnorm(mu.B, tau.B)
mu.B <- log(median.B)
tau.B <- pow(log(EF.B)/1.645, -2)
lambda.ftr~dlnorm(mu.ftr, tau.ftr)
mu.ftr <- log(mean.ftr)-pow(log(EF.ftr)/1.645, 2)/2
tau.ftr <- pow(log(EF.ftr)/1.645, -2)
}
data
list(x.TE = 1, n.TE = 13, x.Gate.E = 3, n.Gate.E = 20,
   time.miss = 100)
list(mean.A = 0.001, EF.A = 5, median.B = 0.005, EF.B = 5,
   mean.ftr = 0.0005, EF.ftr = 10)
```

Table 28
Analysis results for example fault tree model

Parameter	Prior mean	Posterior mean	90% credible interval
p A	0.001	9.9×10^{-4}	(1.2×10^{-4}, 3.1×10^{-3})
Lambda B	8.1×10^{-3}/h	2.4×10^{-3}/h	(1.1×10^{-3}, 4.3×10^{-3})
Lambda.ftr (shared by event C and D)	5×10^{-4}/h	3.7×10^{-3}/h	(4.5×10^{-4}, 1.0×10^{-2})

If we ask experts to predict what might happen in the future—for example to predict how many component failures we can expect to see over the next 10 yr—this is not a data collection activity. Instead, we view these opinions as information, not data, since they have not yet been observed. However, it would be incorrect to assume that such information elicited from experts has less value than available data or that this information cannot be used for quantitative inference.

Information, in all its various forms, is just as useful as data when performing inference and may have more organizational value than data, depending on the type, quality, and quantity of the data.

In the risk and reliability domain, we represent processes using both aleatory and deterministic models, where we further subdivide the aleatory classification into simple parametric models (e.g., the Bernoulli process) and more complex models, such as reliability-physics models. These models require information for support, but the type of information varies from one model to the next. Further, what is known about these models and information, including data, may be less than complete, leading to a layer of epistemic uncertainty. In order to use the results produced by such models for decision-making, we need to identify, describe, and understand this uncertainty. In our view this is best done via Bayesian inference with modern computational tools, which eliminate the need for the approximations and *ad hoc* approaches of the past. The methods described in this paper illustrate these modern Bayesian inference techniques now available to analysts.

References

[1] Lunn DJ, et al. WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. Stat Comput 2000:325–37.
[2] R Development Core Team. R: a language and environment for statistical computing [Online] 2.6.1, R Foundation for Statistical Computing, 2008 ⟨www.r-project.org⟩.
[3] Gill Jeff. Bayesian methods: a social and behavioral sciences approach. Boca Raton, FL: CRC Press; 2002.
[4] Geweke John. Contemporary Bayesian econometrics and statistics. New York: Wiley; 2005.
[5] Lawson Andrew B. Statistical methods in spatial epidemiology. second ed. New York: Wiley; 2006.
[6] Castillo Enrique D, Colosimo Bianca M. Bayesian process monitoring, control, and optimization. Boca Raton, FL: CRC Press; 2006.
[7] Clark James, Gelfand Alan. Hierarchical modelling for the environmental sciences. Oxford: Oxford University Press; 2006.
[8] Siu Nathan, O, Kelly Dana L. Bayesian parameter estimation in probabilistic risk assessment. Reliab Eng Syst Saf 1998:89–116.
[9] Jaynes Edward. Probability theory—the logic of science. Cambridge: Cambridge University Press; 2003.
[10] Atwood C, et al. Handbook of parameter estimation for probabilistic risk assessment. US Nuclear Regulatory Commission; 2003 NUREG/CR-6823.
[11] Winkler Robert L. An introduction to Bayesian inference and decision. second ed. Probabilistic Publishing; 2003.
[12] Apostolakis George E. A commentary on model uncertainty. In: Ali Mosleh, et al., editors. Proceedings of workshop I in advanced topics in risk and reliability analysis, model uncertainty: its characterization and quantification. US Nuclear Regulatory Commission, 1994. NUREG/CP-0138.
[13] Apostolakis George. The concept of probability in safety assessments of technological systems. Science 1990;250:1359–64 [December 7].
[14] Kaplan Stan. On a two-stage Bayesian procedure for determining failure rates. IEEE Trans Power Apparatus Syst 1983:195–262.
[15] Pearl J. Probabilistic reasoning in intelligent systems: networks of plausible inference. Los Altos, CA: Morgan Kaufmann; 1988.
[16] Kelly DL, Atwood CL. Bayesian modeling of population variability: practical guidance and pitfalls. Ninth international conference on probabilistic safety assessment and management, Hong Kong, 2008.
[17] Gelman Andrew. Inference and monitoring convergence. In: Gilks WR, et al., editors. Markov chain Monte Carlo in practice. Chapman & Hall/CRC; 1996. p. 131–43.
[18] Eide Steven AL, et al. Industry-average performance for components and initiating events at US commercial nuclear power plants. US Nuclear Regulatory Commission; 2007 NUREG/CR-6928.
[19] Pulkkinen U, Simola K. Bayesian models and ageing indicators for analysing random changes in failure occurrence. Reliab Eng Syst Saf 2000;68:255–68.
[20] Ascher H, Feingold H. Repairable systems reliability. New York: Marcel-Dekker; 1984.
[21] Musa John D, Iannino Anthony, Okumoto Kazuhira. Software reliability: measurement, prediction, application. New York: McGraw-Hill; 1987.
[22] Pham Hoang. System software reliability. Berlin: Springer; 2006.
[23] Guida M, Calabria R, Pulcini G. Bayes inference for a non-homogeneous Poisson process with power intensity law. 5. IEEE Trans Reliab 1989;38 [December].
[24] Chen Zhao. Bayes and empirical Bayes approaches to power law process and microarray analysis. PhD. dissertation, University of South Florida, 2004.
[25] Kelly DL. Bayeisan modeling of time trends in component reliability data via Markov chain Monte Carlo simulation. In: Aven T, Vinnen JE, editors. Stavanger: risk, reliability, and societal safety, 2007.
[26] Bain L, Engelhardt M. Statistical theory of reliability and life-testing models. New York: Marcel-Dekker; 1991.
[27] US Nuclear Regulatory Commission and Electric Power Research Institute. EPRI/NRC-RES Fire PRA Methodology for Nuclear Power Facilities, 2005. NUREG/CR-6850.
[28] Atwood CL, et al. Evaluation of loss of offsite power events at nuclear power plants: 1980–1996. US Nuclear Regulatory Commission; 1997 [NUREG/CR-5496].
[29] Eide S, et al. Reevaluation of station blackout risk at nuclear power plants. US Nuclear Regulatory Commission; 2005 [NUREG/CR-6890].
[30] Martz H, Picard R. Uncertainty in poisson event counts and exposure time in rate estimation. Reliab Eng Syst Saf 1995;48:181–93.
[31] Martz H, Hamada M. Uncertainty in counts and operating times in estimating poisson occurrence rates. Reliab Eng Syst Saf 2003;80:75–9.
[32] Tan Zhibin, Xi Wei. Bayesian analysis with consideration of data uncertainty in a specific scenario. Reliab Eng Syst Saf 2003;79:17–31.
[33] Groen Frank G, Mosleh Ali. Foundations of probabilistic inference with uncertain evidence. Int J Approx Reason 2005;39:49–83.
[34] Embrey DE, et al. SLIM-MAUD: an approach to assessing human error probabilities using structured expert judgment. US Nuclear Regulatory Commission; 1984 [NUREG/CR-3518].
[35] Johnson Valen E, Moosman Ann, Cotter Paul. A hierarchical model for estimating the early reliability of complex systems. IEEE Trans Reliab 2005;54. [2 June].
[36] Dalal SR, Fowlkes EB, Hoadley B. Risk analysis of the space shuttle: pre-challenger prediction of failure. J Am Stat Assoc 1989;84:945–57 [December].
[37] Efron B. Bootstrap methods: another look at the jackknife. Ann Stat 1979;7:1–26.
[38] Kelly DL, Smith CL. Risk analysis of the space shuttle: pre-challenger bayesian prediction of failure. Los Angeles, CA: NASA Space Systems Engineering and Risk Management; 2008.
[39] Gelman Andrew, Meng XL, Stern H. Posterior predictive assessment of model fitness via realized discrepancies. Stat Sin 1996:733–806.
[40] Bayarri MJ, Berger JO. Quantifying surprise in the data and model verification. In: Bernardo JM, et al., editors. Bayesian statistics 6. Oxford: Oxford University Press; 1999. p. 53–82.
[41] Bayarri MJ, Berger JO. *P*-values for composite null models. J Am Stat Assoc 2000;95:1127–42.
[42] Johnson Valen E. A Bayesian χ^2 test for goodness-of-fit. Ann Stat 2004;32:2361–84. [6 December].
[43] Martz H, Picard R. On comparing PRA results with operating experience. Reliab Eng Syst Saf 1998;59:197–9.
[44] Gelman Andrew, et al. Bayesian data analysis. second ed. Chapman & Hall/CRC; 2004.
[45] Akaike H. Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csaki F, editors. Proceedings of the second international symposium on information theory. Budapest: Akademiai Kiado; 1973. p. 267–81.
[46] Spiegelhalter D, et al. Bayesian meaures of model complexity and fit. J R Stat Soc B 2002;64:583–640.
[47] Kjaerulff Uffe B, Madsen Anders L. Bayesian networks and influence diagrams: a guide to construction and analysis. Berlin: Springer; 2008.
[48] Wilson Alyson G, Huzurbazar Aparna V. Bayesian networks for multilevel system reliability. Reliab Eng Syst Saf 2007;92:1413–20.
[49] Martz M, Almond R. Using higher-level failure data in fault tree quantification. Reliab Eng Syst Saf 1997;56:29–42.
[50] Hamada M, et al. A fully bayesian approach for combining multilevel failure information in fault tree quantification and optimal follow-on resource allocation. Reliab Eng Syst Saf 2004;86:297–305.
[51] Graves TL, et al. Using simultaneous higher-level and partial lower-level data in reliability assessments. Reliab Eng Syst Saf, 2008:93;1273–9.
[52] Stamatelatos M, et al. Probabilistic risk assessment procedures guide for NASA managers and practitioners. National Aeronautics and Space Administration; 2002.
[53] Apostolakis G, Kaplan S. Pitfalls in risk calculations. Reliab Eng 1981;2:135–45.