IOWA STATE UNIVERSITY
Digital Repository

5-2013

Application of Bayesian Methods in Reliability Data Analyses

Ming Li
General Electric Global Research Center

William Q. Meeker
Iowa State University, wqmeeker@iastate.edu

Follow this and additional works at: http://lib.dr.iastate.edu/stat_las_preprints

Part of the Statistics and Probability Commons

Application of Bayesian Methods in Reliability Data Analyses

Abstract

The development of the theory and application of Monte Carlo Markov Chain methods, vast improvements in computational capabilities and emerging software alternatives have made it possible for more frequent use of Bayesian methods in reliability applications. Bayesian methods, however, remain controversial in Reliability (and some other applications) because of the concern about where the needed prior distributions should come from. On the other hand, there are many applications where engineers have solid prior information on certain aspects of their reliability problems based on physics of failure or previous experience with the same failure mechanism. For example, engineers often have useful but imprecise knowledge about the effective activation energy in a temperature-accelerated life test or about the Weibull shape parameter in the analysis of fatigue failure data. In such applications, the use of Bayesian methods is compelling as it offers an appropriate compromise between assuming that such quantities are known and assuming that nothing is known. In this paper we compare the use of Bayesian methods with the traditional maximum likelihood methods for a group of examples including the analysis of field data with multiple censoring, accelerated life test data, and accelerated degradation test data.

Disciplines

Statistics and Probability

Comments

This preprint was published as Li Ming and William Q. Meeker, "Application of Bayesian Methods in Reliability Data Analyses", *Journal of Quality Technology* (2014): 1-23.

Application of Bayesian Methods in Reliability Data Analyses

Ming Li
Applied Statistics Laboratory
General Electric Global Research Center
Niskayuna NY

William Q. Meeker
Department of Statistics
Iowa State University
Ames, Iowa 50011

Abstract

The development of the theory and application of Monte Carlo Markov Chain methods, vast improvements in computational capabilities and emerging software alternatives have made it possible for more frequent use of Bayesian methods in reliability applications. Bayesian methods, however, remain controversial in Reliability (and some other applications) because of the concern about where the needed prior distributions should come from. On the other hand, there are many applications where engineers have solid prior information on certain aspects of their reliability problems based on physics of failure or previous experience with the same failure mechanism. For example, engineers often have useful but imprecise knowledge about the effective activation energy in a temperature-accelerated life test or about the Weibull shape parameter in the analysis of fatigue failure data. In such applications, the use of Bayesian methods is compelling as it offers an appropriate compromise between assuming that such quantities are known and assuming that nothing is known. In this paper we compare the use of Bayesian methods with the traditional maximum likelihood methods for a group of examples including the analysis of field data with multiple censoring, accelerated life test data, and accelerated degradation test data.

Key Words: Accelerated Test, Degradation, MCMC, Likelihood, OpenBUGS, Prior Information

Dr. Li is a Statistical Leader with the Applied Statistics Laboratory, General Electric Global Research Center. His email address is limi@ge.com.

Dr. Meeker is a Distinguished Professor of Liberal Arts and Sciences and Professor of Statistics at Iowa State University. He is a Fellow of the American Society for Quality. His email address wqmeeker@iastate.edu.

1. Introduction and Motivation

1.1 Background

In the past twenty years, there has been a rapid evolution in statistical practice. Increases in computer power and developments in the theory and applications of Monte Carlo Markov Chain (MCMC) methods have made it possible to apply Bayesian statistical estimation methods to an increasingly wide range of applications. In addition, there are a number of freely-available software packages that have been developed to facilitate Bayesian computations and some commercial statistical software packages are beginning to provide easy-to-use capabilities for doing Bayesian statistical analysis.

1.2 Motivation

From a practical point of view, depending on the application, there can be three strong motivators for the use of Bayesian methods:

- Bayesian methods allow an analyst to incorporate prior information into a data analysis/modeling problem to supplement limited data, often providing important improvements in precision (or cost savings).

- Bayesian methods can handle, with relative ease, complicated data-model combinations for which no maximum likelihood (ML) software exists or for which implementing ML would be difficult. For example, available software for doing Bayesian computations can handle combinations of nonlinear relationships, random effects, and censored data that cannot be handled easily by available commercial software.

- When using Bayesian methods it is easy to produce estimates and credible intervals for complicated functions of the model parameters such as the probability of failure or quantiles of a lifetime distribution.

1.3 Related work

Although there have been many technical papers written to describe the use of Bayesian methods in reliability inference, there has been relatively little written to describe the use of the more modern computationally-based methods in this important area of application. Several exceptions are worth mentioning. Olwell and Sorell (2001) show how to supplement limited field failure data with prior information to estimate a Weibull lifetime distribution. Hamada (2005) describes a Bayesian analysis of repeated-measures degradation data from a laser life test in which the slopes of the degradation paths vary from laser to laser. León et al. (2007) describe a Bayesian analysis of complicated accelerated life data that include random batch effects. Azarkhail and Modarres (2007) also use Bayesian methods in the analysis of accelerated test data. Hamada et

al. (2008) describe Bayesian methods for a wide variety of reliability applications including accelerated testing, demonstration testing, and the estimation of system reliability. Dezfuli et al. (2009) describe the use of modern Bayesian methods to various reliability and probabilistic risk problems. This paper is an expansion and further application of some ideas that were described in Chapter 14 of Meeker and Escobar (1998).

1.4 Overview

The rest of this paper is organized as follows. Section 2 describes the relationship between likelihood and Bayesian inference methods, including discussion of some of the advantages and a brief introduction to the challenges of using Bayesian inference methods. Section 3 presents a basic example that uses simple Monte Carlo simulation to illustrate the ideas behind combining data with prior information and to compare the results obtained from using diffuse and informative prior information. Section 4 explains some of the issues that arise in the specification of prior distributions. Section 5 gives a high-level description of MCMC methods for making inferences from MCMC outputs, the popular OpenBUGS program, and some special considerations required for the effective handling of heavy censoring that arises in many reliability applications. Section 6 presents brief descriptions of several other reliability data analysis examples using other kinds of reliability data. We conclude with some remarks about the future of Bayesian methods in reliability analysis.

2. Background for the Bayesian Inference Method

2.1 The Relationship between Bayesian Inference and non-Bayesian Likelihood Inference

The left-hand side of Figure 1 shows the components of a likelihood-based non-Bayesian inference procedure. Inputs are the data and a model for the data. The inference outputs would be, for example, point estimates and confidence intervals for quantities of interest (e.g., a quantile of a failure probability associated with a failure-time distribution). The right-hand side of Figure 1 is a similar diagram for the Bayesian inference procedure. In addition to the model and the data, one must also specify a joint prior distribution that describes one's knowledge about the unknown parameters of the model. Bayes' theorem is used to combine the prior information with the likelihood to produce a posterior distribution. Similar to the non-Bayesian inference, outputs would be point estimates and credible intervals (the name commonly used to describe the Bayesian analog to non-Bayesian confidence intervals).
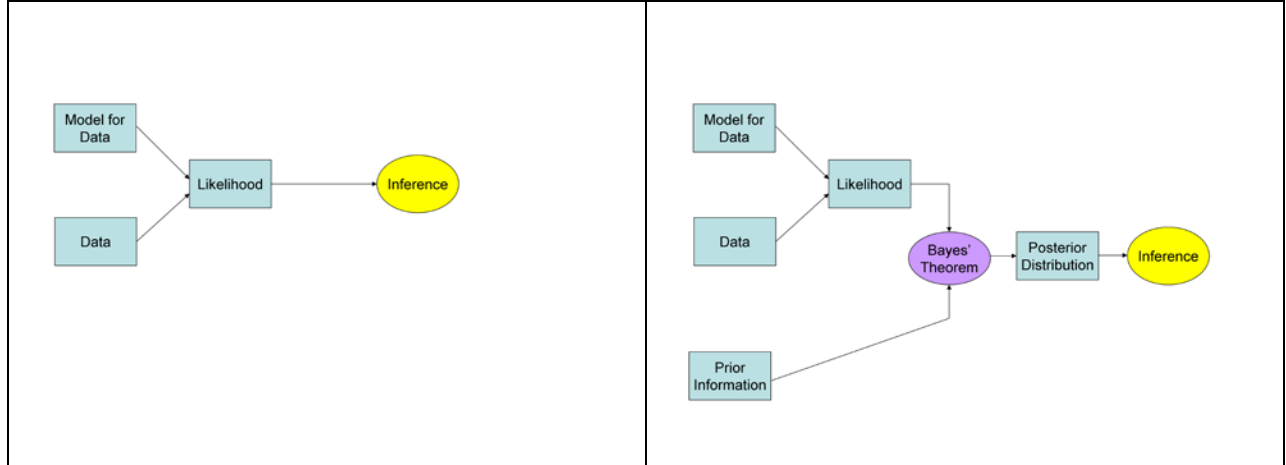
Figure 1   Comparison of the structure of the non-Bayesian likelihood and Bayesian inference methods

2.2 Bayes' Theorem and Bayesian Data Analysis

Bayes' theorem is a well-known probability rule that relates different kinds of conditional probabilities (or conditional probability density functions) to one another. This probability rule is also the basis for the Bayesian method of statistical inference which allows one to combine available data with prior information to obtain a posterior (or updated) distribution that can be used for inference. In this regard, Bayes' theorem can be written as

$$f(\boldsymbol{\theta} \,|\, \text{DATA}) = \frac{L(\text{DATA} \,|\, \boldsymbol{\theta}) f(\boldsymbol{\theta})}{\int L(\text{DATA} \,|\, \boldsymbol{\theta}) f(\boldsymbol{\theta}) d\boldsymbol{\theta}} \tag{1}$$

where $\boldsymbol{\theta}$ is a vector of unknown parameters that need to be estimated. The likelihood $L(\text{DATA} \,|\, \boldsymbol{\theta})$ is a function of the assumed model for the data and the data itself and must be proportional to the probability of the data. The likelihood quantifies the information in the data. The joint prior distribution $f(\boldsymbol{\theta})$ quantifies the available prior information about the unknown parameters in $\boldsymbol{\theta}$. The output of (1) is $f(\boldsymbol{\theta} \,|\, \text{DATA})$, the resulting joint posterior distribution for $\boldsymbol{\theta}$, reflecting knowledge of $\boldsymbol{\theta}$ after the information in the data and the prior distribution have been combined. The denominator in (1) is a normalizing constant that assures that the joint posterior distribution is a proper probability distribution (i.e., that it integrates to 1). In many practical statistical problems, computing this normalizing constant would be intractable. Modern computation methods for Bayesian analysis avoid this computational difficulty by basing inferences on a large, relatively easy to compute sample from the joint posterior distribution, making Bayesian inference computations relatively simple to conduct.

When the joint prior distribution $f(\boldsymbol{\theta})$ is diffuse and relatively flat over the range of $\boldsymbol{\theta}$ values where the likelihood $L(\text{DATA} \,|\, \boldsymbol{\theta})$ is non-negligible and the data dominates the joint prior distribution, the likelihood is approximately proportional to the joint posterior distribution.

In such situations, one can expect to have Bayesian inferences that are similar to what would make using non-Bayesian methods of analysis like ML. The important topic of prior distribution specification is discussed more fully in Section 4.

2.3 Parameterization

Parametric statistical models have unknown parameters that are to be estimated from data (sometimes with the aid of prior information). For example, the Weibull distribution cumulative distribution function (cdf) is often written as

$$\Pr(T \le t) = F(t; \eta, \beta) = 1 - \exp\left[-\left(\frac{t}{\eta}\right)^{\beta}\right], \quad t > 0$$

where β is a unitless shape parameter and η is a scale parameter that has the same units as T. This scale parameter can be interpreted as the approximate 0.632 quantile of the failure-time distribution (i.e., the time at which 63.2% of the population will fail).

For the following reasons, the parameters β and η are not the best ones to use.

- When using Bayesian methods in reliability applications, engineers are more likely to have prior information on the Weibull shape parameter and a quantile other than the 0.632 quantile (often, interest centers on a quantile in the lower tail of the distribution).

- When there is heavy censoring, the likelihood surface for η and β will tend to have an elongated shape reflecting the strong correlation between the ML estimators of η and β. This strong correlation can make the computation of ML or Bayesian estimates more difficult (increasing the amount of computer time needed or increasing the probability of algorithmic failure). As an alternative, we would like to identify and use what Ross (1972) called *stable parameters* which generally correspond to quantities that one can effectively identify in a plot of the data.

Although it is possible to use one parameterization for purposes of prior specification and a different parameterization for computing parameter estimates, we generally find that one alternative parameterization serves both purposes.

A useful reparameterization for the Weibull distribution would replace η with a particular distribution quantile that could be estimated nonparametrically directly from the available data. For heavily right censored data from a life test of a high-reliability component, this would be a lower tail quantile of the failure-time distribution. For example, the p quantile of the Weibull distribution can be written as $t_p = \eta[-\log(1-p)]^{1/\beta}$. Thus a reparameterized version of the Weibull distribution is

5

$$\Pr(T \le t) = F(t; t_p, \beta) = 1 - \exp\left[-\left(\frac{t}{t_p / [-\log(1-p)]^{1/\beta}}\right)^{\beta}\right], \quad t > 0.$$

Especially when there is heavy censoring, t_p and β will be more stable than η and β, for some appropriately chosen value of p. Moreover, graphical estimates of the chosen t_p and β that are within the range of the data (e.g., estimated by fitting a simple linear regression line through the points on a probability plot) will provide excellent starting values for either ML or Bayesian estimation methods. A useful rule of thumb for choosing p is to choose a value that is near to the center of the data. For example, if the nonparametric estimate fraction failing at the largest failure time is 0.10, then choosing $p = 0.05$ would be expected to work well.

We note that changes in parameterization are commonly used in high-quality software to produce stable estimation methods. In user-friendly software (especially commercial software) such changes are often hidden from the user. The commonly-used free software packages for doing Bayesian analysis (e.g., OpenBUGS) have not evolved to the point where such reparameterizations are automatic. Thus in the examples in this paper we will suggest and illustrate the use of appropriate reparameterizations that should be used to provide more stable estimation algorithms.

2.4 Prior information

The use of Bayesian methods for statistical modeling and inference requires one to specify a joint prior distribution $f(\boldsymbol{\theta})$ to describe the prior knowledge that is available about the unknown parameters in $\boldsymbol{\theta}$. One of the reasons that the use of Bayesian methods has been controversial in reliability (and other) applications is that it is possible that the joint prior distribution can have a strong influence on final answers, especially when the amount of data is limited (as is common in reliability applications). Section 4 provides a more detailed discussion about the practical aspects of prior distribution specification.

3. Example 1 Weibull Analysis for the Bearing Cage Failure Data

This section uses a relatively simple example of the analysis of multiply-censored field-failure data to illustrate the basic ideas and computational methods behind the use of Bayesian methods in reliability applications.

3.1 The data

To illustrate the basic ideas of Bayesian inference we will use a simple example of fitting a Weibull distribution to field failure data for an aircraft engine bearing cage. The data were first given in the Weibull Handbook (Abernethy et al. 1984) and were also analyzed in Meeker and Escobar (1998, Chapters 8 and 14).

Over time, 1,703 similar aircraft engines with this type of bearing cage had been introduced into service. The design life specification for the bearing cage was that the 0.10 quantile of bearing life (also known as B10) should be at least 8,000 hours of operation. The longest running units had seen only 2,220 hours of operation. At the time of the analysis of the data, there had been six failures and a preliminary Weibull analysis of these limited data suggested that the reliability goal had not been met. Managers needed to know if a redesign of the bearing cage would be required and also wanted to predict how many spare parts would be needed over the coming years to keep the fleet of aircraft flight-ready.
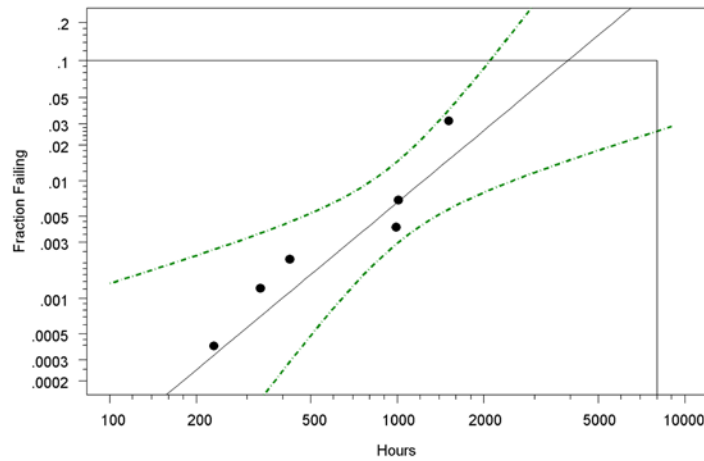
3.2 Weibull ML analysis

Figure 2 Weibull probability plot of the bearing cage failure data showing the ML estimate of fraction failing as a function of time and pointwise 95% likelihood-based confidence interval.

Figure 2 is a Weibull probability plot of the bearing cage failure data. The north-east corner of the inner rectangle corresponds to the bearing cage reliability goal of no more than 10% failing at 8,000 hours of operation. The ML estimate of the Weibull distribution cdf lies above this corner, indicating that there is a good chance that the reliability goal has not been met. On the other hand, the lower endpoints of the set of 95% pointwise two-sided confidence intervals suggest that it is possible that the true fraction failing is well below 0.10 allowing for an argument to postpone decision making until more information is available. Of course the true fraction failing at 8,000 hours could also be well above 0.10 and waiting for more information could cause a bad situation to become much worse. The problem is that there is very little information in the available data. The use of Bayesian methods with informative prior information could provide more precision and a better basis for decision making.

7

3.3 Prior information for the Bearing Cage example

For the sake of comparison, we will first do a Bayesian analysis with a diffuse joint prior distribution. An analysis with a diffuse joint prior distribution can be expected to provide estimates and credible intervals that are close to the non-Bayesian ML estimates and confidence intervals (note that "credible intervals from a Bayesian analysis are similar to confidence intervals in a non-Bayesian analysis) . This will be followed by an analysis that uses a joint prior distribution with an informative prior distribution containing information about the Weibull shape parameter β. In both cases, as described in Section 2.3, we will specify the prior distribution in terms of the more meaningful parameters of the 0.10 quantile of the life distribution (i.e., $t_{0.10}$) and the Weibull shape parameter β. We do this because it is easier to elicit prior information on these two parameters, the information on these two parameters is more likely to be approximately independent (allowing the joint prior distribution to be specified more simply by two marginal distributions), and because the 0.10 quantile is the primary quantity of interest in the study of the bearing cage life. Table 1 contains a summary of the prior distributions that we will use in this example.

Table 1 Prior distribution specification for the Bearing Cage data analyses

Analysis	Weibull Distribution Stable Parameters	
	$t_{0.10}$	β
Diffuse Prior	Log-uniform(1,000, 50,000)	Log-uniform(0.30, 8.0)
Informative Prior	Log-uniform(1,000, 50,000)	Lognormal(1.5, 3.0)

There is little or no prior information available for $t_{0.10}$. Thus to specify our prior lack of knowledge about $t_{0.10}$, we use a log-uniform distribution over the wide range between 1,000 and 50,000 hours. Choosing an even wider range for this prior distribution would have little practical effect on final answers. This prior distribution is equivalent to a uniform distribution for $\log(t_{0.10})$ between $\log(1,000)$ and $\log(50,000)$. We use this prior specification for both the diffuse-prior analysis and the informative-prior analysis in the bearing cage example.

For the diffuse-prior analysis we use a log-uniform distribution between 0.30 and 8 to describe our lack of knowledge about the Weibull shape parameter β (for field failure data a value of β outside of this range would not be expected). Again, choosing an even wider range for this prior distribution would have little practical effect on final answers. For the informative-marginal prior distribution for β we use a lognormal distribution with 99% of its probability

between 1.5 and 3.0 [denoted by Lognormal(1.5, 3.0)]. The justification for this informative prior distribution would come from previous field-data experience with fatigue failures in similar bearing cages.

3.4 Generating a sample from the joint posterior distribution via simple simulation

The basic output of modern Bayesian analysis computational tools is generally a sample from the joint posterior distribution of the model parameters. For a given likelihood and prior distribution shown in (1), there are a number of algorithms that can be used to generate such a sample. In this section we use a particularly simple method that can be employed in situations where the likelihood is easy to compute and there are only a few parameters. The more versatile MCMC methods described in Section 5 can be used with models that are much more complicated.

The points in the left-hand plot in Figure 3 are a sample from the diffuse-analysis joint prior distribution (which is actually the product of marginal distributions for both parameters). The contours in Figure 3 are relative likelihood contours, obtained by dividing the likelihood values by the value of the maximum of the likelihood. Because the likelihood is proportional to the probability of the data, we can, for example, say that the probability of the data at the ML estimate (where the relative likelihood is 1) is 100 times (10 times) larger than the probability of the data at any point on the 0.01 contour (the 0.10 contour). Also, as described in Section 8.3 of Meeker and Escobar (1998), the region enclosed by the 0.01 relative likelihood contour is an approximate 99% joint confidence region for B10 life and the Weibull shape parameter β. Similar statements can be made for any of the other relative likelihood contours. Although the upper endpoint of the prior distribution for the Weibull shape parameter exceeds 5, the plots in Figure 3 show only points below 5 to provide a better view of the interesting features of the interaction between the joint prior distribution and the relative likelihood contours.
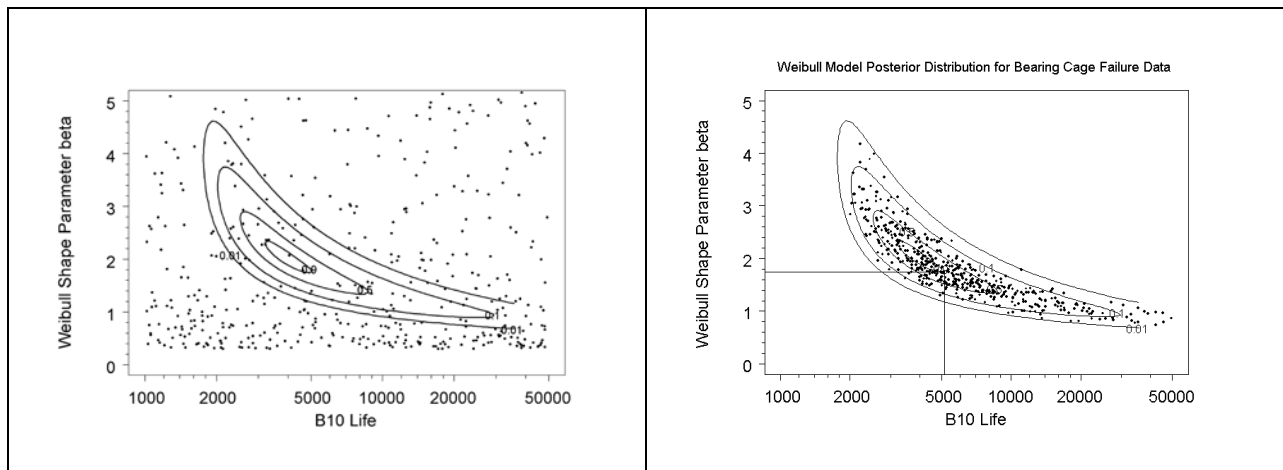

Figure 3 A sample from the diffuse joint prior distribution with likelihood contours (left) and the corresponding sample from the joint posterior distribution (right).

The right-hand side of Figure 3 shows a sample from the diffuse-prior analysis joint posterior distribution, along with the relative likelihood contours. As expected, the sample from the joint posterior agrees well with the likelihood contours because the posterior is proportional to the likelihood when the prior distribution is constant (recall that the actual prior used here is constant on the log scale).

Figure 4 is similar to Figure 3 but is based on the prior distribution that is informative for the Weibull shape parameter β. The points plotted in the left-hand side of Figure 4 are a sample from the joint prior distribution, restricting the Weibull shape parameter β to the range 1.5 to 3. The plot on the right-hand side of Figure 4 shows the corresponding sample from the joint posterior distribution. Visually we can see that the joint posterior is concentrated where the joint prior distribution and the likelihood overlap.
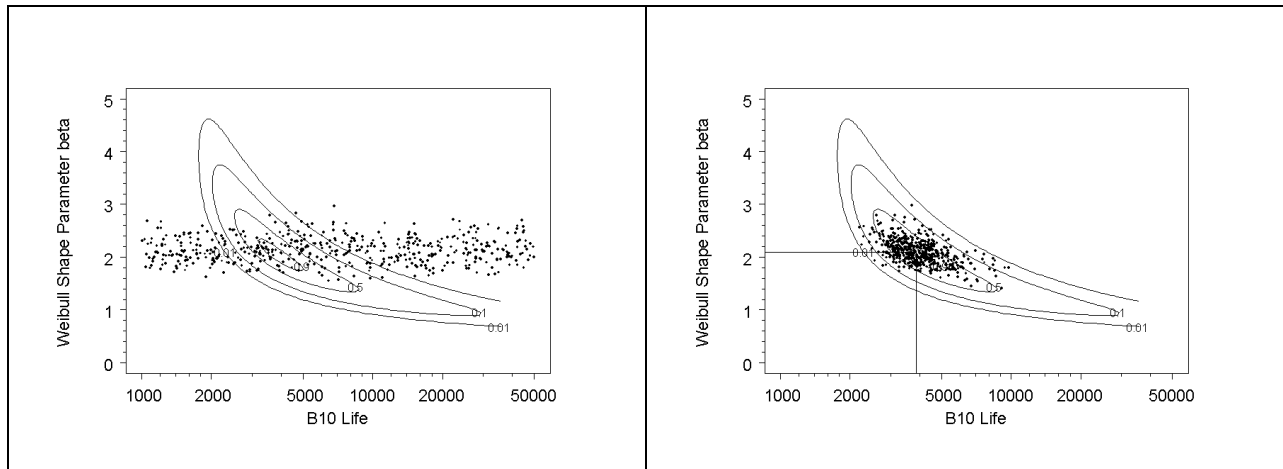

Figure 4 A sample from the informative joint prior distribution (on the Weibull shape parameter β) with likelihood contours showing the ML estimate (left) and the corresponding sample from the joint posterior distribution showing the Bayesian estimates (right).

The sample from the posterior is easy to generate in applications such as this one where the likelihood is easy to compute and when there is only a small number of parameters to be estimated. The method we used is a slight modification of the method described in Smith and Gelfand (1992). In particular, the sample from the joint posterior is obtained by randomly filtering the sample from the joint prior distribution. A point from the joint prior distribution is accepted with a probability corresponding to the value of the relative likelihood at the point. Thus points on the 0.01 contour would be kept with probability 0.01 and points on the 0.90 contour would be kept with probability 0.90. Samples of size 10,000 to 20,000 from the posterior distribution are sufficient for most practical purposes when the observations are i.i.d. (as they are in this simple simulation but not in the MCMC type simulations described in Section 5 and used in the other examples in Section 6).

3.5 Using the sample from the joint posterior distribution to make Bayesian inferences

The sample from the joint posterior is usually organized as a matrix with columns corresponding to marginal posterior distributions for the unknown parameters or other quantities of interest (e.g., functions of the parameters) that are to be estimated and rows corresponding to a large number of samples from the joint posterior. If one is interested in estimating a particular function of the parameters, new columns corresponding to the marginal posterior can be added to the matrix by simply computing the function of interest row-wise.

Point estimates for quantities of interest can be obtained by using a measure of central tendency for a given column of the matrix (i.e., the marginal posterior distribution of the quantity of interest). Theoretically, the mean of the marginal posterior distribution will provide (under an assumption of a correct model) a Bayesian estimate that minimizes squared-error loss, relative to the true quantity being estimated. The median of the marginal posterior distribution is, however, less affected by the long tail of a skewed posterior distribution and will generally agree better with the ML estimate when using a diffuse joint prior distribution. We will use the sample median of the marginal posterior distributions in all of our examples. Bayesian credible intervals can be obtained by using the appropriate quantiles of the sample from the same marginal posterior distribution. For example, a 95% credible interval for a quantity is obtained by using the 0.025 and the 0.975 quantiles of the sample from the marginal posterior distribution.

Figure 5 shows two Weibull probability plots with Bayesian estimates of the fraction failing as a function of time. On the left are the estimates using the diffuse joint prior distribution and on the right are the estimates using the joint prior distribution that is informative for the Weibull shape parameter. The width of the credible bounds is much narrower for the informative-prior analysis. The reason for this can be seen by looking at the informative analysis joint prior and joint posterior distributions in Figure 4. In particular, focusing on the lower endpoints of the credible intervals, by using the prior information that the Weibull shape parameter is larger than 1.5, the optimism that $t_{0.10}$ could be larger than 8,000 hours (or equivalently that the fraction failing by 8,000 hours could be less than 0.10) disappears. There is a similar, but less dramatic change in the upper credible bound for $t_{0.10}$.
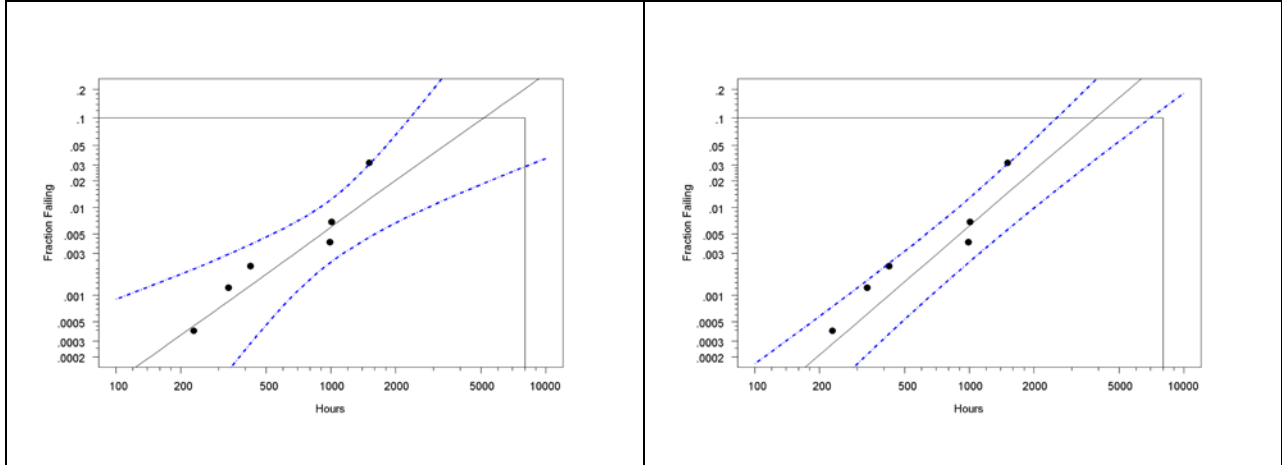
Figure 5 Weibull probability plot of the bearing cage failure data showing the Bayesian estimates of fraction failing as a function of time and pointwise 95% credible interval for the diffuse prior analysis (left) and the informative prior analysis (right).

Figure 6 provides another view of this same conclusion. These plots show density estimates for the marginal posterior distributions for $t_{0.10}$ with the diffuse-prior analysis on the left and the informative-prior analysis on the right. The vertical lines on these plots show the lower and upper 95% credible bounds for $t_{0.10}$. Again, our conclusion for the informative-prior analysis is that $t_{0.10}$ is less than 8,000 hours.
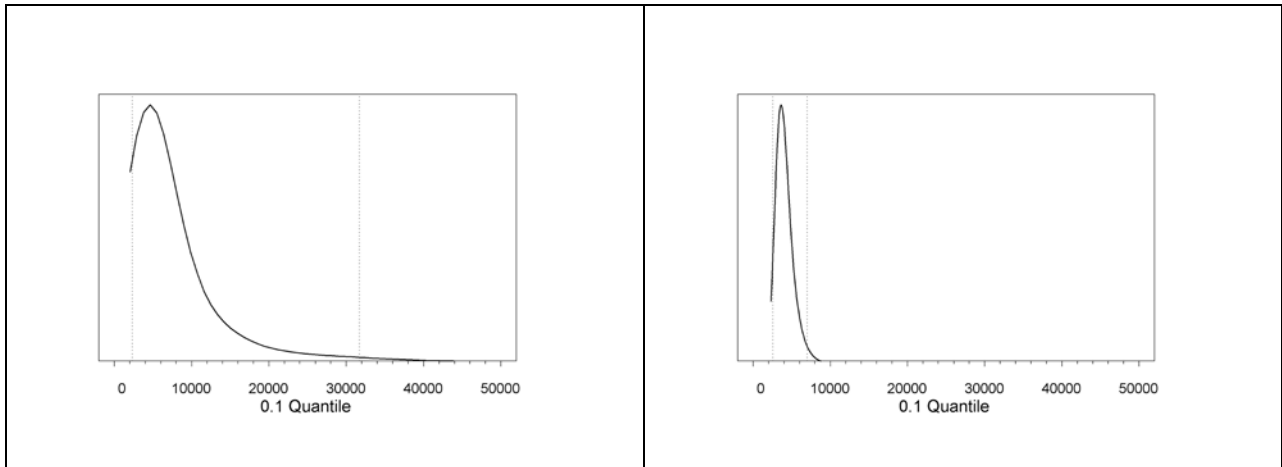

Figure 6 Plots of the marginal posterior distribution for $t_{0.10}$ of bearing cage life corresponding to the diffuse prior (left) and the informative prior (right) distributions.

12

4. Prior information

4.1 Diffuse versus informative prior distributions

A statistical model generally can be described by a relatively small number of parameters. Recall that reparameterization is often used to describe a model with quantities that are of particular interest to the analyst and we will also refer to these quantities of interest as parameters.

As described earlier, one of the important motivations for using Bayesian methods is that the analysis provides a formal mechanism for including prior information into the analysis. Thus if we have information for one or more of the model parameters and if the definition of the parameters (i.e., the particular parameterization) has been chosen such that the information about the parameters is approximately independent, we can specify a joint prior distribution with separate marginal distributions for each parameter. Informative marginal prior distributions can be used for those parameters for which there is prior information and diffuse marginal prior distributions can be specified for the other parameters.

A commonly-used practice is to specify diffuse (or vague) marginal prior distributions for each of the parameters when there is little or no prior information about the parameters. A prior distribution that is flat (or uniform) over the entire range of a parameter space is sometimes referred to as a *noninformative* prior distribution. A difficulty with this term is that a flat (noninformative) prior distribution for a parameter will imply a non-flat (and thus informative) prior distribution for any nonlinear transformation of that parameter (e.g., a prior that is noninformative (flat) for the standard deviation σ is informative (not flat) for the variance σ^2. In applications for which there is little or no useful prior information on any of the parameters, one can specify a diffuse joint prior distribution (i.e., a distribution that is flat or approximately flat over the range of the parameters where the likelihood is non-negligible). Commonly-used proper diffuse prior distribution include uniform distributions with a wide (but finite) range or a normal distribution with a large variance. It is important to note that with limited data, the choice of a prior distribution (even a diffuse prior distribution) can have strong influence on inferences. It is important when attempting to use a diffuse prior distribution that an analyst experiments with different specifications of the prior to assess sensitivity to the specification, especially when the information in the data is limited.

4.2 Who's prior should we use?

As mentioned earlier, the use of Bayesian methods has been controversial in reliability and other areas of application due to the need to specify a prior distribution. Analysts are faced with the question of which or who's prior distribution should be used in the analysis. One generally accepted principle for answering this question is that whoever is assuming the risks associated with decisions resulting from the Bayesian analysis should be allowed to choose the prior distribution. If, however, different people, groups of people, or organizations have difference risk

functions, there will be a conflict. For example, for an experiment where the results will be used to determine if a product is safe or not, customers who will use the product and managers who will benefit from the development and sale of the product will have different risk functions. In such cases it may be necessary to use a diffuse joint prior distribution or to use a non-Bayesian method that does not require specification of a prior distribution.

4.3 Sources of prior information

In some applications solid prior information, based on a combination of physics of failure and previous empirical experience, is available. This is particularly true in some engineering applications such as reliability, for which there are known, well-understood failure mechanisms. For example, engineers working in certain areas of industry will have previous experience with particular failure mechanisms and test and product operating environments that will allow them, in some situations, to provide strong prior information about aspects of the failure-time model.

For a given failure mode (e.g., fatigue), engineers will, for example, typically have useful knowledge about the Weibull distribution shape parameter. For example, if the primary failure mode for a component is caused by wear out, we immediately know that the Weibull shape parameter is greater than 1. Previous field data with similar products (as with the bearing cage) may however allow tighter bounding of the parameter. Similar knowledge is often available for the shape parameter of a Weibull or a lognormal distribution when it is used to model the failure-time distribution of microelectronic devices that fail due to known certain causes. For example, if a component will fail only when it receives an external shock that arrives according to a homogeneous Poisson process, the failure-time distribution would be exponential, corresponding to a Weibull shape parameter equal to one. Because such an assumption would be only an approximation, describing this knowledge with a prior distribution would be more appropriate.

In accelerated testing applications there is often available knowledge about the parameter describing the relationship between life and an accelerating variable. In the case of temperature acceleration of a particular chemical reaction or other mechanism relating to failure, there is often strong knowledge about the effective activation energy in the Arrhenius relationship that is commonly used to describe how temperature affects the rate of the chemical reaction (or other mechanism). Indeed, some reliability handbooks on electronic reliability (e.g., Klinger, Nakada, and Menendez 1990, page 59) provide values of the effective activation energy as a function of the failure mechanism (e.g., metalization, electromigration, or corrosion).

Yet another source of prior information is so-called "expert opinion." Individual or groups of individuals with knowledge about reliability in particular situations may be able to provide subjective information that can be used to construct appropriate prior distributions. Technical descriptions of methods of elicitation of prior distributions have been provided, for example, in Meyer and Booker (1991), O'Hagan (1998), O'Hagan and Oakley (2004), and Garthwaite, Kadane, and O'Hagan (2005).

14

There is, however, the ever-present danger that such subjective prior information is contaminated with biases arising from the risks and rewards associated with decisions that are made on the basis of a Bayesian analysis. Pressure from top management within a company or political pressure in situations where there is government oversight or funding involved could, for example, affect opinions on what prior distribution to use. *We must beware of such pressures or wishful thinking masquerading as prior information.*

4.4 Specification of prior distributions

As mentioned in Section 2.3, if appropriate definitions of stable parameters are used, it is possible to specify an appropriate joint prior distribution by specifying individual marginal distributions for each of the unknown model parameters. At the higher level, the parameters of a prior distribution for the model parameters are known as hyperparameters.

Historically within the area of Bayesian statistics there was a great deal of emphasis on the use of conjugate prior distributions. Such distributions have a posterior distribution that is in the same family as the prior distribution. Conjugate prior distributions generally afford advantages in the development of theory, ease of computation, and may even allow the use of closed form expressions for the posterior distribution. In spite of such advantages, the use of conjugate prior distributions is, in general, overly constraining. Due to recent advances in MCMC theory, related computational methods, and faster computer hardware, the use of conjugate prior distributions is no longer such an important consideration in practical applications.

Different text books and software packages use different parameterizations for the same distribution. For example most text books characterize the normal (Gaussian) distribution in terms of it mean μ and variance σ^2. The R package (R Development Core Team 2011) uses the mean and standard deviation σ. OpenBUGS uses the mean and *precision* which is defined to be $1/\sigma^2$. When programming Bayesian estimation methods in OpenBUGS or in R one must take such differences into consideration.

There are a number of ways that one could specify a marginal prior distribution to describe uncertainty in a particular model parameter. One simple method is to specify the form of the prior distribution and the corresponding hyperparameters of the prior distribution using the distribution's usual parameterization. Such an approach may not, however, be user-friendly because some of the hyperparameters may not have an easy-to-understand interpretation. Also, parameterizations for a distribution are not unique. Instead of requiring the user to specify the actual hyperparameters for a particular marginal prior distribution, one could require specification of the mean and standard deviation of the prior distribution, Such a specification may not, however, be meaningful when a distribution is highly skewed. When a prior distribution has a finite range, an alternative would be to specify the range of the distribution

along with a shape parameter or parameters. The beta and uniform distributions have a finite range and are popular for prior specification (the uniform distribution is a special case of the beta distribution). When the distribution does not have a finite range, another user-friendly alternative is to ask the user to specify a range of the distribution that contains some large proportion of the distribution's probability content. For example 99% of a marginal prior distribution is between the 0.005 and 0.995 quantiles of the distribution.

There are some special challenges in the specification of diffuse prior distributions. Relieved from the need to specify conjugate prior distributions, a uniform distribution over a wide range of potential parameter values (wide enough that the likelihood is near to 0 at the extremes of the joint prior distribution) often works well. Another popular diffuse prior distribution alternative is to use a normal distribution with a large variance (poor precision).

The conjugate distribution for the precision (i.e., the reciprocal of the variance) of a normal distribution is the gamma distribution. To specify a diffuse prior distribution for a variance parameter, it has been suggested to use a gamma distribution with a large variance. As described by Gelman (2006), many of the examples in OpenBUGS use the specification dgamma(.001,.001) as a prior distribution for the variance parameters. He points out, however, that this type of prior can be far from noninformative and can present difficulties in estimation. For example this specification implies that the precision parameter has a gamma distribution with a mean of 1.0 and a standard deviation of 31.62. The particular gamma distribution implied by this specification is highly skewed to the right, having most of its mass extremely close to 0. The distribution median is 5.24×10^{-299} and the 0.99 quantile is 0.024, implying that large values of precision (small values of variance) are unlikely. Gelman (2006) describes alternative ways to specify diffuse prior distributions for variances and variance components in hierarchical models. In our examples we use a finite-range uniform prior distribution (over some sensible range of values) to specify a diffuse prior distribution for parameters that must be positive. Then we do sensitivity analyses to assess the effect that the choice of the range has on posterior inferences.

One might specify a diffuse prior distribution by using a uniform distribution over a very wide range of values. For example, for a particular parameter one could specify the marginal prior distribution for that parameter by using $\text{uniform}\left(10^{-5},\ 10^{4}\right)$. If, however, the values of the profile likelihood for that parameter are essentially 0 outside the range from 10^{-3} to 10^{3}, then using $\text{uniform}\left(10^{-3},\ 10^{3}\right)$ instead would have little or no effect on the resulting joint posterior distribution and but would generally result in faster, more efficient computation of the joint posterior distribution.

5. Monte Carlo Markov Chain Simulation

5.1 Basic ideas of MCMC simulation

The purpose of this section is to provide a general overview of the MCMC method. MCMC is a powerful, versatile method of simulating a sample from a particular joint posterior distribution (i.e., the joint posterior distribution corresponding to a given model, data, and joint prior distribution). MCMC methods are particularly important in inference problems that have a large number of parameters. During the past 20 years there have been many developments and much has been written about MCMC. For technical details and examples, we refer the reader to some of the many books on Bayesian inference and MCMC simulation that are now available. These include, for example, Gilks, Richardson, and Spiegelhalter (1996), Congdon (2003), Gelman et al. (2013), Robert and Casella (2004), and Carlin and Louis (2008).

A Markov chain is a well known stochastic process model that can be used to characterize the probability of moving from one state to another. An important property of a Markov chain is that the probability of going from one state to another depends only on the current state and not on any of the other history of the process. Numerous algorithms have been developed that will simulate samples from a discrete-time continuous-space Markov chain such that after reaching a steady-state, the sequence of samples constitutes a sample from the desired joint posterior distribution. The most well known methods are Gibbs sampling and the Metropolis-Hastings algorithm, although combinations of these two methods and other MCMC algorithms also exist.

Because the probability of being in a particular state at time i depends on the state at time $i-1$, simulated samples from a Markov Chain are not, in general, independent (as they were in the simple Monte Carlo simulation used in Section 3). Technically, this is not a problem, as estimates of marginal posterior quantities computed from autocorrelated sample are still statistically consistent. The autocorrelation does, however, imply that larger samples from the joint posterior are needed to adequately estimate the mean and, especially, the quantiles (used to compute credible bounds) of the marginal posterior distributions for the parameters of interest. In some cases (especially when it is desired to save the joint posterior for later use), analysts will "thin" the sample by retaining every k^{th} value in the sequence, where k would be larger if autocorrelation is stronger.

It is also common practice to drop some number of the initial values from the sample, so that the remaining points more accurately represent a sample from the Markov chain after steady state has been reached. The dropped values are referred to as "burn-in" values.

The plots in the NE, NW, and SW corners of Figure 7, based on the Bearing Cage problem from Section 3 with the informative prior distribution, show Metropolis-Hastings

MCMC sample paths for three different relatively short chains (1,000 samples) that had different starting values. The plot in the SE corner shows, on the same scales, a "final" sample obtained after initial "burn-in" samples have been dropped and the sample was thinned. For this plot, 20,000 samples were drawn, the first 2,000 were discarded, and then every 20th sample was retained. The resulting thinned samples are approximately independent and it is easier to store and process the smaller thinned sample. In addition, because autocorrelation is generally small or non-existent in the thinned samples, it is relatively easy to assess the amount of Monte Carlo error in estimates computed from the samples.
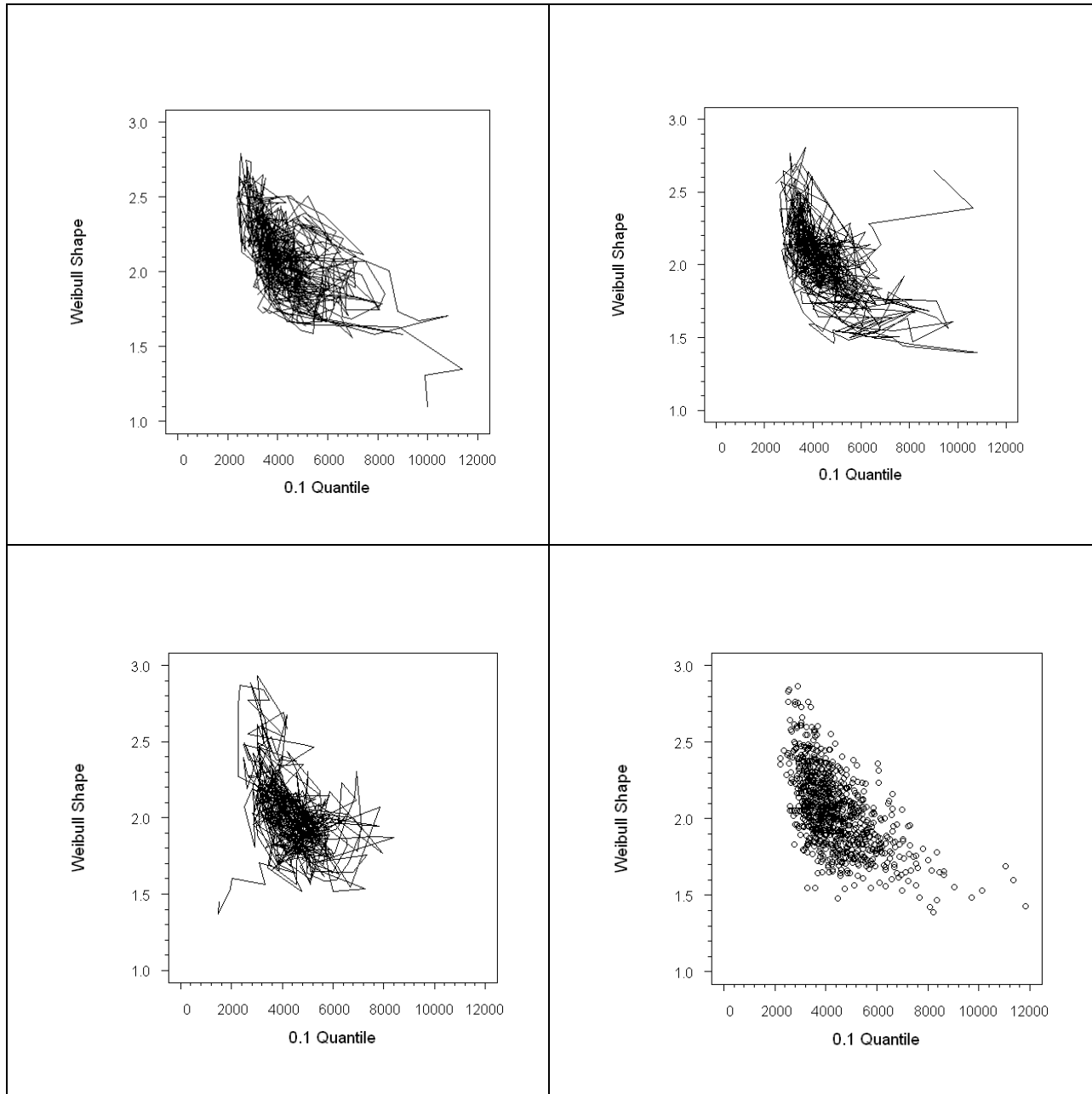
Figure 7 Simple illustrations of samples from a Markov chain for a 2-dimensional (two-parameter) distribution.

5.2 Risks of misuse of MCMC simulation

Although modern MCMC methods are extremely versatile and powerful, it is possible for a naïve user to misuse the methods and obtain seriously incorrect results. Putting programming errors aside, MCMC application problems are especially likely to occur with combinations of an attempt to use diffuse prior distributions when there are only limited data or when a poor

parameterization is used. If an improper prior distribution is used and the data are not sufficient to identify the unknown model parameters, joint posterior distribution will be improper. An MCMC algorithm will still give answers, but they will generally be wrong. If a proper joint prior distribution is specified, then the joint posterior distribution will be proper. If, however, the proper prior is diffuse and there is limited information in the data, then posterior inferences can be highly sensitive to the exact way in which the diffuse prior was specified. Sensitivity analysis is then recommended (i.e., try different methods to specify the diffuse joint prior distribution to see the effect that the choice has on posterior inferences).

If a proper joint posterior distribution exists for a given model, data, and joint prior distribution, then MCMC theory assures us that eventually a properly chosen MCMC algorithm will converge to a sample from the joint posterior distribution. In practice, however, there is no guarantee that a given MCMC simulation, run for a finite number of iterations, has converged. To gain some degree of assurance, it is necessary to use appropriate diagnostics to assess whether the sample from the chain has reached a steady state and that a sufficient number of samples have been obtained to properly estimate quantities of interest. This assessment will be more difficult to make when the sequence of samples have a high degree of autocorrelation. Useful graphical diagnostics include trace (time series) and autocorrelation function plots of the MCMC samples from the joint poster distribution for each of the model parameters. It is common practice to run the MCMC algorithm three or four times, using different starting values, and then check (using plots and numerical diagnostics) that all of them have converged to the same distribution. There are also numerical summary diagnostic tools that complement the graphical approach. These diagnostics are described in the books referenced at the beginning of Section 5.1.

A common, but difficult-to-answer question is "How many MCMC samples do I need?" The answer to this question will depend on the strength of the autocorrelation in the samples (the autocorrelation will not be the same in all dimensions of the parameter space) and on the inherent variability in the MCMC output. The autocorrelation functions (ACF) of the simulated chains are also helpful to assess how long the simulated chain should be (after the burn-in samples have been removed). One useful guideline is that if the MCMC sample is thinned sufficiently such that there is little remaining autocorrelation, then standard methods of sample size determination for independent samples can be used as a guideline. Raftery and Lewis (1992) provide some guidance on this subject.

Figure 8 gives some examples of some simple MCMC diagnostic plots from one chain corresponding to the Bearing Cage example and the MCMC output shown in Figure 7. The top row shows a trace plot of first 1000 of the Weibull 0.10 quantile sample draws provided by the chain's output, after the initial 2000 "burn-in" samples were dropped. Figure 8 suggests that the time series contains autocorrelation. The autocorrelation can be seen more clearly in the plot of the sample autocorrelation function (ACF) on the right. The bottom part of the plot is similar, but

shows first 1000 of the Weibull 0.10 quantile sample draws provided by the chain's output after the "burn-in" samples were dropped and after the output was "thinned" by keeping every 20th value. The effect of the thinning is to provide samples that have little or no autocorrelation.
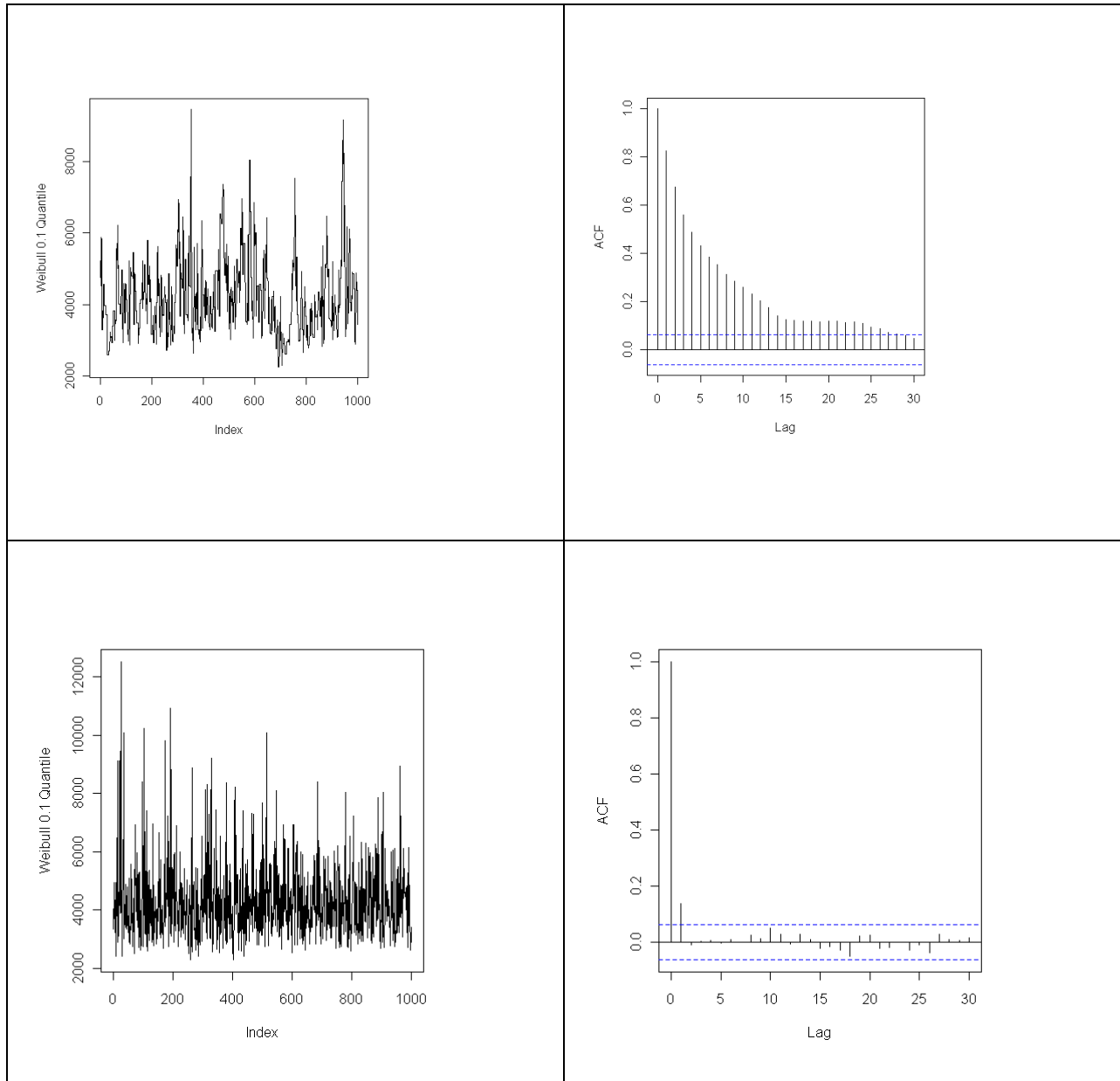

Figure 8 MCMC diagnostic plots corresponding to the chain in Figure 7 showing a trace plot (on the left) and a plot of the ACF (on the right) for both raw samples from the chain (top) and a thinned sample (bottom).

5.3 Software for MCMC simulation

Although we can expect commercial packages to eventually have easy-to-use procedures for doing a Bayesian analysis, not much is available today. In the past, many analysts would program their own analyses in a low-level programming language like C or within programming environments like R or MATLAB. Such programming can be tedious and provides a barrier that would make Bayesian analysis available only to experts. Books, such as Albert (2009), however, provide useful guidance for programming Bayesian computations.

A popular alternative to writing one's own code has been to use WinBUGS or OpenBUGS. Although OpenBUGS requires knowledge of their general Bayesian model-specification language, using OpenBUGS is far easier than lower-level programming. The documentation for OpenBUGS provides a wide range of examples with data and codes necessary to do the analyses. This collection is a valuable resource for users of the package. We used OpenBUGS for the examples in Section 6 and the codes for these examples should be useful for those wanting to use OpenBUGS for reliability data analyses. For further information about the BUGS project (past, present and future), see Lunn et al. (2009).

The proper and safe use of OpenBUGS will, however, still require a high degree of expertise that will come only with hard work and accumulated experience. For our examples, we had to pay special attention to parameterization as described in Section 2.3 and in the examples in Section 6. OpenBUGS has built-in methods for handling censored observations. Their approach assumes that there is an unknown failure time for each censored observation. Then, using a natural Bayesian approach, a parameter is added to the model for each unknown failure times. For problems involving heavy censoring, we found this approach to be computationally inefficient. As an alternative, we used the OpenBUGS "zero trick" that allows one to specify a likelihood directly. In our case we specified the usual likelihood for censored data, providing a large improvement in speed of convergence.

There exist several interfaces that can be used to run OpenBUGS from other programs such as R or Excel. For more information on these, see the OpenBUGS webpage. Another package, which is similar to OpenBUGS, is JAGS. JAGS uses the same model-specification language and capabilities. Some analysts like to use OpenBUGS and JAGS to make sure that the final answers agree. The R package coda (Plummer et al. 2006) contains useful diagnostics for studying MCMC output.

Finally, summarizing some of what we have said, the WinBUGS website provides the following "Health Warning" about the use of packages like WinBUGS, OpenBUGS, and JAGS.

> The programs are reasonably easy to use and come with a wide range of examples. There is, however, a need for caution. A knowledge of Bayesian statistics is assumed, including recognition of the potential importance of prior

distributions, and MCMC is inherently less robust than analytic statistical methods. There is no in-built protection against misuse.

6. Other Examples

In Section 3 we used an example of field-failure data. In this section we present three more examples illustrating the types of data that reliability analysts encounter in accelerated testing applications. The first example is an accelerated life test. The other two are examples of accelerated degradation tests. One of these is an example of an accelerated repeated measures degradation test and the other is an example of an accelerated destructive degradation test. All of these examples were done using OpenBUGS to simulate samples from the joint posterior distributions with post-processing of these results using specially written R functions to compute needed estimates, credible intervals and to produce the graphical outputs in this section.

6.1 Example 2 Analysis of the Device-A Accelerated Test Data

Meeker and Escobar (1998) describe the analysis of accelerated life test (ALT) data of an electronic device that they call Device A. The data were first presented and analyzed by Hooper and Amster (1990). The purpose of the ALT was to evaluate the suitability of the device for use in the first undersea digital telecommunications cable between North America and Europe. Units were tested for 5,000 hours at 10, 40, 60, and 80 Degrees C. There were 0 failures out of 30 units, 10 failures out of 100 units, 9 failures out of 20 units, and 14 failures out of 15 units tested, respectively, at these four levels of temperature. The complete data are given in Table C.10 of Meeker and Escobar (1998). The analysts needed to obtain an estimate of the failure-time distribution at the nominal use conditions of 10 Degrees C (the approximate temperature at the bottom of the Atlantic Ocean). We will focus on the estimation of $t_{0.01}(10)$, the 0.01 quantile of the failure-time distribution at 10 Degrees C.

The lognormal/Arrhenius model (as identified in both Hooper and Amster 1990 and Meeker and Escobar 1998) for the data implies that $Y = \log(\text{Hours})$ follows a normal distribution with mean $\mu(\text{Temperature}) = \beta_0 + \beta_1 x$, variance σ^2, and $x = 11605/(\text{Degrees C} + 273.15)$ is the Arrhenius transformation on temperature where 11605 is the reciprocal of Boltzmann's constant in units of eV. Then β_1 can be interpreted as the effective activation energy of the failure-causing reaction in units of eV per Kelvin.

Because the ML and Bayesian estimators of the parameters β_0 and β_1 can tend to be highly correlated, potentially causing estimation problems, we will use an alternative stable parameterization. In particular, we replace β_0 with the 0.01 quantile of the failure-time

distribution at 40 Degrees C (chosen to be near to the center of the data). More generally, the p quantile at a given temperature can be expressed as

$$\log[t_p(\text{Temperature})] = \beta_0 + \beta_1 \frac{11605}{\text{Degrees C}+273.15} + z_p\sigma$$

where for $t_{0.01}(40)$, $z_{0.01} = -2.326$ is the 0.01 quantile of the standard normal distribution. Then the reparameterized model can be expressed as

$$\mu(\text{Temperature}) = \log[t_{0.01}(40)] - z_{0.01}\sigma + \beta_1\left(\frac{11605}{\text{Degrees C}+273.15} - \frac{11605}{40+273.15}\right).$$

The effective activation energy β_1 generally depends on the particular acceleration mechanism and perhaps other factors such as material properties. Prior information for β_1 is often available from sources like previous experience testing similar components or products, general engineering knowledge, and physical-chemical knowledge about the failure mechanism. As mentioned in Section 4.3, approximate values of the effective activation energy values are available as functions of the failure mechanism.

For the Device-A example we will do an ML analysis and also do Bayesian analyses with both a diffuse joint prior distribution and a joint prior distribution that uses commonly-available information on the effective activation energy for the Arrhenius relationship. For the Bayesian analyses, we will use the prior distributions in Table 2. Here normal (0.5, 0.8) implies a normal distribution with 99% of it probability between 0.5 and 0.8. The informative prior distribution for β_1 would, in other applications, be obtained by looking at the estimates and intervals reflecting uncertainty of the effective activation energy in other experiments involving temperature acceleration of devices of the same type as Device A. For the diffuse marginal prior distributions, we use a uniform distribution between a number that is close to 0 (10^{-5}) and a very large number (10^4).

Table 2 Prior specification for the Device A accelerated life test example

Analysis	Lognormal/Arrhenius Failure-time Model Stable Parameters		
	$\log[t_{0.01}(40)]$	β_1	σ
Diffuse Prior	uniform$\left(10^{-5},\ 10^4\right)$	uniform$\left(10^{-5},\ 10^4\right)$	uniform$\left(10^{-5},\ 10^4\right)$
Informative Prior	uniform$\left(10^{-5},\ 10^4\right)$	normal (0.5, 0.8)	uniform$\left(10^{-5},\ 10^4\right)$

The plot on the left-hand side of Figure 9 is a multiple probability plot in which the plotted points are adjusted Kaplan-Meier (nonparametric) estimates of the fraction failing as a function of hours of exposure for the different levels of ambient temperature. The lines are ML estimates of fraction failing as a function of hours of exposure for the lognormal-Arrhenius acceleration model described above. The dotted lines are a set of pointwise approximate 95% confidence intervals for the 10 degrees C estimates. The ML estimate for $t_{0.01}(10)$ is 21,793 hours with a corresponding approximate 95% confidence interval of [9962.0,   47,676] hours. The plot on the right-hand side of Figure 9 is a model plot showing quantiles (0.10, 0.50, and 0.90) of the Device-A failure-time distribution as a function of temperature, reflecting the linear relationship between log life and reciprocal temperature kelvin. The horizontal line shows the censoring time.
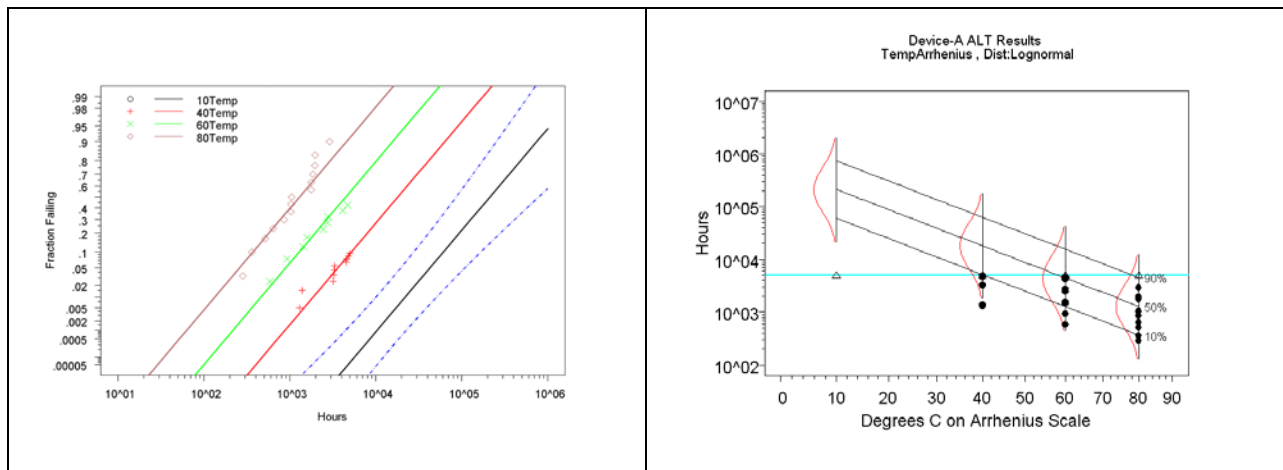

Figure 9 Lognormal multiple probability plot showing ML lognormal-Arrhenius model estimates of fraction failing as a function of time at different temperature levels (left) and a Lognormal-Arrhenius model plot of the Device-A accelerated life test data (right).

The plots in Figure 10 are similar to the plot on the left-hand side of Figure 9, except that the model estimates are from the diffuse-prior Bayesian analysis (left) and the informative-prior Bayesian analysis (right). Using the diffuse prior distribution, the 95% Bayesian credible interval for $t_{0.01}(10)$ is [9,902.9,   56,223] hours. The analysis based on the informative-prior Bayesian provides somewhat more precision, owing to the information about β_1, the effective activation energy, which was included in the informative joint prior distribution. Using the informative prior distribution, the 95% Bayesian credible interval for $t_{0.01}(10)$ is [11,933,   39,039] hours.
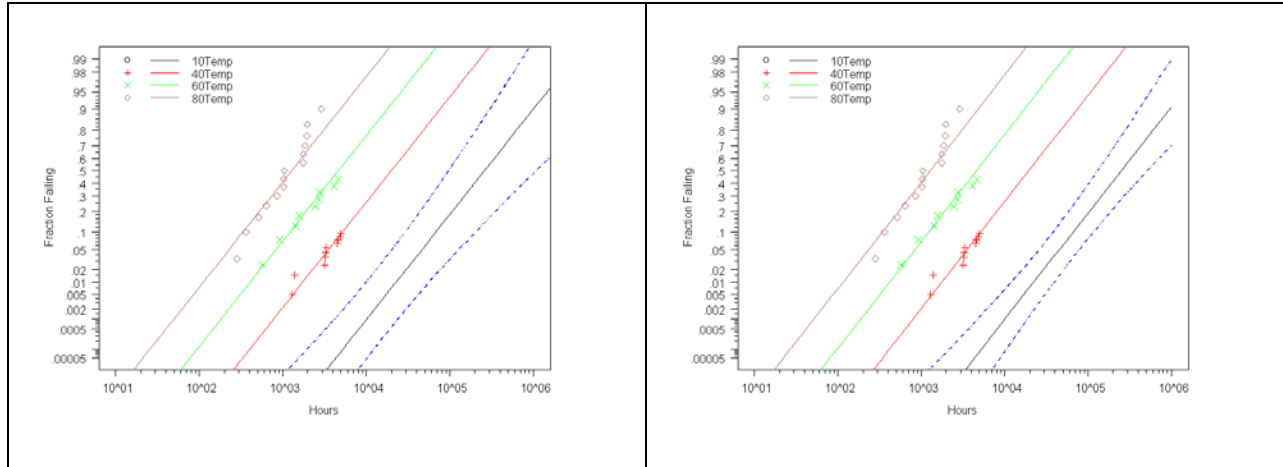
Figure 10 Lognormal multiple probability plot showing the lognormal-Arrhenius model Bayesian estimates of fraction failing as a function of time at different temperature levels and 95% credible intervals for 10 degrees C using diffuse prior distributions (left) and corresponding plot that incorporated informative prior information on the effective activation energy (right).

6.2 Example 3 Analysis of the Adhesive Bond B Accelerated Destructive Degradation Test to Evaluate the Failure-time Distribution of an Adhesive Bond

Escobar et al. (2003) describe an accelerated destructive degradation test (ADDT) that was used to evaluate the strength of an adhesive bond called Adhesive Bond B. They show how to use estimates of the degradation distribution to compute ML estimates of the adhesive bond failure-time distribution.

In the experiment, adhesive bond specimens were placed into chambers operating at 50, 60, and 70 degrees C and were removed at particular points in time, as indicated in Table 3, and subjected to a destructive strength test in which a steadily increasing force was applied until the adhesive failed. Eight baseline units were tested for strength without any aging.

Table 3 Adhesive Bond B test plan showing the number of specimens allocated to different temperature/time combinations

Temperature Degrees C	Weeks Aged					
	0	2	4	6	12	16
70		6	6	4	9	0
60		6	0	6	6	6

26

50		6	0	8	8	7
Not aged	8					

The development in this example follows that of Escobar et al. (2003), except that we use Bayesian estimation instead of ML. Figure 11 shows the measured strength data for the different combinations of Weeks and Degrees C. On the left the data are plotted on linear axes and on the right the same data are plotted on transformed axes using a log transformation for strength (newtons) and a square root transformation for Weeks (along with a fitted model described below). These transformations were identified empirically to provide an approximate linear relationship between the transformed strength and time. The horizontal line at $\mu_f = 40$ newtons is the soft-failure definition for the adhesive.
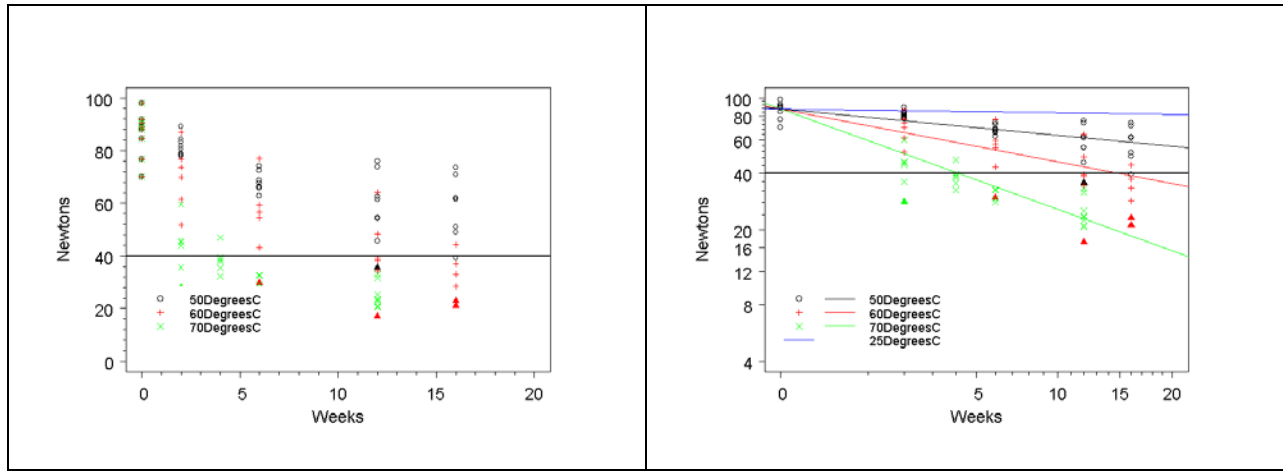

Figure 11 Scatter plot of the Adhesive Bond B accelerated destructive degradation data on linear scales (left) and transformed scales (right). The transformation scales are log for newtons and square root for Weeks. The horizontal line at 40 newtons is the soft failure definition. On the right the other lines are the ML estimates of the mean degradation level as a function of Weeks for the four different levels of temperature.

The model suggested by the plots of the data is $Y \sim \mathrm{Normal}(\mu, \sigma)$ where $Y = \log(\text{newtons})$ and

$$\mu(x, \tau; \boldsymbol{\beta}) = \beta_0 + \beta_1 \exp(-\beta_2 x) \tau \qquad (2)$$

The explanatory variables in this nonlinear regression model are $\tau = \sqrt{\text{Weeks}}$ and $x = 11605/(\text{Degrees C} + 273.15)$ is again the Arrhenius transformation on temperature. The model implies that at a given level of temperature, log strength is linearly decreasing in the square root of time. β_0 can be interpreted as the mean log strength of units tested at time 0 and the degradation rate of change at a given temperature is $\beta_1 \exp(-\beta_2 x)$ where β_2 is interpreted as

27

the effective activation energy, describing the effect that temperature has on the degradation rate of change of transformed strength versus transformed time. In the original parameterization using $(\beta_0, \beta_1, \beta_2, \sigma)$, correlations among the parameter estimates can be extremely strong (in the Adhesive Bond B data $\hat{\beta}_1$ and $\hat{\beta}_2$ have a correlation that is -0.9997. Thus, as in the other examples, it is important to use a stable parameterization. An alternative stable parameterization can be obtained that uses parameters that describe features of the data that are observed directly. One way to do this is to use the alternative form of the model with stable parameters:

$$\mu(x, \tau; \boldsymbol{\gamma}) = \gamma_0 + \gamma_1 \left\{ \exp\left[-\gamma_2 (x - \bar{x}) \right] \tau - \bar{\tau} \right\}$$

where

$$\gamma_0 = \beta_0 + \beta_1 \exp\left(-\beta_2 \bar{x} \right) \bar{\tau}$$

is the mean (transformed) strength at mean (transformed) temperature $\bar{x}$ and at mean (transformed) time $\bar{\tau}$ and

$$\gamma_1 = \beta_1 \exp\left(-\beta_2 \bar{x} \right)$$

is the slope of the linear relationship between μ and τ at the mean (transformed) temperature $\bar{x}$. The parameters $\gamma_2 = \beta_2$ and σ do not change.

Table 4 gives the diffuse and informative prior distributions that we use in the analysis of the Adhesive Bond B data. As in the Device A analysis, for the informative prior analysis, we specify an informative marginal prior distribution only for the effective activation energy, $\gamma_2 = \beta_2$ and we use diffuse marginal priors for the other stable parameters. In particular, $\text{normal}(0.65,\ 0.75)$ implies a normal distribution with 99% of its probability between 0.65 and 0.75. We do this because there would generally be useful prior information for $\gamma_2 = \beta_2$ but not necessarily for the other two parameters. For the diffuse marginal prior distributions we again use a uniform distribution between 10^{-5} and 10^4. For the regression parameters (unrestricted in sign), we use a "flat" prior, corresponding to a uniform distribution between $-\infty$ and ∞. Again, the cautious analyst would use informative prior information only when it is grounded in solid knowledge of the physics or chemistry of failure for the particular failure mode, combined with previous similar experience.

Table 4 Prior distribution specification for the Adhesive Bond B accelerated destructive degradation Arrhenius model.

Analysis	Lognormal/Arrhenius Destructive Degradation Model Stable Parameters			
	γ_0	γ_1	γ_2	σ
Diffuse Prior	flat	flat	flat	$\text{uniform}\left(10^{-5},\ 10^{4}\right)$
Informative Prior	flat	flat	$\text{normal}\left(0.65,\ 0.75\right)$	$\text{uniform}\left(10^{-5},10^{4}\right)$

The plot on the right-hand side of Figure 11shows the transformed data (using log by square root axes to show the data in the untransformed units) along with the diffuse-prior analysis Bayesian estimates of mean log strength μ as a function of time and temperature. These estimates are close to the ML estimates given in Meeker, Kugler, and Kramer (2003).

The purpose of the Adhesive Bond B accelerated test was to estimate the fraction of bonds that would fail at 2 years and at 4 years (104 and 208 weeks) of operation. As shown in Meeker, Kugler, and Kramer (2003), the degradation model in (2) induces a failure time distribution with fraction failing as a function of time given by

$$F(t;x) = \Phi\left(\frac{\tau - \nu}{\varsigma}\right), \qquad t \geq 0$$

where

$$\nu = \frac{(\beta_0 - \mu_f)\exp(-\beta_2 x)}{|\beta_1|} \quad \text{and} \quad \varsigma = \frac{\sigma \exp(-\beta_2 x)}{|\beta_1|}.$$

Figure 12 shows estimates of the fraction failing as a function of time for different temperature levels. The plot on the left is based on the diffuse-prior analysis and is almost exactly the same as the ML analysis given in Escobar, Meeker, Kugler, and Kramer (2003). The estimate for the nominal use conditions of 25 degrees C is accompanied by a set of pointwise 95% credible intervals. The plot on the right-hand side of Figure 12 is similar, but based on the informative-prior distribution analysis and we can see that this analysis provides an important improvement in estimation precision. Numerically, for the diffuse-prior analysis, the upper endpoint of the 95% credible interval for fraction failing at 25 degrees C is 0.00039 and 0.0020 respectively at 104 and 208 weeks. The corresponding values for the informative-prior analysis

are 0.00020 and 0.00054. For either analysis the conclusion would be that the probability of failure during the technological life of the product in which the adhesive was to be used would certainly be acceptably small (assuming that the acceleration model adequately describes the degradation process).
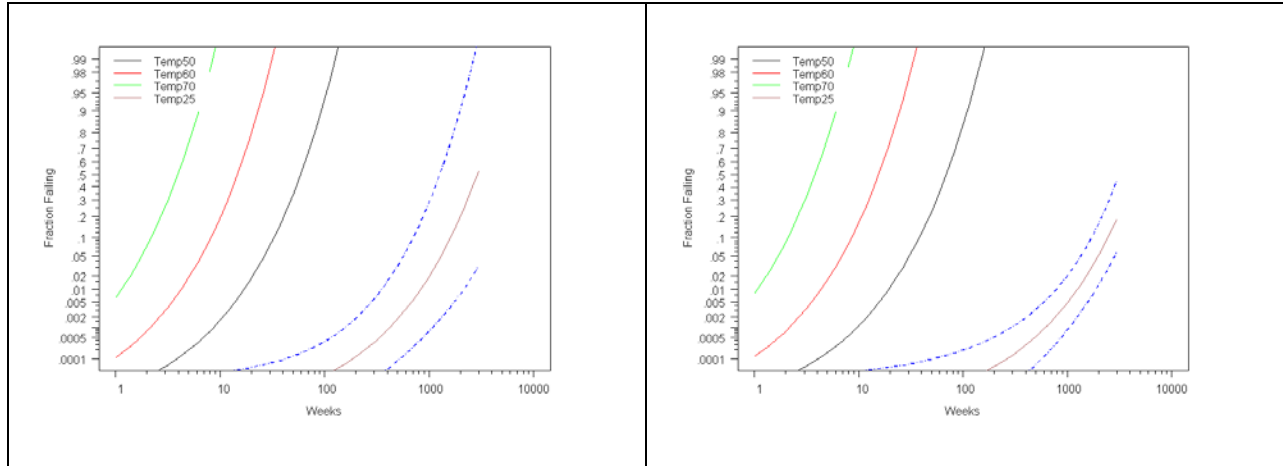

Figure 12 Lognormal probability plots showing the Bayesian estimates of the failure-time distribution of Adhesive Bond B along with 95% credible intervals at 25 decrees C, based on the accelerated destructive degradation data, an Arrhenius model and diffuse (left) and informative (right) prior distributions on the activation energy.

6.3 Example 4 LED Accelerated Repeated Measures Degradation Test

Pascual, Meeker, and Escobar (2006) describe the analysis of data from a repeated measures accelerated degradation test on LEDs. Here we reanalyze these data using a Bayesian approach. Samples of 50 LEDs were tested at six different combinations of junction temperature (temperature at the junction of the diode in the LED) in degrees Celsius and current in milliamps (mA). The light output of the LEDs was measured at the beginning of the test and at 10 other approximately equally spaced times up to 906 hours. The response used by the engineers was light output, relative to an initial measurement taken on each unit at time 0, so that all units start at 1. The plots in Figure 13 show a decrease in light output as a function of operating time. LED failure was defined as the time at which the relative light output drops to 60% of the initial value.

As can be seen on the left-hand plot in Figure 13, the sample paths in the first 138 hours had a complicated irregular behavior that LED experts were unable to explain. Because primary interest was in the long-run behavior of the LEDs, it was decided to omit the first 138 hours of data and to renormalize so that all units start with a value of 1. The renormalized data are shown in the right-hand plot in Figure 13.
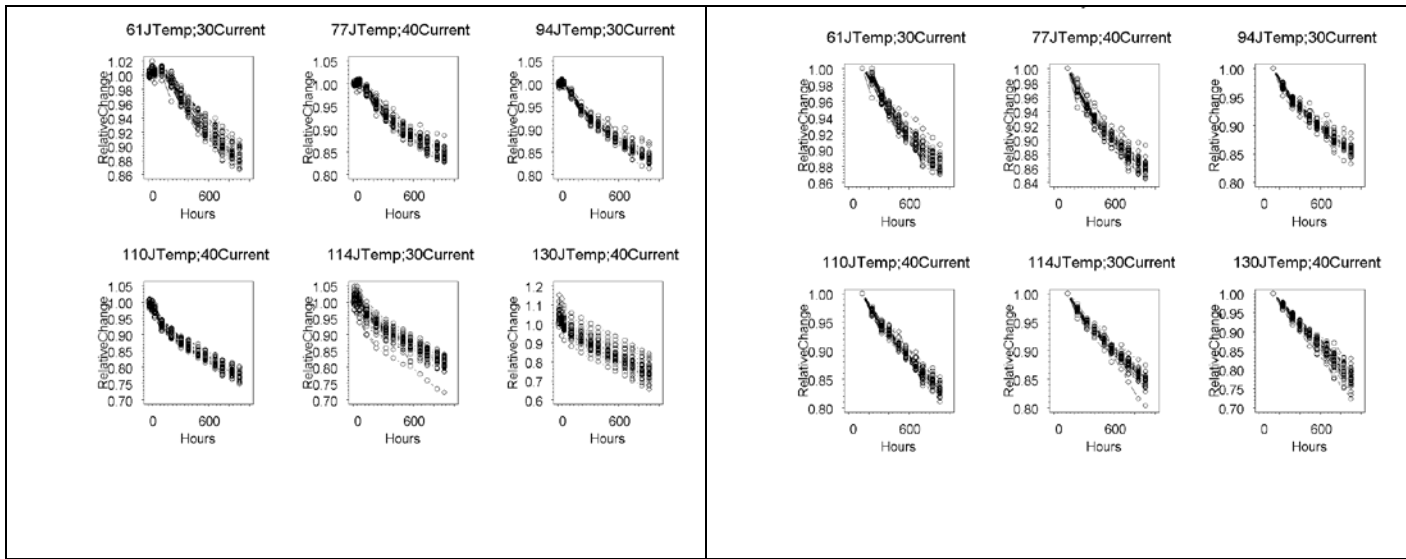
Figure 13 Plots of the original data on the left and the truncated renormalized data on the right.

Different transformations were used on both the relative output response and hours of operation in order to find a combination for which the degradation paths would be approximately linear. With no transformation on the response but a square root transformation on hours, the sample degradation paths could be well-approximated by linear relationships with different slopes for each path. Figure 14 is a plot of the data with a square root axis for time and also showing separate least-squares line fit to each sample path, extended to the 0.60 failure definition line. Although it is possible to use the individual linear regression lines to define pseudo failure times which can then be analyzed with ordinary failure-time regression model (as done in Pascual, Meeker, and Escobar 2006), here we will use a mixed effect model in which the distribution of the slopes of the degradation curves is modeled as a function of temperature and current. Both of these approaches to the estimation of lifetime from repeated measures degradation data are described in Chapter 21 of Meeker and Escobar (1998).
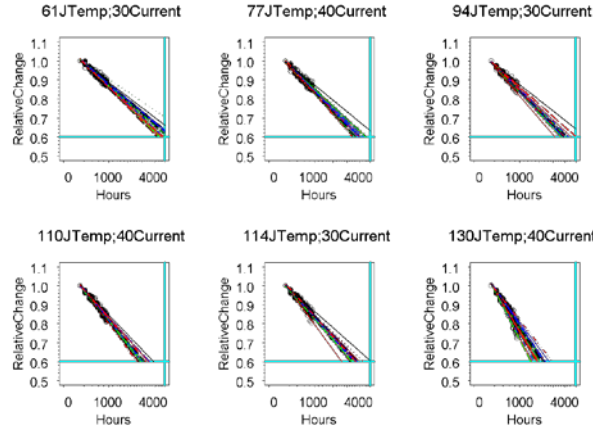
Figure 14 Plot of the truncated renormalized data with a square root time axis showing fitted linear regression lines extrapolated to reach the failure threshold.

In their analysis, Pascual, Meeker, and Escobar (2006) showed that data at the combination of 130 degrees C junction temperature and 40 mA current caused serious model fit problems, suggesting the test at this combination had caused the occurrence of a new failure mechanism. When the data at 130 degrees C junction temperature and 40 mA current were removed, the standard accelerated test model (described below and used in our analysis) fit well. All of our analyses omit the data at these conditions.

Following the standard acceleration models for temperature and current (e.g., Chapter 2 of Nelson 1990, Chapter 18 of Meeker and Escobar 1998, or Escobar and Meeker 2006) we use $x_1 = 11605 / (T_C + 273.15)$, the Arrhenius transformation on junction temperature Celsius and $x_2 = \log(I_{mA})$ is known as Black's law for current acceleration.

In the mixed effect model used here for the LED light-output degradation, motivated by the plot of the data in Figure 14, is

$$D_{ijk} = 1 + \beta_1\left(x_{1k} - x_1^0\right)\tau_j + \beta_2\left(x_{2k} - x_2^0\right)\tau_j + b_i\tau_j$$
$$= 1 + \beta_1 x_{1k}\tau_j + \beta_2 x_{2k}\tau_j + \left(b_i - \beta_1 x_1^0 - \beta_2 x_2^0\right)\tau_j$$

(3)

where x_{1k} and x_{2k} are the values x_1 and x_2 for test condition k, $\tau = \sqrt{t} - \sqrt{138}$ (recall that the first 138 hours of data had been removed before renormalization of the response, and the transformed time variable τ is defined such that $\tau = 0$ when $t = 138$), $b_i \sim N\left(\mu_b, \sigma_b^2\right), i = 1,...,180$ describes the randomness in the slopes for the 180 different LEDs, and $x_1^0 = 11605 / (94 + 273.15)$ and

32

$x_2^0 = \log(35)$. The values of x_1^0 and x_2^0 were chosen to be near the center of the respective transformed variables and the centered x values were used in the model to improve numerical stability (resulting in reduced autocorrelation in the MCMC samples). Then the model for the observed degradation is

$$Y_{ijk} = D_{ijk} + \varepsilon_{ijk} = 1 + \beta_1 x_{1k}\tau_j + \beta_2 x_{2k}\tau_j + \left(b_i - \beta_1 x_1^0 - \beta_2 x_2^0\right)\tau_j + \varepsilon_{ijk}$$

where $\varepsilon_{ijk} \sim N\left(0, \sigma_\varepsilon^2\right)$, $j = 1,...,10$ (time points), and $k = 1,...,5$ (test conditions) describes measurement error, assumed to be independent of the b_i and across time (a reasonable assumption for well-controlled laboratory experiments).

The engineers who conducted the LED accelerated test were interested in estimating the lifetime distribution of the LEDs at the nominal use conditions of 40 degrees C junction temperature and 20 mA current. Generally operating the LEDs at higher levels of current would provide better performance (i.e., more light output), but shorter life.

As mentioned above, LED failure is defined as the time when light output falls to $D_f = 0.60$ of the original light output. For model (3), the probability of failure by time t at x_1 and x_2 is

$$\Pr(T \leq t) = \Pr(D \leq D_f) = \Phi\left[\frac{D_f - \left(1 + \beta_1\left(x_1 - x_1^0\right)\tau + \beta_2\left(x_2 - x_2^0\right)\tau + \mu_b\tau\right)}{\tau\sigma_b}\right]. \tag{4}$$

Table 5 gives the diffuse and informative prior distributions used in the analysis. The diffuse prior distributions are similar to those used in the previous examples. Typically engineers would have some prior information on the accelerating variable coefficients. For purposes of illustration, the informative prior distributions on these parameters were obtained by using a distribution that has 99% of its probability between the credible limits of the posterior distribution from the diffuse prior distribution analysis. For example, the prior specification normal$\left(3.5 \times 10^{-4}, 4.1 \times 10^{-4}\right)$ implies a normal distribution with 99% of its probability between

3.5×10^{-4} and 4.1×10^{-4}. For the other parameters, the marginal prior distributions are the same as in the diffuse analysis.

Table 5 Prior distribution specifications for the LED repeated measures degradation example.

Analysis	μ_b	β_1	β_2	σ_b	σ_ε
	Lognormal/Arrhenius/Inverse Power Repeated Measures Degradation Model Stable Parameters				
Diffuse Prior	flat	flat	flat	$\text{uniform}\left(10^{-5},10^{4}\right)$	$\text{uniform}\left(10^{-5},10^{4}\right)$
Informative Prior	flat	$\text{normal}\begin{pmatrix}3.5\times10^{-4}, \\ 4.1\times10^{-4}\end{pmatrix}$	$\text{normal}\begin{pmatrix}-2.4\times10^{-3}, \\ -3.0\times10^{-3}\end{pmatrix}$	$\text{uniform}\left(10^{-5},10^{4}\right)$	$\text{uniform}\left(10^{-5},10^{4}\right)$

For both analyses, the MCMC samples from the joint posterior distribution of the model parameters were used with (4) to compute marginal posterior distributions of the failure time distribution for a large number of time values between 2,000 and 15,000 hours. For each of these time points the marginal posterior distribution was used to compute the point estimate and 95% credible bounds and these are shown in the plots in Figure 15 with the diffuse prior results on the left and the informative prior results on the right. We can see that, as expected, the informative-prior analysis improved precision of the estimates. Of course, it should be kept in mind that credible intervals (like confidence intervals in non-Bayesian analyses) reflect only sample uncertainty (due to limited data) and assume that the acceleration model and assumed distributions are correct.
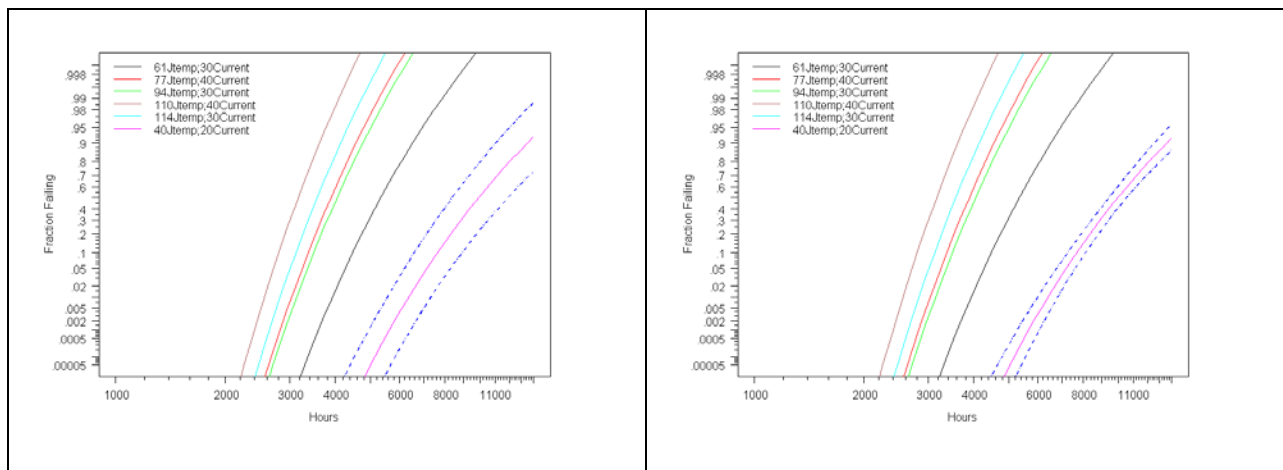

Figure 15 Lognormal probability plots showing the Bayesian estimates of the failure-time distribution of LED accelerated repeated measures degradation test, an Arrhenius model and diffuse (left) and informative (right) prior distributions on the activation energy.

7 Concluding remarks

Engineers often have useful prior information that they should use in an analysis. Bayesian methods provide a convenient formal means to incorporate engineering information or other prior information into a formal statistical analysis.

All statistical analyses involve making certain assumptions about the model (adequacy relative to the truth) and the manner in which data were obtained (e.g., from a random sample). In some applications it is possible to check these assumptions with the data or with other sources of information. The use of Bayesian inference methods requires another layer of assumptions concerning the nature of the prior information to be used in the analysis. Incorrect prior information can lead to seriously misleading inferences. Thus the use of Bayesian methods can lead to higher levels of risk of being misused. Caution is required, not only in the specification of prior information, but also to assure that there are not unrecognized convergence/estimability issues that can arise in the use of the relatively more complicated MCMC methods.

In some applications involving science and engineering, however, there will be strong, verifiable sources of prior information that can be used to construct appropriate joint prior distributions. In such applications the use of Bayesian methods can be compelling. With proper and careful use of the software tools that are available today, there will be little added risk to do such an analysis. A key advantage of a Bayesian analysis when prior information is available can be more precise estimates or smaller sample sizes in reliability tests.

Today, for most practitioners, software like OpenBUGS and JAGS, perhaps used from within the R computing environment, will offer the best alternative for doing the computations needed for a Bayesian analysis. Eventually commercial software will provide the needed functionality in an environment that will be more user-friendly (e.g., making it unnecessary for users to specify and program needed reparameterizations, making it easy for users to specify prior distributions, automatically doing needed diagnostics to check for proper convergence of the MCMC algorithms, and making it easy to do sensitivity analyses with respect to uncertain inputs). After such user-friendly software is available, we can expect to see rapid increases in the use of Bayesian methods in reliability data analysis.

Acknowledgments

References

Abernethy, R. B., Breneman, J. E., Medlin, C. H., and Reinman, G. L. (1983), *Weibull Analysis Handbook*, Air Force Wright Aeronautical Laboratories Technical Report AFWAL-TR-83-2079. Available from the National Technical Information Service, Washington, DC.

Albert, J. (2009), *Bayesian Computation with R*, New York: Springer.

Azarkhail, M. and Modarres, M. (2007), Markov Chain Simulation for Estimating Accelerated Life Model Parameters, *Reliability and Maintainability Symposium (RAMS) 2007 Proceedings*, 214-219.

Carlin, B.P. and Louis, T.A. (2008), *Bayesian Methods for Data Analysis*, Third Edition, New York: Chapman & Hall.

Congdon. P (2006), *Bayesian Statistical Modelling*, Second Edition, West Sussex, England: John Wiley & Sons.

Dezfuli, H., Kelly, D., Smith, C., Vedros, K., and Galyean, W. (2009), Bayesian Inference for NASA Probabilistic Risk and Reliability Analysis, Technical report, NASA/SP-2009-569. Downloaded from http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20090023159_2009022350.pdf on October 16, 2012.

Escobar, L. A., and Meeker, W. Q. (2006), A Review of Accelerated Test Models. *Statistical Science*, 21, 552-577.

Escobar, L.A., Meeker, W.Q., Kugler, D. L., and Kramer, L. L., (2003), Accelerated Destructive Degradation Tests: Data, Models, and Analysis, in *Mathematical and Statistical Methods in Reliability*, B. H. Lindqvist and K. A. Doksum, Editors, World Scientific Publishing Company.

Garthwaite, P. H., Kadane, J. B. and O'Hagan, A. (2005), Statistical Methods for Eliciting Probability Distributions. *Journal of the American Statistical Association*, 100, 680-701.

Gelman, A., Carlin, J.B., Stern, H.S, and Rubin, D.B. (2013), *Bayesian Data Analysis*, Third Edition, New York: Chapman & Hall.

Gelman, A. (2006), Prior Distributions for Variance Parameters in Hierarchical Models, *Bayesian Analysis* 1,515–533.

Gilks, W.R., Richardson, S. and Spiegelhalter, D. (Editors) (1996), *Markov Chain Monte Carlo in Practice*. Boca Raton: Chapman & Hall/CRC.

Hamada, M.S. (2005), Using Degradation Data to Assess Reliability, *Quality Engineering* 17, 615-620.

Hamada, M. S., Wilson, A. G., Reese, C. S., and Martz, H. F. (2008), *Bayesian Reliability*, Springer, New York.

Hooper, J. H., and Amster, S. J. (1990), Analysis and Presentation of Reliability Data, in *Handbook of Statistical Methods for Engineers and Scientists*, Harrison M. Wadsworth, Editor. New York: McGraw Hill.

Klinger, D. J., Nakada, Y., and Menendez, M. A. (1990), *AT&T Reliability Manual*, New York: Van Nostrand Reinhold.

León, R. V., Ramachandran, R. , Ashby, A.J., and Thyagarajan, J. (2007), Bayesian Modeling of Accelerated Life Tests with Random Effects, *Journal of Quality Technology* 39, 3-16.

Lunn, D., Spiegelhalter, D., Thomas, A., and Best, N. (2009), The BUGS Project: Evolution, Critique and Future Directions (with discussion), *Statistics in Medicine* 28, 3049-3082.

Meeker, W. Q. and Escobar, L. A. (1998), *Statistical Methods for Reliability Data*, John Wiley and Sons, New York.

Meyer, M., and J. Booker (1991), *Eliciting and Analyzing Expert Judgment: A Practical Guide*. London: Academic Press.

O'Hagan, A. (1998), Eliciting Expert Beliefs in Substantial Practical Applications. *The Statistician* 47, 21-35 (with discussion), 55-68.

O'Hagan, A. and Oakley, J. E. (2004), Probability is Perfect, but We Can't Elicit it Perfectly. *Reliability Engineering and System Safety* 85, 239-248.

Olwell, D. H. and Sorell, A. A. (2001), Warranty Calculations For Missiles With Only Current-status Data, Using Bayesian Methods, *Reliability and Maintainability Symposium (RAMS) 2001 Proceedings*, 133-138.

Pascual, F.G., Meeker, W.Q., and Escobar, L.A. (2006), Accelerated Life Test Models and Data Analysis. Chapter 19 in *Handbook for Engineering Statistics*, Hoang Pham (Editor), New York: Springer.

Plummer, M., Best, N., Cowles, K. and Vines, K. (2006), CODA: Convergence Diagnosis and Output Analysis for MCMC, *R News*, 6, 7-11.

R Development Core Team (2011), R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Raftery, A.E. and Lewis, S.M. (1992). How Many Iterations in the Gibbs Sampler? In *Bayesian Statistics* 4 (J.M. Bernardo et al., editors), Oxford University Press, pp. 763-773.

Robert, C. and Casella, G. (2004), *Monte Carlo Statistical Methods*, New York: Springer.

Ross, G. J. S. (1990), *Nonlinear Estimation*, New York: Springer-Verlag.

Smith, A. F. M., and Gelfand, A. E. (1992), Bayesian Statistics Without Tears: a Sampling-Resampling Perspective, *The American Statistician*, 46, 84-88.