Bayesian Computations for a Class of Reliability Growth Models

Alaattin ERKANLI
Developmental Epidemiology Program
Duke University Medical Center
Durham, NC 27710
(al@psych.mc.duke.edu)

Thomas A. MAZZUCHI
Department of Operations Research and
Department of Engineering Management
The George Washington University
Washington, DC 20052
(mazzu@gwis2.circ.gwu.edu)

Refik SOYER
Department of Management Science
The George Washington University
Washington, DC 20052
(soyer@gwis2.circ.gwu.edu)

In this article, we consider the development and analysis of both attribute- and variable-data reliability growth models from a Bayesian perspective. We begin with an overview of a Bayesian attribute-data reliability growth model and illustrate how this model can be extended to cover the variable-data growth models as well. Bayesian analysis of these models requires inference over ordered regions, and even though closed-form results for posterior quantities can be obtained in the attribute-data case, variable-data models prove difficult. In general, when the number of test stages gets large, computations become burdensome and, more importantly, the results may become inaccurate due to computational difficulties. We illustrate how the difficulties in the posterior and predictive analyses can be overcome using Markov-chain Monte Carlo methods. We illustrate the implementation of the proposed models by using examples from both attribute and variable reliability growth data.

KEY WORDS: Bayesian inference; Gibbs sampling; Markov-chain Monte Carlo methods; Ordered Dirichlet distribution.

During the development phase of a new system, testing is performed in stages, and at the end of each test stage, design changes/modifications are made to the system with the intention of improving its performance. Development of models, termed reliability growth models, for this test/modify process has been the focus of much attention in the reliability literature.

It is common to classify reliability growth models by the type of data that testing yields—that is, attribute or variable. The attribute-testing scenario has been considered in the literature from both the sample-theoretic (e.g., see Lloyd and Lipow 1962; Barlow and Scheuer 1966; Finkelstein 1983; Johnson 1991) and Bayesian (e.g., see Pollock 1968; Smith 1977; Mazzuchi and Soyer 1992, 1993) perspectives. The variable-data models are most often based on point processes, more specifically the nonhomogeneous Poisson processes [see, for example, Crow (1982) and Sen and Bhattacharyya (1993) for sampling theory and Kyparisis and Singpurwalla (1984) for Bayesian approaches]. Other proposed variable-data reliability growth models are based on assumptions of a monotonic (nondecreasing) structure on the mean failure times associated with each stage of testing and the use of isotonic regression for estimation [see, for example, McWilliams (1979), Padgett and McNichols (1982), and Robinson and Dietrich (1987) for sampling theory and Robinson and Dietrich (1989) for Bayesian approaches].

As noted by Meth (1992, p. 339), reliability growth models, “can potentially provide the basis for (1) planning reliability growth tests, (2) monitoring progress and estimating current (instantaneous) reliability, and (3) forecasting future reliability improvements.” Due to today’s complex high-cost, high-reliability systems—such as hydromechanical devices, aircraft generators, jet engines, weapon systems—these are genuine concerns.

Most of the aforementioned models are based on the sample-theoretic paradigm and often require large amounts of test data for implementation. Not only is this in conflict with today’s testing environment in which test data is “usually of limited quantity” (see Crow 1974, p. 3), but in addition such models are of no use in the planning decisions prior to initiation of the development phase.

One of the strengths of a Bayesian approach is that it allows for the input of prior perceptions about the reliability growth process into the analysis and provides a formal procedure for combination of this subjective input with available test data via the calculus of probability. In addition, a Bayesian approach provides a coherent framework for making decisions (e.g., see Lindley 1990) with regards

© 1998 American Statistical Association and the American Society for Quality
TECHNOMETRICS, FEBRUARY 1998, VOL. 40, NO. 1
to planning during the development phase. The Bayesian approaches of Pollock (1968), Smith (1977), and Robinson and Dietrich (1989), however, have either not developed an explicit procedure for the incorporation of prior information into the prior distribution and/or have not provided a fully Bayesian analysis with full posterior and predictive distributions. Both of these drawbacks are addressed with the class of models presented here.

For attribute data, the Bayesian model of Mazzuchi and Soyer (1992) was developed to address the shortcomings of previously developed attribute models. Explicit procedures were discussed for incorporation of prior information into the prior distribution, and full predictive and posterior distribution expressions were presented. Use of this model in addressing reliability growth management issues (see Department of Defense 1981) such as how to assess the duration and cost of the development phase both prior to and during testing, how to monitor the growth process and compare it with initial estimates, and when to terminate the development process and declare the system ready for field use, was discussed by van Dorp, Mazzuchi, and Soyer (1997).

Consideration of variable reliability growth models from a Bayesian perspective is technically a more difficult issue in terms of both the incorporation of prior information and the development of distributional results. Robinson and Dietrich (1989), for example, considered an exponential time-to-failure model and focused on the use of uniform or gamma prior distributions with little discussion on the method of incorporating prior information. In addition, the authors noted that development of full posterior distributions is a difficult task, so they provided only posterior moments in their analysis. Without full distributional results, predictive analysis is not available. The preceding limitations make addressing reliability growth management issues a formidable task.

In this article, we extend the results of the Mazzuchi–Soyer model to the variable-data scenario considering both the exponential and Weibull time-to-failure models. We consider a unified framework for reliability growth models that includes both the attribute- and variable-data scenarios. This framework allows for incorporation of expert judgment, the coherent combination of subjective and test information, and addressing important reliability growth management issues through predictive analysis. Computational difficulties associated with posterior analysis in the variable models are handled through the use of Markov-chain Monte Carlo (MCMC) methods.

In Section 1, we review the attribute-data reliability growth model introduced by Mazzuchi and Soyer (1992) and discuss the difficulties in computing the posterior and predictive inferences. In Section 2, we introduce extensions of this model to the variable-data case in which the failure distribution in each testing stage is assumed to be either exponential or Weibull. We again discuss the difficulties in inference for each case. In Section 3, we present inference for these models using MCMC methods based on a Gibbs-rejection sampling algorithm. In Section 4, we first illustrate the accuracy of the approach by comparing MCMC results to analytical results obtained for the attribute-data scenario by van Dorp et al. (1997). Then, we compare our approach for the variable-data case to a representative non-Bayesian model presented by Sen and Bhattacharyya (1993).

1. A BAYESIAN ATTRIBUTE RELIABILITY GROWTH MODEL

Consider a testing–modification scenario in which identical replications of the system are tested until a failure is observed. On the discovery of a failure, a modification is made to the system to remove the cause of failure and therefore increase system reliability. The test–modification scenario is repeated for some specified number of times. Let \( m \) denote the total number of test/modify stages, \( R_i (i = 1, \ldots, m) \) denote the system reliability for the \( i \)-th stage of testing (i.e., prior to the \( i \)-th system modification), and \( R_{m+1} \) denote the final (field) reliability; then the distribution for \( N_i \), the number of items tested at each stage, is given by the geometric distribution

\[
Pr\{N_i = n_i | R_i\} = (1 - R_i) R_i^{n_i - 1},
\]

\( n_i = 1, 2, \ldots \) \hspace{1cm} (1.1)

Because engineering modifications are made to the system on the discovery of a failure, it is reasonable to assume that

\[
0 < R_1 \leq \cdots \leq R_{m+1} < 1.
\]

In the Bayesian paradigm, uncertainty about the parameters \( R_1, \ldots, R_{m+1} \) is expressed by a probability distribution. Following Mazzuchi and Soyer (1992), a natural and mathematically tractable (prior) probability distribution for \( R = (R_1, \ldots, R_{m+1}) \) is the ordered Dirichlet distribution given as

\[
II (R | D^{(0)}) = \frac{\Gamma(\beta)}{\prod_{j=1}^{m+2} \Gamma(\beta \alpha_j)} \prod_{j=1}^{m+2} (R_j - R_{j-1})^{\beta \alpha_j - 1}, \hspace{1cm} (1.3)
\]

where \( R_0 = 0, R_{m+2} = 1 \), and \( D^{(0)} \) represents the information prior to any testing as captured by the prior parameters \( \beta, \alpha_j > 0, \sum_{j=1}^{m+2} \alpha_j = 1 \). Note that the distribution is defined over the simplex \( \{R | 0 < R_1 \leq \cdots \leq R_{m+1} < 1\} \) and thus embodies the restrictions in (1.2) and imposes no additional restrictions in the analysis.

An attractive feature of the ordered Dirichlet distribution is that incorporation of expert judgment is facilitated because all relevant marginal and conditional distributions are beta distributions. For example, defining \( \alpha^*_j = \sum_{\ell = 1}^{j} \alpha_{\ell} \), it can be shown that the marginal distributions are given by

\[
(R_i | D^{(0)}) \sim \text{beta}(\beta \alpha^*_j, \beta(1 - \alpha^*_j)), \hspace{1cm} (1.4)
\]

where \( E[R_i | D^{(0)}] = \alpha^*_j \) and \( \beta \) is a degree-of-belief parameter with lower values of \( \beta \) reflecting more spread in the

\[\text{TECHNOMETRICS, FEBRUARY 1998, VOL. 40, NO. 1}\]
distribution. Similarly, it can be shown that

$$[R_j - R_i \mid D^{(0)}] \sim \text{beta}(\beta(\alpha_j - \alpha_i), \beta(1 - \alpha_j + \alpha_i))$$

for \( i < j \). (1.5)

implying that \( E[R_i - R_{i-1} \mid D^{(0)}] = \alpha_i \), and thus the parameter \( \alpha_i \) can be interpreted as the expected one-step improvement in reliability from stage \( i - 1 \) to \( i \). Finally, it can be shown that

$$[R_i \mid D^{(0)}] \sim \text{beta}(\alpha_i, \beta(\alpha_i - \alpha_i))$$

for \( i < j \). (1.6)

It has been common practice to elicit measures such as the mean, mode, and/or variance to specify the prior distribution parameters. That is, the elicited quantities would be compared with the analytical expressions to specify the prior parameters. The expressions for mean, median, variance, and so forth for the beta distribution are well known and can be obtained in closed form for all of these. Other elicitation measures based on observables were proposed for the beta distribution by Chaloner and Duncan (1983). The various forms of expressing the reliability growth process, either piecewise, by additive increase, or by relative increase, facilitate expert input. In the case in which prior information is not available, the ordered uniform distribution, obtained as (1.3) with \( \beta \alpha_j = 1 \) for all \( j \), can be used.

After the specification of the prior parameters, probability statements concerning the reliability growth process can be made using (1.4), (1.5), and (1.6) prior to initiating testing. In addition, the testing effort can be assessed through probability statements about the \( N_i \)'s for any stage \( i \), via the predictive distribution \( \Pr\{N_i = n_i \mid D^{(0)}\} \), which is a Polya distribution whose expected value can be determined analytically (see van Dorp et al. 1997).

Defining \( D^{(i)} = \{D^{(0)}, n_1, \ldots, n_i\} \) and assuming that the \( N_i \)'s are conditionally independent given \( R_i \), the likelihood function of \( R \) after \( i \) stages of testing is given by

$$L(R; D^{(i)}) = \prod_{j=1}^{i} (1 - R_j) R_j^{n_j-1}. \quad (1.7)$$

The posterior distribution of \( R \) is proportional to the product of (1.3) and (1.7). After the \( i \)th testing stage, the interest will be in the current and future system reliabilities \( R_i, \ldots, R_{m+1} \). As shown by van Dorp et al. (1997), by integrating out \((R_1, \ldots, R_{i-1})\) over the region \( 0 < R_1 \leq \cdots \leq R_{i-1} < R_i \), the posterior distribution \( \Pi(R_1, \ldots, R_{m+1} \mid D^{(i)}) \) can be obtained as a mixture of ordered Dirichlet distributions. Thus, given \( D^{(i)} \), marginal distributions for the current and future stage reliabilities \( R_j, j = i, \ldots, m+1 \), and predictive distributions can be obtained by taking weighted averages of the prior expressions evaluated for revised parameter values. The mixture involves nested sums of ratios of gamma functions, however, and when \( m \) is large (say, \( m \geq 10 \)), the evaluation of the weights becomes cumbersome and the results become inaccurate due to the problems in evaluation of gamma functions. In addition, posterior distribution results for \( R_j, j = 1, \ldots, i-1 \), do not exist in closed form. As will be discussed in Section 3, the computations for all posterior and predictive forms can be done more efficiently and accurately by using MCMC methods.

2. EXTENSIONS OF THE MODEL FOR THE CASE OF VARIABLE DATA

Though other sampling models are easily accommodated, we adopt a sampling scheme similar to that for the attribute reliability growth model of Section 1. It is assumed that identical copies of the system will be tested one at a time for a fixed period of time \( \tau \) until a failure is observed before \( \tau \). On the discovery of the failure, a modification is made to remove the cause and thus improve the system’s reliability. Thus, if \( F(\tau \mid \theta_i) \) is the underlying cdf of the system in stage \( i \) conditioned on the parameters \( \theta_i \), then the distribution for \( N_i \), the number of items tested at stage \( i \), is given by

$$\Pr\{N_i = n_i \mid F(\tau \mid \theta_i)\} = (F(\tau \mid \theta_i))^{n_i-1} F(\tau \mid \theta_i),$$

$$n_i = 1, 2, \ldots, \quad (2.1)$$

where \( F(\tau \mid \theta_i) = 1 - F(\tau \mid \theta_i) \) is the system reliability for a mission time \( \tau \) at stage \( i \). Note that the case in which we test each item to failure is easily accommodated by letting \( \tau \rightarrow \infty \). As before, it is assumed that the test–modification process is repeated for \( m \) stages.

2.1 Exponential Failure Data

The following model will be referred to as the exponential reliability growth (ERG) model. Let \( X_i, i = 1, 2, \ldots, \), denote the life-length of the system during the \( i \)th stage of testing and \( \lambda_i \) denote the failure rate of the system during the \( i \)th stage. The quantity, \( \lambda_{m+1} \), denotes the system failure rate in the field. If the system does not exhibit any aging, then, given \( \lambda \), the failure behavior of the system is described by the exponential failure model with density

$$p(x_i \mid \lambda_i) = \lambda_i \exp(-\lambda_i x_i). \quad (2.2)$$

Under the reliability growth scenario, it is expected that \( \{\lambda_i\} \) is a nonincreasing sequence in \( i \); that is,

$$\infty > \lambda_1 \geq \lambda_2 \cdots \geq \lambda_{m+1} > 0. \quad (2.3)$$

By reparameterizing (2.2) as

$$p(x_i \mid R_i) = \ln(1/R_i) R_i^{-x_i}, \quad (2.4)$$

where

$$R_i = \exp(-\lambda_i), \quad (2.5)$$

(2.3) is reflected as the restriction of (1.2) for the \( R_i \)'s.

Following the development in Section 1, the prior on \( R \) is given by an ordered Dirichlet as in (1.3), and thus the prior on the \( \lambda_i \)'s is defined over the simplex (2.3) due to the one-to-one transformation in (2.5). In incorporating expert judgment, the prior distribution for \( \lambda_i \) can be obtained via
this transformation in conjunction with (1.4) as
\[
\Pi \{ \lambda_i | D^{(0)} \} = \frac{\Gamma(\beta)}{\Gamma(\beta(1 - \alpha_i^*)^*)} e^{-\lambda_i(\beta(1 - \alpha_i^*)^*)^*} \\
\times (1 - e^{-\lambda_i})^{\beta(1 - \alpha_i^*)^{x_i}} - 1, \tag{2.6}
\]
where \( \alpha_i^* \) and \( D^{(0)} \) are defined as in Section 1. It is more convenient, however, to work with the distributions involving \( R_i \) and the inverse transform of (2.5). Again alternative strategies exist for incorporating expert judgment. If direct elicitation of failure rates is possible, then (1.4) can be used to obtain appropriate expressions for incorporating expert judgment. Expressions for additive differences in \( x_i \) can also be obtained using \( x_i - x_j = -\ln(R_i/R_j)(i < j) \) and (1.6). Equating such expressions to elicited information will require numerical routines to obtain prior parameter values.

Often it is more convenient for experts to think in terms of reliability. It can be shown that, for a fixed mission time \( x_i \), the reliability in the \( i \)th testing stage is \( R_i \) and the relative improvements in reliability from stage \( i \) to stage \( j(i < j) \) is expressed as \((R_i/R_j)^{x_i}\). Thus, again (1.4) and (1.6) can be used to obtain the appropriate expressions.

Similar to the attribute-data case, once the prior parameters have been specified, several quantities can be determined for use in pre-testing analysis. The predictive density of \( x_i \) can be obtained as
\[
p(x_i|D^{(0)}) = \frac{\Gamma(\beta)}{\Gamma(\beta | x_i)} \frac{\Gamma(\beta \alpha_i^* + x_i)}{\Gamma(\beta \alpha_i^*)!} \times \left[ \Psi(\beta + x_i) - \Psi(\beta \alpha_i^* + x_i) \right], \tag{2.7}
\]
where \( \Psi(\cdot) \) is the digamma function and the predictive reliability at stage \( i \) is given as
\[
F(x_i|D^{(0)}) = \frac{\Gamma(\beta)}{\Gamma(\beta | x_i)} \frac{\Gamma(\beta \alpha_i^* + x_i)}{\Gamma(\beta \alpha_i^*)!} \times \left[ \frac{1}{\beta} \frac{\beta + (n_i - 1)\tau}{\beta + (n_i - 1)\tau} - \frac{1}{\beta} \frac{\beta + (n_i - 1)\tau}{\beta + (n_i - 1)\tau} \right]. \tag{2.8}
\]
For a given test time \( \tau \), assessment of the number of items tested in stage \( i \) is made via the predictive distribution
\[
Pr\{N_i = n_i|D^{(0)}\} = \frac{\Gamma(\beta)}{\Gamma(\beta | x_i)} \times \left\{ \frac{\Gamma(\beta \alpha_i^* + (n_i - 1)\tau)}{\Gamma(\beta + (n_i - 1)\tau)} - \frac{\Gamma(\beta \alpha_i^* + n_i\tau)}{\Gamma(\beta + n_i\tau)} \right\}. \tag{2.9}
\]
This distribution can be used to study the effects of truncation time versus expected number of items to be tested when designing the reliability growth test. For example, it is easy to show in (2.9) that, as \( \tau \to \infty \), \( Pr\{N_i = 1|D^{(0)}\} \to 1 \).

Under the specified testing scenario as implied via (2.1), the likelihood of \( R_j \) given a single failure at \( x_j < \tau \) and \( (n_j - 1) \) survivals at the \( j \)th testing stage is
\[
L(R_j; x_j, n_j) = \ln(1/R_j)R_j^{T_j}, \tag{2.10}
\]
where \( T_j = x_j + \tau(n_j - 1) \). Defining \( D^{(i)} \equiv \{D^{(0)}, x_1, n_1, x_2, n_2, \ldots, x_i, n_i\} \), the posterior distribution of \( R \) can be obtained as
\[
\Pi(R|D^{(i)}) \propto \prod_{j=1}^{m+2} \left( R_j - R_{j-1} \right)^{\beta \alpha_j^{(i)}} \times \left( \prod_{j=1}^{i} \ln(1/R_j)R_j^{T_j} \right). \tag{2.11}
\]
In this case, the posterior distribution \( \Pi(R_i; \ldots, R_{m+1}|D^{(i)}) \) and therefore \( \Pi(x_i; \ldots, x_{m+1}|D^{(i)}) \), along with the marginals, cannot be obtained in any type of tractable analytical form (e.g., see Robinson and Dietrich 1989). Thus, one way to make full Bayesian inference in this case is to use the MCMC methods, as will be discussed in Section 3.

2.2 Weibull Failure Data

Results similar to those of Section 2.1 can be obtained for the case in which \( x_i \), the life-length of the system during the \( i \)th stage of testing, is described by the Weibull failure model
\[
p(x_i|\lambda_i, \phi) = \lambda_i \phi x_i^{-\phi - 1} \exp(-\lambda_i x_i^{\phi}), \tag{2.12}
\]
where \( \lambda_i > 0 \) and \( \phi > 0 \). Note that this model assumes that the aging characteristics, as modeled through \( \phi \), are constant from stage to stage but that the scale parameter, \( \lambda_i \), may change from one stage to another. Reliability growth then, is modeled by assuming the ordering (2.3) on the \( \lambda_i \)'s. Again using the parameterization defined in (2.5), the ordered Dirichlet prior (1.3) can be employed. Note that, because \( \phi \) is fixed over the testing stages, this implies that \{\( \lambda_i \)\} and therefore the failure rates over the testing stages, \( \lambda_i \phi x_i^{\phi - 1} \), are ordered (nonincreasing) in \( i \). Within each stage, however, the failure characteristics can be either decreasing (\( \phi < 1 \)), constant (\( \phi = 1 \)), or increasing (\( \phi > 1 \)). This provides more flexibility in the modeling process. The preceding model will be referred to as the Weibull reliability growth (WRG) model.

Development of prior parameters follows along the lines of Section 2.1 but conditioned on \( \phi \). A priori, \( \phi \) is assumed independent of \{\( R_i \)\}, so we can separately define a prior distribution on \( \phi \) reflecting the believed aging characteristics. Next we focus on reliability. For the WRG model, the reliability in the \( i \)th testing stage is \( R_i^{\phi} \), and the relative improvements in reliability from stage \( i \) to stage \( j(i < j) \) are expressed as \((R_i/R_j)^{x_i}\). Again the distributions of (1.4) and (1.6) can be used to obtain the appropriate expressions for mean, variances, and so forth. These expressions, however, will be conditioned on \( \phi \) and numerical or Monte Carlo techniques will have to be used for solving for prior parameters.

Using the inverse transform of (2.5), the likelihood of \( R_j \) and \( \phi \), given a single failure and \( (n_j - 1) \) survivals at the \( j \)th testing stage, is
\[
L(R_j; \phi, x_j, n_j) = \ln(1/R_j)\phi x_j^{-\phi - 1} R_j^{T_j(\phi)}, \tag{2.13}
\]
where \( T_j(\phi) = x_j^{\phi} + (n_j - 1)\tau^{\phi} \). Again, defining \( D^{(i)} \equiv \{D^{(0)}, x_1, n_1, x_2, n_2, \ldots, x_i, n_i\} \), the posterior distribution
of $\mathbf{R}$ and $\phi$ can be obtained as

$$
\Pi(\mathbf{R}, \phi | D^{(i)})
\propto \prod_{j=1}^{m+2} (R_j - R_{j-1})^{\beta \alpha_j - 1}
\times \left( \prod_{j=1}^{i} \ln(1/R_j) \phi x_j^{\phi - 1} R_j^{\phi (\phi - 1)} \right) \Pi(\phi | D^{(0)}),
$$

(2.14)

where $\Pi(\phi | D^{(0)})$ is the prior of $\phi$. The posterior distributions of $\mathbf{R}$ and $\phi$ cannot be obtained in analytical form for any choice of $\Pi(\phi | D^{(0)})$ (e.g., see Soland 1969) but, as shown in Section 3, the MCMC methods can be easily adapted to this case.

### 3. BAYESIAN INFERENCE VIA MONTE CARLO METHODS

In this section, we present an overview of the use of the Gibbs sampler for making posterior and predictive inference with the reliability growth models of Sections 1 and 2. The attractive feature of the Gibbs sampler is that it enables the generation of samples from the posterior distributions without having to obtain the exact distributional forms. For a more detailed discussion of the Gibbs sampler and other related Monte Carlo methods, see Gelfand and Smith (1990) and Casella and George (1992).

#### 3.1 Implementation of Gibbs Sampling for the Attribute-Data Reliability Growth Model

Consider inference for the attribute data reliability growth model of Section 1. After $i$ stages of testing, it is desired to obtain the posterior distribution $\Pi(\mathbf{R} | D^{(i)})$ and its marginal distributions $\Pi(R_j | D^{(i)})$, $j = 1, \ldots, m + 1$. The Gibbs sampler enables the drawing of samples from $\Pi(\mathbf{R} | D^{(i)})$ without actually computing the exact distributional form. This is achieved by successive drawings from the full conditional distributions $\Pi(R_j | R(-j), D^{(i)})$, where $R(-j) = \{R_l | l \neq j, l = 1, 2, \ldots, m + 1\}$.

The process starts with a vector of arbitrary starting values $\mathbf{R}^0 = (R_1^0, R_2^0, \ldots, R_{m+1}^0)$ and

- draws $R_1^1$ from $\Pi(R_1 | R(-1), R_2^0, \ldots, R_m^0, R_{m+1}^0, D^{(i)})$
- draws $R_2^1$ from $\Pi(R_2 | R_1^1, R_3^0, \ldots, R_m^0, R_{m+1}^0, D^{(i)})$
- $\vdots$
- draws $R_{m+1}^1$ from $\Pi(R_{m+1} | R_1^1, R_2^1, \ldots, R_m^1, D^{(i)})$.

(3.1)

As a result of this single iteration of the Gibbs sampler in (3.1), a single vector that represents a transition from the starting value $\mathbf{R}^0 = (R_1^0, R_2^0, \ldots, R_{m+1}^0)$ to $\mathbf{R}^1 = (R_1^1, R_2^1, \ldots, R_{m+1}^1)$ has been generated. If this iteration is repeated $k$ times (i.e., next starting with $\mathbf{R}^1$ and iterating to $\mathbf{R}^2$, and so on), then the Gibbs sequence,

$$
\mathbf{R}^0, \mathbf{R}^1, \mathbf{R}^2, \ldots, \mathbf{R}^k,
$$

(3.2)

is generated. Under some mild regularity conditions (see Smith and Roberts 1993), distribution of $\mathbf{R}^k$ converges to $\Pi(\mathbf{R} | D^{(i)})$ as $k \to \infty$, and thus $\mathbf{R}^k$ is a sample point from $\Pi(\mathbf{R} | D^{(i)})$.

To generate a sample from $\Pi(\mathbf{R} | D^{(i)})$, one alternative is to generate $s$ independent Gibbs sequences of $k$ iterations each and use the $k$th value from each sequence. An alternate approach is to generate a single long Gibbs sequence and, after disregarding an initial burn-in sample, collect every $ith$ value from the remaining sequence (see Smith and Roberts (1993) for a detailed discussion of other implementation issues and Gilks, Richardson, and Spiegelhalter (1996) for various applications of MCMC methods).

For $j = 1, \ldots, i$, the exact form of $\Pi(R_j | R(-j), D^{(i)})$ is not known and thus, to facilitate the implementation of the Gibbs sampler, some random-variable generation method must be employed. Such methods include the standard rejection sampling method of the type described by Smith and Gelfand (1992), the importance-sampling-type methods discussed by Mueller (1992), the adaptive rejection procedure of Gilks and Wild (1992), or the Metropolis algorithm of Chib and Greenberg (1995).

Due to the structure of our models, the standard rejection sampling method can be easily implemented. Following Smith and Gelfand (1992), to sample from the conditional posterior $\Pi(R_j | R(-j), D^{(i)})$, $j = 1, \ldots, i$, as required in (3.1), we proceed as follows:

**Step 1:** For $j = 1, \ldots, i$, generate $R_j$ from the prior conditional distribution $\Pi(R_j | R(-j), D^{(i)})$.

**Step 2:** Generate independently a $\beta$ uniform random variate $u$.

**Step 3:** Compute the ratio

$$
\frac{\mathcal{L}(R_j; \mathbf{R}(-j), D^{(i)})}{\mathcal{L}(R_j; \mathbf{R}(-j), D^{(i)})}
$$

(3.5)

where, for $j = 1, \ldots, i$, the conditional likelihood

$$
\mathcal{L}(R_j; \mathbf{R}(-j), D^{(i)})
\propto \begin{cases} 
(1 - R_j) R_j^{\phi - 1} & R_{j-1} \leq R_j \leq R_{j+1} \\
0 & \text{otherwise}
\end{cases}
$$

(3.6)
and for \( n_j > 1 \), \( \hat{R}_j \), the maximum of (3.6), is available analytically as

\[
\hat{R}_j = \begin{cases} 
  R_{j-1} & \text{if } \frac{n_{j-1}}{n_j} \leq R_{j-1} \\
  \frac{n_{j-1}}{n_j} & \text{if } R_{j-1} < \frac{n_{j-1}}{n_j} < R_{j+1} \\
  R_{j+1} & \text{if } R_{j+1} \leq \frac{n_{j-1}}{n_j}.
\end{cases}
\]  

(3.7)

For the case \( n_j = 1 \), \( \hat{R}_j \) is the value of \( R_{j-1} \) generated at the previous iteration of the Gibbs sampler.

**Step 4:** Accept \( R_j \) if

\[
u \leq \frac{L(R_j; R^{(-j)}, D^{(i)})}{L(R_j; R^{(-j)}, D^{(i)})};
\]

otherwise reject \( R_j \) and go to step 1 and repeat the process.

The likelihood ratio (3.5) is the acceptance probability, implying that those values of \( R_j \)'s drawn from the prior will be more likely to be included in the posterior sample if their likelihood contribution is high. The preceding rejection method uses the conditional prior as the importance function and therefore can be viewed as a special case of the Gibbs-importance-sampling algorithm of Mueller (1992). This procedure is reasonably efficient for well-behaved likelihoods as in our case; however, more efficient and robust techniques were discussed by Gilks and Wild (1992) and Mueller (1992).

For \( j = i + 1, \ldots, m + 1 \), the exact form of the full conditionals is known because they are identical to the prior conditionals

\[
\Pi(R_j|R^{(-j)}, D^{(i)}) = \Pi(R_j|R_{j-1}, R_{j+1}, D^{(i)})
\]

for \( j = i + 1, \ldots, m + 1 \), (3.8)

whose distribution is given by (3.4). This simplifies drawing from these distributions for \( j = i + 1, \ldots, m + 1 \) because the standard rejection sampling method need not be employed.

Once the sample \( R_1, R_2, \ldots, R^r \) from \( \Pi(R_j|D^{(i)}) \) is formed, one can obtain all the marginal posterior distributions of \( \lambda_j \)'s by using (2.5) for \( j = 1, \ldots, m + 1 \). The predictive distribution for future \( X_j \)'s \( p(x_j|D^{(i)}), j = i + 1, \ldots, m + 1 \), is computed using (2.4) as

\[
p(x_j|D^{(i)}) \approx \frac{1}{r} \sum_{k=1}^{r} \ln(1/R^k_j)(R^k_j)^{x_j}.
\]

(3.14)

The posterior distribution of the reliability at any stage \( j \), for a given mission time \( x \), is also straightforward via the identity \( e^{-\lambda_j x} = R^x_j \). The mean time to failure can be approximated by

\[
E[X_j|D^{(i)}] \approx \frac{1}{r} \sum_{k=1}^{r} \frac{1}{1 - \ln(R^k_j)}, \quad j > i.
\]

(3.15)

Predictions about \( N_j \)'s for \( j = i + 1, \ldots, m + 1 \) can also be made via

\[
Pr\{N_j = n_j|D^{(i)}\} \approx \frac{1}{r} \sum_{k=1}^{r} (1 - R^k_j)^{n_j-1}(R^k_j)^{n_j},
\]

(3.16)

and the expectation is given by

\[
E[N_j|D^{(i)}] \approx \frac{1}{r} \sum_{k=1}^{r} \frac{1}{1 - (R^k_j)^{n_j}}.
\]

(3.17)

### 3.3 Inference for the WRG Model

In implementing the Gibbs sampler for the WRG model, it is necessary to draw from the full conditional distributions \( \Pi(R_j|R^{(-j)}, \phi, D^{(i)}) \), for \( j = 1, \ldots, m + 1 \), and

| Table 1. Actual Number of Copies Tested in Each Stage |
|-----------------|------|------|------|------|------|------|------|
| Testing stage j | 1    | 2    | 3    | 4    | 5    | 6    | 7    |
| \( n_j \)       | 1    | 1    | 1    | 1    | 3    | 4    | 7    | 10   |

TECHNOMETRICS, FEBRUARY 1998, VOL. 40, NO. 1
Table 2. Comparison of Posterior Expected Reliabilities

<table>
<thead>
<tr>
<th>Testing stage j</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>(E[\theta</td>
<td>D(j)]) - exact</td>
<td>0.324</td>
<td>0.638</td>
<td>0.737</td>
<td>0.840</td>
<td>0.856</td>
<td>0.873</td>
<td>0.892</td>
<td>0.912</td>
<td>0.935</td>
<td>0.957</td>
</tr>
<tr>
<td>(E[\theta</td>
<td>D(j)]) - Gibbs</td>
<td>0.326</td>
<td>0.638</td>
<td>0.737</td>
<td>0.840</td>
<td>0.855</td>
<td>0.873</td>
<td>0.892</td>
<td>0.912</td>
<td>0.934</td>
<td>0.957</td>
</tr>
</tbody>
</table>

As before, for stages \(j = 1, \ldots, i\), the conditional distribution for \(R_j\) is not of a familiar form and therefore the standard rejection sampling method of Section 3.1 will be employed. It can be shown that the conditional likelihoods for \(R_j, j = 1, \ldots, i\), are given by (3.12) with maximums (3.13), where \(T_j(\phi)\) replaces \(T_j(\phi)\). For stages \(j = i + 1, \ldots, m + 1\), the full conditionals are given by (3.4) because \(\phi\) and \(R\) are assumed independent, a priori.

An additional step for this model is needed to complete the Gibbs sequence. This step involves generating \(\phi\) from its full conditional posterior distribution, \(\Pi(\phi|R, D(i)) = \Pi(\phi|R_1, R_2, \ldots, R_i, D(i))\). To draw a sample from \(\Pi(\phi|R, D(i))\), one can again use the standard rejection sampling method with conditional likelihood

\[L(\phi; R, D(i)) \propto \left( \prod_{j=1}^i \phi^{x_j} \frac{R_j^{T_j(\phi)} - 1}{R_j^{T_j(\phi)}} \right). \tag{3.18}\]

The maximum of the preceding conditional likelihood must be obtained numerically. This is a straightforward task, because \(L(\phi; R, D(i))\) is concave in \(\phi\). The more efficient adaptive rejection procedure of Gilks and Wild (1992) can also be used. Distributional results similar to those obtained in (3.14)-(3.17) can also be obtained for the WRG model.

4. ILLUSTRATIONS

Example 1: Comparison of Gibbs and Analytical Results for Attribute Reliability Growth Data

Van Dorp et al. (1997) assumed the reliability growth model of Section 1 and used it for developing optimal stopping rules during a system-development phase. The inferences for stage reliabilities and the number of items to be tested in future stages were obtained using the closed-form expressions involving ratios of gamma functions.

The authors considered a testing scenario of \(m = 10\) stages. Using cost-based utility functions, it was shown that the optimal decision was to stop after the eighth testing stage. The prior parameters were specified as \(\alpha = (0.36, 0.34, 0.102, 0.0985, 0.0981, 0.0128, 0.0127, 0.0125, 0.0124, 0.0123, 0.0122, 0.0120, 0.0120)\) and \(\beta = 50\). The test data from the development process are given in Table 1.

After eight stages of testing, a comparison of the analytical results of van Dorp et al. (1997) and the Gibbs sampling results using the techniques of Section 3.1 is given in Table 2. The Gibbs sampling results are obtained after 5,000 iterations. The convergence of the Gibbs sampler was determined by using different starting values for \(R_j\)’s and by monitoring the ergodic averages of the \(R_j\)’s. We note that our results are exactly the same as the analytical results except in a few instances in which the differences are most likely due to round-off errors. Similar accuracy was obtained at the distributional level for posterior distributions of \(R_j|D(j), j = 8, \ldots, 11\), and the predictive distributions for \(N_j|D(8), j = 8, \ldots, 11\). Distributional results for \(R_j|D(j), j = 1, \ldots, 8\), are not obtainable in analytical form but are available from the Gibbs sampling results.

Example 2: Use of the Gibbs Sampling Approach for Exponential Reliability Growth Data

We next consider the tank-failure data analyzed by Sen and Bhattacharyya (1993). The authors’ analysis of the data favors both reliability growth and exponentiality of the time between failures. In this example, “time” is measured in number of miles accumulated. In what follows, we will consider the data from the 24 stages of testing and at intermittent stages make inferences about the failure behavior during the 25th stage (or field results). That is, it is assumed that the development process lasts for 24 stages, and we make predictive inference about the failure behavior during the 25th stage prior to testing and after 5, 10, and 24 stages of testing. In so doing, we will use the exponential reliability growth model of Section 2.1. The data are given in Table 3.

Table 3. Tank-Failure Data in Miles to Failure in Stage j

<table>
<thead>
<tr>
<th>Testing stage j</th>
<th>(x_j)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>56</td>
</tr>
<tr>
<td>3</td>
<td>195</td>
</tr>
<tr>
<td>4</td>
<td>58</td>
</tr>
<tr>
<td>5</td>
<td>175</td>
</tr>
<tr>
<td>6</td>
<td>208</td>
</tr>
<tr>
<td>7</td>
<td>27</td>
</tr>
<tr>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>9</td>
<td>52</td>
</tr>
<tr>
<td>10</td>
<td>249</td>
</tr>
<tr>
<td>11</td>
<td>533</td>
</tr>
<tr>
<td>12</td>
<td>205</td>
</tr>
<tr>
<td>13</td>
<td>27</td>
</tr>
<tr>
<td>14</td>
<td>145</td>
</tr>
<tr>
<td>15</td>
<td>92</td>
</tr>
<tr>
<td>16</td>
<td>35</td>
</tr>
<tr>
<td>17</td>
<td>224</td>
</tr>
<tr>
<td>18</td>
<td>134</td>
</tr>
<tr>
<td>19</td>
<td>137</td>
</tr>
<tr>
<td>20</td>
<td>526</td>
</tr>
<tr>
<td>21</td>
<td>372</td>
</tr>
<tr>
<td>22</td>
<td>168</td>
</tr>
<tr>
<td>23</td>
<td>620</td>
</tr>
<tr>
<td>24</td>
<td>330</td>
</tr>
</tbody>
</table>

Table 4. Comparison of \(P(R_{25} > .90|D(i))\) for \(x = 25, 50, 100\), and 200 Miles

| \(P(R_{25} > .90|D(i))\) | \(x = 25\) | \(x = 50\) | \(x = 100\) | \(x = 200\) |
|--------------------------|-------------|-------------|-------------|-------------|
| \(P(R_{25} > .90|D(0))\) | .1020       | .0512       | .0266       | .0138       |
| \(P(R_{25} > .90|D(3))\) | .8164       | .5804       | .3582       | .1964       |
| \(P(R_{25} > .90|D(6))\) | .9022       | .7308       | .4932       | .2906       |
| \(P(R_{25} > .90|D(9))\) | .9996       | .9384       | .6782       | .3916       |

TECHNOMETRICS, FEBRUARY 1998, VOL. 40, NO. 1
in terms of number of miles accumulated to failure in each stage.

In this example we illustrate both the usefulness of our approach and its comparability with the discrete power law model of Sen and Bhattacharyya (1993). The discrete power law model reflects the so-called “Duane postulate,” which is the basis of all the nonhomogeneous Poisson process (NHPP) reliability growth models. The discrete nature of this model, however, makes it more realistic than the NHPP models, which fail to explicitly reflect the effect of the test/modify process.

To compare our results with those of Sen and Bhattacharyya (1993), we assume a joint ordered uniform prior for $R_j$'s by specifying $\beta R_j = 1$ for all $j$ in (1.3). Such a choice implies $E[R_j|D^{(0)}] = j/(m + 2)$ (where $m = 24$)—that is, constant growth in $R_j$'s. Using the transformation in (2.5), the prior means of $\lambda_j$'s can be obtained as $E[\lambda_j|D^{(0)}] \approx \ln((m + 2)/j)$, implying geometrically decreasing failure rates as in the discrete power law model of Sen and Bhattacharyya (1993).

The reliability growth test for this example is designed with $\tau \to \infty$, and thus only one system is tested to failure at each stage. If this were not the case, we could have used (2.9) to explore the number of systems to be tested at each stage versus the required test truncation time. Although many requirements may be imposed, for illustrative purposes we assume that interest centers on mission reliabilities for 25, 50, 100, and 200 miles. The prior predictive (prior to testing) mission reliability assessments for stage 25 are given in the first row of Table 4 using the ordered uniform prior on the $R_j$'s. Specifically, the first row shows our prior probabilities (before testing) that the reliability at the 25th stage of testing is at least .90 for a given mission time. Similar predictive probability statements can be made for failure rates as well as other mission values. As shown by van Dorp et al. (1997), these are useful for determining when to terminate the development process.

Once test data become available, they may be used to update the prior information and provide a revision of the uncertainty regarding the system reliability. Using the Gibbs sampler, after any stage of testing we can make probability statements about the reliability for any mission time at any given stage of testing (past, current, or future) using the generated marginal posterior distributions. Table 4 illustrates how our predictive probability statements about $R_{25}$, the reliability at mission length $x$, can change after observing data from each stage of testing. We note that as the testing progresses we assess a better-performing system.

In Figure 1, we show the posterior predictive distributions for the reliability for a mission length of 50 miles in stage 25, after 1, 5, 10, and 24 stages of testing. These results are based on 5,000 iterations of the Gibbs sampler for the 25-dimensional joint distribution of $R_j$'s. The entire analysis was completed in a few minutes using a Pentium PC. The change in the uncertainty about the final-stage reliability and the concentration of the posterior distribution is clearly seen from the distribution plots. In Figure 2, we illustrate the reliability growth curves—that is, the plots of $E[R_j|D^{(i)}]$ versus $j$—prior to any testing ($i = 0$) and after $i = 5, 10,$ and 24 stages of testing. This figure shows that there was more steady reliability growth than expected prior to testing.
Prior and smoothed posterior predictive densities of \( \lambda_{25} \) after 24 stages of testing are shown in Figure 3. The posterior density was obtained using the transformation \( \lambda_{25} = \ln(1/R_{25}) \) and smoothing the probability histogram. Again, the figure shows the decrease in uncertainty about the failure rate and the improvement in reliability as a result of the 24 stages of testing.

In Figure 4, we compare the expected posterior failure rates, \( E[\lambda_j|D^{(24)}], j = 1, 2, \ldots, 24 \), from the ERG model with the maximum likelihood estimates of \( \lambda_j \) obtained by Sen and Bhattacharyya (1993). We note that the failure-rate estimates from both models are fairly close except for stage 1 for which the Sen and Bhattacharyya estimate is extremely low in light of the failure time of one mile in stage 1. The difference in the results for this stage is due to the increased flexibility of the ERG model. After several stages of testing, however, the models produce very similar point estimates. In addition to incorporation of prior information, an advantage of the ERG model is its development of full posterior and predictive distributions for the relevant quantities of interest.

ACKNOWLEDGMENTS

This work was completed while Refik Soyer was visiting the Institute of Statistics and Decision Sciences, Duke University.

[Received June 1995. Revised April 1997.]
Bayesian Computations for a Class of Reliability Growth Models


