BAYESIAN RELIABILITY MODELING FOR MASKED SYSTEM LIFETIME DATA

Lynn KUO*

Department of Statistics, University of Connecticut, Storrs, CT 06269-3120, USA

Tae Young YANG †

Department of Mathematics, Myongji University, Kyunggi, 449-728 South Korea.

Abstract

In the masked system lifetime data, the exact component that causes the system's failure is often unknown. For each series system at test, we observe its system's failure time and a set of components that includes the component actually causing the system to fail. The objective is to make inferences for the reliability of the components. In this paper we consider various probability models for the conditional masking probabilities that identify the set of possible failed components given the true cause of failure and the system's failure time. In addition to exponential distributions for the component lifetimes, we consider Weibull distributions. A Bayesian approach that uses Gibbs sampling will be developed for each of the models. Model selection by a predictive approach will also be developed. We show that improved inference can be obtained by modeling the masking probabilities.

Key Words: Competing risk; Gibbs sampling; Incomplete data; Latent variable model

Abbreviated Title: Reliability Inference for Masked System Lifetime Data.

1. Introduction

In the masked system lifetime data, the exact component that causes the system to fail is often unknown. Masking is usually due to limited resources for diagnosing the causes of failures. More importantly, it is due to the modular nature of present systems. For example, in computer repair shops, the cause of failure is often isolated to a module of several components. The system is brought back into operation by replacing the whole module. For each system at test, we observe its system's failure time and a set of components that includes the component actually causing the system to fail. The objective is to make inferences for the

*Correspondence to: Lynn Kuo, Department of Statistics, University of Connecticut, Storrs, CT 06269-3120, USA.

†Research supported by Korea Science and Engineering Foundation 971-0105-027-2.

reliability of the components. The system reliability is usually derived from the component reliability and the system's configuration.

Recently, several authors studied this problem for the series system using exponential distributions for the component lifetimes. Most authors make an equiprobable assumption for the masking probabilities where the conditional masking probabilities take the same value given each cause of failure in the masked set. This assumption (called the symmetry assumption by some authors) allows one to consider only the reduced likelihood. The work includes Miyakawa (1984), Usher and Hodgson (1988), Guess *et al.* (1991), Lin *et al.* (1993, 1996), Doganaksoy (1991), Gastaldi (1994), Reiser *et al.* (1995), Usher (1996), and Mukhopadhyay and Basu (1997).

To relax the equiprobable assumption, Lin and Guess (1994), and Guttman *et al.* (1995) consider a proportional probability model for a dependent masking probability for the two-component series system. They assume that the masking probability conditional on the first component's failure and the system's failure time is proportional to that of the second component. Although both masking probabilities are functions of the system's failure time, they assume the proportionality is independent of the system's failure time. We explore two different stochastic models for the masking probabilities. Both models relax the equiprobable assumption. One model assumes that the masking probabilities depend on the true cause of failure, not on the system's failure time. The other model assumes the masking probabilities are not only functions of the true cause of failures but also monotone functions of the system's failure time. In addition to the exponential distribution for the component's failure time, we also consider Weibull distributions. To ensure identifiability, we assume the components have independent failure distributions. The main theme of this paper is to show that better predictive power can be achieved by modeling the masking probabilities without the equiprobable assumption. We demonstrate this with a simulated data set and with a real data set.

With the recent advances in Gibbs sampling (Geman and Geman, 1984, Tanner and Wang, 1987, and Gelfand and Smith, 1990), Bayesian computation even for quite complex stochastic systems can be carried out by the Markov chain Monte Carlo (MCMC) method. We develop MCMC algorithms for Bayesian inference for the problems posed here. A Gibbs sampling procedure augmented with latent variables that simulates the true causes of failures will be used for Bayesian computation of component reliability. In addition, we develop an EM algorithm for this problem.

Sections 2 and 3 describe our models and the Bayesian computation techniques. Section 4 discusses the model selection criterion that uses the pseudo-Bayes factor. Some numerical results for both simulated and real data are given to illustrate that our model can improve upon the existing models (with the equiprobable assumption) in Sections 5 and 6.

2. Full Likelihood

In the full likelihood formulation, we not only consider the system lifetimes but also the labels that identify the possible failed components. Let t_{ij} denote the random lifetime of the j^{th} component in the i^{th} system. Assume we observe a sample of n series systems each having J components; therefore we observe $t_i = \min(t_{i1}, \ldots, t_{iJ})$, for $i = 1, \ldots, n$. In addition, for each system, we observe a set of labels of the components that include the failed component. Let S_i denote the set of components (labels) that possibly cause the system i to fail, and let s_i denote the realized set for S_i. The observed data denoted by $\mathbf{d}$ consist of $(t_1, s_1), \ldots, (t_n, s_n)$. If each set s_i is a singleton, then this reduces to the usual competing risk study where the exact cause of failure is identified. The set of labels is often masked because the true cause of failure is hiding with other components. Let K_i denote the index of the component actually causing the i^{th} system to fail. It is a singleton because of the series system and continuous lifetimes. Let f_j denote the density for the j^{th} component's lifetime and $\overline{F}_j$ denote its survival function. We assume the component lifetime distributions are independent. For the ith system, the likelihood contribution coming from the datum (t_i, s_i) consists of $p(t_i, s_i)$. This can be written as $\sum_{j=1}^{J} p(t_i, K_i = j) p(s_i|t_i, K_i = j)$. Note the masking probability $p(s_i|t_i, K_i = j) = 0$, if $j \notin s_i$. Moreover, we have: $p(t_i, K_i = j) = f_j(t_i) \prod_{l=1, l\neq j}^{J} \overline{F}_l(t_i)$. Therefore the full likelihood of this data set is

$$L_F = \prod_{i=1}^{n} \left[\sum_{j \in s_i} \left\{ \left(f_j(t_i) \prod_{l=1, l\neq j}^{J} \overline{F}_l(t_i) \right) P(S_i = s_i | T_i = t_i, K_i = j) \right\} \right].$$

It is often of interest to compute the diagnostic probability, that is the probability that the j^{th} component actually causing the system to fail given the possible malfunctioning subset and the system's failure time. For the i^{th} system and $j \in s_i$, the diagnostic probability for the j^{th} cause is

$$p(k_i = j | s_i, t_i) = \frac{p(k_i = j, s_i, t_i)}{p(s_i, t_i)} = \frac{P(S_i = s_i | T_i = t_i, K_i = j) f_j(t_i) \prod_{l=1, l\neq j}^{J} \overline{F}_l(t_i)}{\sum_{j \in s_i} P(S_i = s_i | T_i = t_i, K_i = j) f_j(t_i) \prod_{l=1, l\neq j}^{J} \overline{F}_l(t_i)}. \tag{1}$$

Several authors assume an equiprobable assumption where the masking probabilities take on the same value irrelevant of the particular cause of failure. That is, they make the following assumption for a fixed j:

$$P(S_i = s_i | T_i = t_i, K_i = j') = P(S_i = s_i | T_i = t_i, K_i = j) \ \text{ for all } \ j' \in s_i.$$

Then they proceed with their analyses with a reduced likelihood that does not depend on the masking probabilities:

$$L_R = \prod_{i=1}^{n} \left[\sum_{j \in s_i} \left\{ f_j(t_i) \prod_{l=1, l\neq j}^{J} \overline{F}_l(t_i) \right\} \right].$$

We model the masking probabilities from the simple classification point of view. One of the simple assumptions is to make the probabilities independent of the time and dependent on the component. Let us look at the two-components system. Assume $P(S_i = \{1\} | t_i, K_i = 1) = \theta_1$, and $P(S_i = \{2\} | t_i, K_i = 2) = \theta_2$.

Consequently, $P(S_i = \{1,2\}|t_i, K_i = 1) = 1 - \theta_1$, and $P(S_i = \{1,2\}|t_i, K_i = 2) = 1 - \theta_2$. Then the full likelihood in this case is

$$L_F = \prod_{i|s_i=\{1\}} \theta_1 f_1(t_i)\overline{F}_2(t_i) \prod_{i|s_i=\{2\}} \theta_2 f_2(t_i)\overline{F}_1(t_i) \prod_{i|s_i=\{1,2\}} \left\{(1-\theta_1)f_1(t_i)\overline{F}_2(t_i) + (1-\theta_2)f_2(t_i)\overline{F}_1(t_i)\right\}. \tag{2}$$

Let n_1, n_2, and n_{12} denote the numbers of systems in the sample with true causes identified to be the first component, the second component, or masked (either component). Then the sample sizes n_1, n_2, and n_{12}, that are also the cardinalities of the indices in each product in (2), convey information for the θ's and consequently for the component lifetime distributions.

The same idea can be generalized to an arbitrary number of components. For example for $J = 3$, we define the following masking probabilities: $P(S_i = \{j\}|t_i, K_i = j) = \theta_j$ for all j, $P(S_i = \{1,2\}|t_i, K_i = 1) = \theta_{12|1}$, and $P(S_i = \{1,3\}|t_i, K_i = 1) = \theta_{13|1}$. Consequently, $P(S_i = \{1,2,3\}|t_i, K_i = 1) = 1 - \theta_1 - \theta_{12|1} - \theta_{13|1}$. Similarly for causes 2 and 3. These statements can be simplified if we know more about the configuration of the system. For example, components 1 and 2 are mounted on the same card separately from component 3. Then we can assume $\theta_{13|1} = \theta_{23|2} = \theta_{13|3} = \theta_{23|3} = 0$. There will be fewer parameters to be estimated. Aboul-Séoud and Usher (1996) describe an example of the modular system structures.

If the replaced modules are further sent to the factories for identifying the exact component causing the failure, we can produce fairly accurate estimates of the θ's from the follow-up study. Therefore, we can estimate f_j by assuming the θ's are known. Alternatively, we can combine the original data $(t_i, s_i), i = 1, \ldots, n$, with the follow-up data to obtain better inferences as done in Flehinger *et al.* (1996), Flehinger *et al.* (1998), and Reiser *et al.* (1996).

In addition to exponential distributions for the component lifetimes, we will also consider Weibull distributions for F_j. Gibbs sampling will be developed for Bayesian inference for the θ's and the parameters in f_j. Although we concentrate on the two-components system in this paper, the Gibbs sampler can be extended straightforwardly to any number of components. The method is akin to the Gibbs sampling used in many aggregated models (cf. Yang and Kuo, 1999).

2.1 Gibbs Sampling for the Full Model

Let us define the hazard function $h_j(t) = f_j(t)/\overline{F_j}(t)$ for all j. Then (2) can be rewritten as

$$L_F = \left(\prod_{i|s_i=\{1\}} \theta_1 h_1(t_i) \prod_{i|s_i=\{2\}} \theta_2 h_2(t_i) \prod_{i|s_i=\{1,2\}} \{(1-\theta_1)h_1(t_i) + (1-\theta_2)h_2(t_i)\} \right) \left(\prod_{i=1}^{n} \overline{F}_1(t_i)\overline{F}_2(t_i) \right). \tag{3}$$

The summation in the third product of this likelihood prevents us from updating the parameters in F easily. Therefore, we introduce latent variables z_i for the systems where the cause of failure is masked. Let z_i be a Bernoulli variable with the success $(z_i = 1)$ probability to be $p_i = (1-\theta_1)h_1(t_i)/[(1-\theta_1)h_1(t_i)+(1-\theta_2)h_2(t_i)]$ for i with $s_i = \{1,2\}$. It is interesting to note that this probability of success p_i is exactly the diagnostic probability $p(k_i = 1|s_i, t_i)$ in (1) of the true cause due to the first component conditional on $s_i = \{1,2\}$ and t_i being observed. Let $z_+ = \sum_{i|s_i=\{1,2\}} z_i$. Then the conditional probability of $\boldsymbol{\theta} = (\theta_1, \theta_2)$ and the unknown parameters $\boldsymbol{\lambda}$ in F's given $\mathbf{d}$ and the latent variables $\mathbf{z}$ is

$$\begin{aligned} f(\boldsymbol{\theta}, \boldsymbol{\lambda}|\mathbf{z}, \mathbf{d}) \quad \propto \quad & \theta_1^{n_1}(1-\theta_1)^{z_+}\theta_2^{n_2}(1-\theta_2)^{n_{12}-z_+} \\ & \times \left(\prod_{i|s_i=\{1\}} h_1(t_i) \prod_{i|s_i=\{2\}} h_2(t_i) \prod_{i|s_i=\{1,2\}} h_1(t_i)^{z_i} h_2(t_i)^{1-z_i} \right) \left(\prod_{i=1}^{n} \overline{F}_1(t_i)\overline{F}_2(t_i) \right) \pi(\boldsymbol{\theta}, \boldsymbol{\lambda}), \end{aligned}$$

where π denotes the prior density of $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$. Therefore, updating $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$ can be done easily and independently for independent prior on $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$. We are assuming that readers are familiar with the basics of MCMC algorithms as described in Gelfand and Smith (1990) and Tanner and Wong (1987). In the following, we only describe the transitional kernel in the MCMC algorithm. Now we write the Gibbs steps for a few special cases:

(1) The exponential component: $\overline{F}_j(t) = \exp(-\lambda_j t)$. We consider a prior with the following independent components for $j = 1$ and 2: $\theta_j \sim Beta(\alpha_j, \beta_j)$ and $\lambda_j \sim G(a_j, b_j)$, where $G(a_j, b_j)$ denotes a gamma density with mean a_j/b_j. The conjugate priors for θ_j and λ_j are chosen so that we can easily specify the conditional densities in the Gibbs algorithm. Gamma prior density and Beta prior density are also versatile enough to incorporate various shapes for the distributions of the unknown parameters.

Step a (data augmentation step): Given $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$, we generate $z_i \sim B(1, (1-\theta_1)\lambda_1/[(1-\theta_1)\lambda_1 + (1-\theta_2)\lambda_2])$, independently for i with $s_i = \{1,2\}$. Then we compute $z_+ = \sum_{i|s_i=\{1,2\}} z_i$.

Step b (parameter updating step): given $\mathbf{z}$ and $\mathbf{d}$, we update the parameters by:

$$\begin{aligned} \theta_1|\mathbf{z}, \mathbf{d}, \theta_2, \boldsymbol{\lambda} &\sim Beta(\alpha_1 + n_1, \ \beta_1 + z_+); \\ \theta_2|\mathbf{z}, \mathbf{d}, \theta_1, \boldsymbol{\lambda} &\sim Beta(\alpha_2 + n_2, \ \beta_2 + n_{12} - z_+); \\ \lambda_1|\mathbf{z}, \mathbf{d}, \lambda_2, \boldsymbol{\theta} &\sim G(a_1 + n_1 + z_+, \ b_1 + \sum_{i=1}^{n} t_i); \\ \lambda_2|\mathbf{z}, \mathbf{d}, \lambda_1, \boldsymbol{\theta} &\sim G(a_2 + n_2 + n_{12} - z_+, \ b_2 + \sum_{i=1}^{n} t_i). \end{aligned}$$

Note that the parameters $\theta_1, \theta_2, \lambda_1$ and λ_2 are conditionally independent given $\mathbf{z}$ and $\mathbf{d}$ from the above expressions. Therefore, updating the four parameters can be done in parallel.

(2) The Weibull component: $\overline{F}_j(t) = \exp(-\lambda_j t^{\rho_j})$. We consider a prior with the following independent components: $\theta_j \sim Beta(\alpha_j, \beta_j)$, $\lambda_j \sim G(a_j, b_j)$, and ρ_j has density $\pi(\rho_j)$ that can be arbitrary. The same ideas can be generalized to a system with more than two components. The assumption of independent components in the prior is chosen for convenience. We have chosen the gamma distribution for λ_j for conjugacy. We can choose the prior on ρ_j to be uniformly distributed over $(0, m_j)$ as suggested in Sinha (1986, p.160). We can also explore the choice $\pi(\rho_j, \lambda_j) \propto 1/(\rho_j\lambda_j)$ as given in Sinha and Zellner (1990). Let $\boldsymbol{\rho} = (\rho_1, \rho_2)$.

Step a (data augmentation step): Given $\boldsymbol{\theta}$, $\boldsymbol{\lambda}$ and $\boldsymbol{\rho}$, we generate $z_i \sim B(1, (1-\theta_1)\lambda_1\rho_1 t_i^{\rho_1 - 1}/[(1-\theta_1)\lambda_1\rho_1 t_i^{\rho_1-1} + (1-\theta_2)\lambda_2\rho_2 t_i^{\rho_2-1}])$, independently for i with $s_i = \{1,2\}$. Then we compute $z_+ = \sum_{i|s_i=\{1,2\}} z_i$.

Step b (parameter updating step): given $\mathbf{z}$ and $\mathbf{d}$, we update the parameters by:

$$\begin{aligned}
\theta_1|\mathbf{z},\mathbf{d},\boldsymbol{\lambda},\boldsymbol{\rho},\theta_2 &\sim Beta(\alpha_1+n_1,\ \ \beta_1+z_+);\\
\theta_2|\mathbf{z},\mathbf{d},\boldsymbol{\lambda},\boldsymbol{\rho},\theta_1 &\sim Beta(\alpha_2+n_2,\ \ \beta_2+n_{12}-z_+);\\
\lambda_1|\mathbf{z},\mathbf{d},\boldsymbol{\rho},\boldsymbol{\theta},\lambda_2 &\sim G(a_1+n_1+z_+,\ \ b_1+\sum_{i=1}^{n} t_i^{\rho_1});\\
\lambda_2|\mathbf{z},\mathbf{d},\boldsymbol{\rho},\boldsymbol{\theta},\lambda_1 &\sim G(a_2+n_2+n_{12}-z_+,\ \ b_2+\sum_{i=1}^{n} t_i^{\rho_2});
\end{aligned}$$

and update the parameters ρ_1 and ρ_2 independently given $\mathbf{z},\mathbf{d},\boldsymbol{\lambda}$ and $\boldsymbol{\theta}$, each by the Metropolis law (cf. Metropolis *et al.*, 1953, and Kuo and Yang, 1995, 1996) with the following density as the target density:

$$\pi(\rho_1|\mathbf{z},\mathbf{d},\boldsymbol{\lambda},\boldsymbol{\theta}) \propto \rho_1^{n_1+z_+}\left(\prod_{i|s_i=\{1\}} t_i^{\rho_1-1}\right)\left(\prod_{i|s_i=\{1,2\}} t_i^{(\rho_1-1)z_i}\right)\exp(-\lambda_1\sum_{i=1}^{n} t_i^{\rho_1})\pi(\rho_1); \tag{4}$$

$$\pi(\rho_2|\mathbf{z},\mathbf{d},\boldsymbol{\lambda},\boldsymbol{\theta}) \propto \rho_2^{n_2+n_{12}-z_+}\left(\prod_{i|s_i=\{2\}} t_i^{\rho_2-1}\right)\left(\prod_{i|s_i=\{1,2\}} t_i^{(\rho_2-1)(1-z_i)}\right)\exp(-\lambda_2\sum_{i=1}^{n} t_i^{\rho_2})\pi(\rho_2). \tag{5}$$

Note in the above updating, the generation of $\boldsymbol{\theta}$ and $(\boldsymbol{\lambda},\boldsymbol{\rho})$ can be done independently.

To generalize the above discussion to three components, we can easily extend the same prior for $\boldsymbol{\lambda}$, $\boldsymbol{\rho}$ in the component lifetimes to more components. For the masking probabilities $\boldsymbol{\theta}_1=(\theta_1,\theta_{12|1},\theta_{13|1},\theta_{123|1})$ with $\theta_1+\theta_{12|1}+\theta_{13|1}+\theta_{123|1}=1$, a Dirichlet prior with parameters $(\alpha_1,\alpha_2,\alpha_3,\alpha_4)$ can be considered. Similarly, we can define independent Dirichlet priors for the other masking probabilities $\boldsymbol{\theta}_2$ and $\boldsymbol{\theta}_3$, where $j=2$ or $j=3$ is the true cause of failure. Then we can employ the same data augmentation idea as before and updating the parameters $\boldsymbol{\theta}_1,\boldsymbol{\theta}_2$, and $\boldsymbol{\theta}_3$ independently, each with an updated Dirichlet distribution.

2.2 EM Algorithm for the Full Model

If we are only interested in the posterior mode or the mode of the likelihood with their error estimates, we can apply the EM algorithm (Dempster *et al.*, 1977). The algorithm may be simpler than the Gibbs sampler in implementation, it also may converge faster. The EM algorithm essentially converts a complicated optimization problem to many simpler iterative steps. The iteration steps are similar to that in the Gibbs algorithm, except the data augmentation step is replaced by an E-step (expectation) where the missing data (latent variables) are replaced by their expectations, and the parameter updating step is replaced by a maximization step. In the following, we make the same prior assumption as in the previous subsection to obtain the posterior mode. The algorithm can be easily modified to obtain the maximum likelihood estimate using the flat prior.

(1) For the exponential component, we make the following changes in the $l+1^{st}$ iteration:

Step a:

$$z_+^l = n_{12}(1-\theta_1^l)\lambda_1^l/[(1-\theta_1^l)\lambda_1^l+(1-\theta_2^l)\lambda_2^l].$$

Step b:

$$\begin{aligned}
\theta_1^{l+1} &= \frac{\alpha_1 + n_1 - 1}{\alpha_1 + \beta_1 + n_1 + z_+^l - 2};\\
\theta_2^{l+1} &= \frac{\alpha_2 + n_2 - 1}{\alpha_2 + \beta_2 + n_2 + n_{12} - z_+^l - 2};\\
\lambda_1^{l+1} &= \frac{a_1 + n_1 + z_+^l - 1}{b_1 + \sum_{i=1}^n t_i};\\
\lambda_2^{l+1} &= \frac{a_2 + n_2 + n_{12} - z_+^l - 1}{b_2 + \sum_{i=1}^n t_i}.
\end{aligned}$$

(2) For the Weibull component:

Step a: the generation of z_i is replaced by setting $z_+^l = n_{12}(1-\theta_1^l)\lambda_1^l\rho_1^l t_i^{\rho_1^l - 1}/[(1-\theta_1^l)\lambda_1^l\rho_1^l t_i^{\rho_1^l - 1} + (1-\theta_2^l)\lambda_2^l\rho_2^l t_i^{\rho_2^l - 1}]$.

Step b: the generation of θ_1, θ_2, λ_1, and λ_2 is replaced by

$$\begin{aligned}
\theta_1^{l+1} &= \frac{\alpha_1 + n_1 - 1}{\alpha_1 + \beta_1 + n_1 + z_+^l - 2};\\
\theta_2^{l+1} &= \frac{\alpha_2 + n_2 - 1}{\alpha_2 + \beta_2 + n_2 + n_{12} - z_+^l - 2};\\
\lambda_1^{l+1} &= \frac{a_1 + n_1 + z_+^l - 1}{b_1 + \sum_{i=1}^n t_i^{\rho_1}};\\
\lambda_2^{l+1} &= \frac{a_2 + n_2 + n_{12} - z_+^l - 1}{b_2 + \sum_{i=1}^n t_i^{\rho_2}};
\end{aligned}$$

and the Metropolis algorithm for generating ρ_1 and ρ_2 is replaced by finding ρ_1 that maximizes (4) and finding ρ_2 that maximizes (5).

2.3 Gibbs Sampling for the Reduced Model

Now we write the Gibbs sampler for the reduced model with the equiprobable assumption. Observe $1-\theta_1 = 1-\theta_2 = 1-\theta$ in this case. We assume the same prior as the one in the full model except θ is $Beta(\alpha_1, \beta_1)$ independent of other parameters. So the likelihood in this situation reduces to

$$L_R = \theta^{n_1+n_2}(1-\theta)^{n_{12}} \left(\prod_{i=1}^n \overline{F}_1(t_i)\overline{F}_2(t_i)\right) \prod_{i|s_i=\{1\}} h_1(t_i) \prod_{i|s_i=\{2\}} h_2(t_i) \prod_{i|s_i=\{1,2\}} \{h_1(t_i) + h_2(t_i)\}. \tag{6}$$

So the Gibbs sampling for the exponential model is similar to the full model except z_i is generated by $B(1, \lambda_1/(\lambda_1+\lambda_2))$, and instead of θ_1 and θ_2, we generate a single $\theta \sim Beta(\alpha_1 + n_1 + n_2, \beta_1 + n_{12})$. The Gibbs sampling for the Weibull model is the same as the full model except $\theta_1 = \theta_2 = \theta$ and we generate $\theta \sim Beta(\alpha_1 + n_1 + n_2, \beta_1 + n_{12})$ to replace the generation of θ_1 and θ_2.

2.4 EM Algorithm for the Reduced Model

The maximization for θ and other parameters can be done separately, where $\hat{\theta} = (\alpha_1 + n_1 + n_2 - 1)/(\alpha_1 + \beta_1 + n_1 + n_2 + n_{12} - 2)$ with no iteration. For other parameters we follow the same procedure as the EM algorithm for the full model in Subsection 2.2 with $\theta_1 = \theta_2$.

3. Monotone Masking Probability Functions

In this section, we make the assumption that the masking probabilities are functions of time. We make monotone assumptions for the following masking probabilities. That is, we assume $P(S_i = \{1\}|t_i, K_i = 1)$ and $P(S_i = \{2\}|t_i, K_i = 2)$ are decreasing functions in t, because aging could make it more difficult to distinguish among different causes of failures. We will explore the exponential distribution assumptions for the masking probabilities, where $P(S_i = \{1\}|t_i, K_i = 1) = \exp(-\theta_1 t_i)$ and $P(S_i = \{2\}|t_i, K_i = 2) = \exp(-\theta_2 t_i)$. Other assumptions such as both gamma survival functions, both Weibull survival functions, or mixing them (one exponential, one Weibull, etc.) can be implemented in a similar fashion. The case for the masking probabilities to be increasing functions of time supported by Dinse (1986) is treated in Kuo and Yang (1998). In the followings, we will only consider the exponential component lifetimes.

3.1 Gibbs Sampling for the Full Monotone Function Model with Exponential Component Lifetimes

Let us observe that the full likelihood in this case is

$$\begin{aligned} L_F &= \left(\prod_{i|s_i=\{1\}} \lambda_1 \exp(-\theta_1 t_i)\right)\left(\prod_{i|s_i=\{2\}} \lambda_2 \exp(-\theta_2 t_i)\right) \\ &\quad \times \left(\prod_{i|s_i=\{1,2\}} [\lambda_1(1-\exp(-\theta_1 t_i)) + \lambda_2(1-\exp(-\theta_2 t_i))]\right)\left(\exp[-(\lambda_1+\lambda_2)(\sum_{i=1}^n t_i)]\right). \end{aligned} \tag{7}$$

On developing the Gibbs sampler, the data augmentation idea that uses the latent variable z_i to convert the sum in the third product (where $s_i = \{1,2\}$) in (7) into a product (Subsection 2.1) moves us in the right direction. However, due to the truncated probabilities in this new latent variable generation, we need to introduce two more classes of latent variables. For i with $s_i = \{1,2\}$ and $z_i = 1$, let us consider the exponentially distributed random variate u_i with mean $1/\theta_1$. Then observe $P(u_i < t_i) = 1 - \exp(-\theta_1 t_i)$. This expression is in the likelihood. Therefore, we generate a truncated random variate u_i truncated at t_i from an exponential distribution with mean $1/\theta_1$ for each i with $s_i = \{1,2\}$ and $z_i = 1$. Similarly, we generate a truncated random variate v_i truncated at t_i from an exponential distribution with mean $1/\theta_2$ for i with $s_i = \{1,2\}$ and $z_i = 0$. Let us assume the prior has the following independent components $\lambda_i \sim G(a_i, b_i)$, $\theta_i \sim G(a'_i, b'_i)$ for $i = 1, 2$. Observe the conditional density of $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$ given the classes of latent variables $\mathbf{z}, \mathbf{u}$, and $\mathbf{v}$ can be written as

$$\begin{aligned} f(\boldsymbol{\theta}, \boldsymbol{\lambda}|\mathbf{z}, \mathbf{u}, \mathbf{v}, \mathbf{d}) &\propto \left(\prod_{i|s_i=\{1\}} \lambda_1 \exp(-\theta_1 t_i)\right)\left(\prod_{i|s_i=\{2\}} \lambda_2 \exp(-\theta_2 t_i)\right) \\ &\quad \times \left(\prod_{i|s_i=\{1,2\}} [\lambda_1\theta_1 \exp(-\theta_1 u_i)]^{z_i}[\lambda_2\theta_2 \exp(-\theta_2 v_i)]^{(1-z_i)}\right) \exp[-(\lambda_1+\lambda_2)\sum_{i=1}^n t_i]\pi(\boldsymbol{\theta}, \boldsymbol{\lambda}). \end{aligned}$$

Now we summarize the Gibbs sampler steps:

Step a (data augmentation step): Given $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$, we generate three classes of random variables $\mathbf{z}, \mathbf{u}$, and $\mathbf{v}$. That is, we generate $z_i \sim B(1, \lambda_1(1-\exp(-\theta_1 t_i))/[\lambda_1(1-\exp(-\theta_1 t_i)) + \lambda_2(1-\exp(-\theta_2 t_i))])$, independently, for i with $s_i = \{1,2\}$. Then compute $z_+ = \sum_{i|s_i=\{1,2\}} z_i$. Moreover, we generate u_i independently from an exponential distribution with mean $1/\theta_1$ with truncation at t_i for i with $s_i = \{1,2\}$ and $z_i = 1$. Similarly, we generate v_i independently from an exponential random variable with mean $1/\theta_2$ with truncation at t_i for i with $s_i = \{1,2\}$ and $z_i = 0$. Observe that the truncated random variables can be generated by setting $u_i = -\ln[1 - w_i(1-\exp(-\theta_1 t_i))]/\theta_1$ for i with $s_i = \{1,2\}$ and $z_i = 1$, and setting $v_i = -\ln[1 - w_i(1 - \exp(-\theta_2 t_i))]/\theta_2$ for i with $s_i = \{1,2\}$ and $z_i = 0$, where w_i are independent uniform $U(0,1)$ variates.

Step b (parameter updating step): given $\mathbf{z}$, $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{d}$, we update the parameters independently by

$$\begin{aligned}
\theta_1|\mathbf{z},\mathbf{u},\mathbf{v},\mathbf{d} &\sim G\left(a_1' + z_+,\ b_1' + \sum_{i|s_i=\{1\}} t_i + \sum_{i|s_i=\{1,2\}} u_i z_i\right);\\
\theta_2|\mathbf{z},\mathbf{u},\mathbf{v},\mathbf{d} &\sim G\left(a_2' + n_{12} - z_+,\ b_2' + \sum_{i|s_i=\{2\}} t_i + \sum_{i|s_i=\{1,2\}} v_i(1-z_i)\right);\\
\lambda_1|\mathbf{z},\mathbf{u},\mathbf{v},\mathbf{d} &\sim G(a_1 + n_1 + z_+,\ b_1 + \sum_{i=1}^{n} t_i);\\
\lambda_2|\mathbf{z},\mathbf{u},\mathbf{v},\mathbf{d} &\sim G(a_2 + n_2 + n_{12} - z_+,\ b_2 + \sum_{i=1}^{n} t_i).
\end{aligned}$$

3.2 Gibbs Sampling for the Reduced Model

To apply the equiprobable assumption in the masking probability, we need to assume $\exp(-\theta_1 t_i) = \exp(-\theta_2 t_i)$ for all t_i. Therefore, we need to assume $\theta_1 = \theta_2 = \theta$ in the specification of the masking probability functions. We also assume θ has prior $G(a', b')$ independent of other prior parameters. Then the Gibbs sampler is similar to steps a and b in subsection 3.1 except θ replacing θ_1 and θ_2 for all the steps in step a; and

$$\theta|\mathbf{z},\mathbf{u},\mathbf{v},\mathbf{d} \sim G\left(a' + n_{12},\ b' + \sum_{i|s_i=\{1\}} t_i + \sum_{i|s_i=\{2\}} t_i + \sum_{i|s_i=\{1,2\}} u_i z_i + v_i(1-z_i)\right)$$

replacing the first two equations in step b.

4. Predictive Power

So far, we have presented several models for the masked system life times: full model and reduced model, each with exponential and Weibull component lifetimes in Section 2. Which model is most desirable? We use a predictive approach based on cross-validated data to address the model selection issue.

On model selection, we compare the full and the reduced models in terms of their predictive powers using cross-validated data. Let $(\mathbf{t}_{-i}, \mathbf{s}_{-i})$ denote the vector of ordered pairs of the entire data set with the t_i and s_i deleted. Let $\boldsymbol{\omega}$ denote the set of unknown parameters that includes $\boldsymbol{\theta}$,$\boldsymbol{\lambda}$, and $\boldsymbol{\rho}$. Let $\overline{\theta_1} = 1 - \theta_1$ and

$\overline{\theta_2} = 1 - \theta_2$. Let us define the conditional predictive ordinate (CPO) for the i^{th} data point for the full model to be

$$
\begin{aligned}
p(t_i, s_i|\mathbf{t}_{-i}, \mathbf{s}_{-i}) &= \int p(t_i, s_i|\boldsymbol{\omega})\pi(\boldsymbol{\omega}|\mathbf{t}_{-i}, \mathbf{s}_{-i})d\boldsymbol{\omega} \qquad (8)\\
&= \begin{cases} \int \theta_1 f_1(t_i)\overline{F}_2(t_i)\pi(\boldsymbol{\omega}|\mathbf{t}_{-i}, \mathbf{s}_{-i})d\boldsymbol{\omega}, & \text{for } i \mid s_i = \{1\} \\ \int \theta_2 f_2(t_i)\overline{F}_1(t_i)\pi(\boldsymbol{\omega}|\mathbf{t}_{-i}, \mathbf{s}_{-i})d\boldsymbol{\omega} & \text{for } i \mid s_i = \{2\} \\ \int [\overline{\theta_1} f_1(t_i)\overline{F}_2(t_i) + \overline{\theta_2} f_2(t_i)\overline{F}_1(t_i)]\pi(\boldsymbol{\omega}|\mathbf{t}_{-i}, \mathbf{s}_{-i})d\boldsymbol{\omega} & \text{for } i \mid s_i = \{1, 2\}. \end{cases}
\end{aligned}
$$

The CPO for the reduced model can be constructed from (6). Therefore, it is similar to (8) with both θ_1 and θ_2 replaced by θ.

Then we define the pseudo-marginal predictive likelihood (PL) (Geisser and Eddy, 1979) to be

$$PL = \prod_{i=1}^{n} p(t_i, s_i|\mathbf{t}_{-i}, \mathbf{s}_{-i}).$$

The pseudo-Bayes factor criterion is to select the model with the largest PL.

For models with monotone masking probability functions, we need to modify $p(t_i, s_i|\boldsymbol{\omega})$ in (8), where the classification probabilities θ_1 and θ_2 are replaced by $\exp(-\theta_1 t_i)$ and $\exp(-\theta_2 t_i)$ respectively. For the reduced model, we replace both θ_1 and θ_2 by $\exp(-\theta t_i)$.

5. Simulation Study

We consider a series system with two exponential components, where $f_j(t) = \lambda_j \exp(-\lambda_j t)$, $j = 1, 2$, and $\lambda_1 = 0.01$ and $\lambda_2 = 0.005$. For each system, we simulate a pair of failure times from f_1 and f_2 respectively. Then, we take the minimum of the two failure times as the system's life time and note its label, the cause of the system's failure. If the cause is from the first (second) component, we randomly mask 10 % (20 %) of the observations. That is, the unmasked probabilities are $\theta_1 = 0.9$ and $\theta_2 = 0.8$. The data in Table 1 represent the failure time and cause of failure for a random sample of $n = 50$ systems.

{Insert Table 1 around here.}

We apply our analysis in Sections 2 and 4 to this simulated data set. We fit the data with the full model in (2) and the reduced model in (6) with exponentially distributed component lifetimes. Table 2 summarizes the priors, the log of the pseudo-marginal predictive likelihood, the posterior means, the EM estimates, and the posterior estimates of the diagnostic probability of the cause 1 for each of the full and reduced models. We choose uniform priors for the unmasked probabilities. We monitor the convergence of the Gibbs sampler using the Gelman and Rubin(1992) method that uses the analysis of variance technique to determine whether or not further iterations are needed. We found 3,000 iterations to be large enough for the priors being considered. All the following numerical results are obtained with 3,000 iterations and 50 replications in the Gibbs sampler. The table shows that the log of the pseudo-marginal predictive likelihood in the full model is larger than that in the reduced model as expected. The table also shows that the posterior means using the MCMC algorithm and the EM estimates are comparable. Starting at $\theta_1^0 = 0.5, \theta_2^0 = 0.5, \lambda_1^0 = 0.005, \lambda_2^0 = 0.005$, the EM algorithm converges after the 6^{th} iteration when all the iteration error bounds are met, where each error bound for each parameter is set at the $E-8$ level. The diagnostic probability of the first cause conditional on $s_i = \{1,2\}$ and t_i for $i = 1, \ldots, n_{12}$ is: $p(k_i = 1|t_i, s_i = \{1,2\}) = (1-\theta_1)\lambda_1/((1-\theta_1)\lambda_1 + (1-\theta_2)\lambda_2)$ for the full exponential model and $p(k_i = 1|t_i, s_i = \{1,2\}) = \lambda_1/(\lambda_1 + \lambda_2)$ for the reduced exponential model. Note these diagnostic probabilities are the same independent of t_i. Table 2 also lists the posterior means for these two probabilities evaluated from the Gibbs sample.

{Insert Table 2 around here.}

Figure 1 compares the CPO for each model. The CPOs for the reduced exponential model and the full exponential model are plotted versus failure times. Figure 1 shows that the full model has more predictive power than the reduced model. It also supports the numerical results in Table 1. Figure 2 graphs the posterior densities of λ_1 and λ_2 respectively for the full exponential model based on the last half of the iterations of all replications from the Gibbs sampler.

{Insert Fig. 1 and Fig. 2 around here.}

6. Real Data Study

Table 3 lists the death time and nonrenal vascular disease (NRVD) status for each of 58 female mice given in Dinse (1986). Time to death was recorded in days and a status at death with respect to the NRVD disease was classified as incidental, unknown, or fatal. In Dinse, the status was originally classified as absent, incidental, unknown, or fatal. In this paper, according to the role of the disease, we can combine both the absent and the incidental status into one called the incidental status. Let s_i indicate whether or not the i^{th} death is due to the disease. We let $s_i = \{2\}$ if the disease is fatal and let $s_i = \{1\}$ if the disease is incidental; and we let $s_i = \{1,2\}$ when the role of the disease in causing death is unknown.

We have presented four models in Section 2: the full exponential model, the reduced exponential model, the full Weibull model, and the reduced Weibull model. We select the best model by comparing the predictive powers given in Section 4. In our analysis, we have implicitly assumed that death due to the disease and due to incidental status are independent. Table 4 summarizes the prior, the posterior mean, and the log of the pseudo-marginal predictive power for each of the four models. We consider diffuse proper priors so that the numerical results are more focused on the likelihood. All the numerical results here are obtained with 5,000 iterations and 50 replications in the Gibbs sampler. The EM estimates are quite similar to the posterior means, so we omit these estimates in Table 4. Table 4 shows that the predictive power of the full Weibull model is better than that of each of the other three models. Therefore, the full Weibull model is chosen to be the best among the four models according to the pseudo-Bayes factor.

{Insert Tables 3 and 4 around here.}

Acknowledgement

The authors would like to thank the referee for many constructive comments.

References

[1] Aboul-Séoud, M. H., and Usher, J. S. (1996), The effect of modular system structures on component reliability estimation, 5^{th} *Industrial Engineering Research Conference Proceedings*, 464-467.

[2] Dempster, A., Laird, N and D. Rubin. (1977), Maximum likelihood from incomplete data via the EM algorithm, *J. Roy. Statist. Soc. Ser.* B **39**, 1-38.

[3] Dinse, G. (1986), Nonparametric prevalence and mortality estimators for animal experiments with incomplete cause-of-death data, *J. Amer. Statist. Assoc.,* **81**, 328-336.

[4] Doganaksoy, N. (1991), Interval estimation from censored & masked system-failure data. *IEEE Trans. Reliab.* **40**, 280-286.

[5] Flehinger, B. J., Reiser, B., and Yashchin, E. (1996), Inference about defects in the presence of masking, *Technometrics*, **38**, 247-255.

[6] Flehinger, B. J., Reiser, B., and Yashchin, E. (1998), Survival with competing risks and masked causes of failures, *Biometrika*, **85**, 151-164.

[7] Gastaldi, T. (1994), Improved maximum likelihood estimation for component reliabilities with Miyakawa-Usher-Hodgson-Guess' estimators under censored search for the cause of failure. *Statist. Probab. Lett.* **19**, 5-18.

[8] Geisser, S., and Eddy, W. (1979), A predictive approach to model selection, *J. Amer. Statist. Assoc.* **74**, 153-160.

[9] Gelfand, A., and Smith, A.F.M. (1990), Sampling based approaches to calculating marginal densities, *J. Amer. Statist. Assoc.* **85**, 398-409.

[10] Geman, S., and Geman, D. (1984), Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Trans. Pattern Anal. Machine Intel.* **6**, 721-741.

[11] Guess, F.M., Usher, J.S. and Hodgson, T.J. (1991), Estimating system and component reliabilities under partial information on cause of failure, *J. Statist. Plann. Inference,* **29**, 75-85.

[12] Guttman, I., Lin, D.K.J., Reiser, B., and Usher, J.S. (1995), Dependent masking and system life data analysis: Bayesian inference for two-component systems, *Lifetime Data Anal.* 87-100.

[13] Kuo, L., and Yang, T. (1995), Bayesian computation for software reliability, *J. Comput. Graph. Statist.* **4**, 1-18.

[14] Kuo, L., and Yang, T. (1996), Bayesian computation for nonhomogeneous Poisson processes in software reliability, *J. Amer. Statist. Assoc.* **91**, 763-773.

[15] Kuo, L., and Yang, T. (1998), Bayesian reliability modeling for masked system lifetime data, Technical Report tr9810, Statistics Department, University of Connecticut.

[16] Lin, D. K. J. and Guess, F. M. (1994), System life data analysis with dependent partial knowledge on the exact cause of system failure, *Microelectronics Reliab.* **34**, 535-544.

[17] Lin, D. K. J., Usher, J. S., and Guess, F. M. (1993), Exact maximum likelihood estimation using masked system data, *IEEE Trans. Reliab.* **42**, 631-635.

[18] Lin, D. K. J., Usher, J. S., and Guess, F. M. (1996), Bayes estimation of component-reliability from masked system-life data, *IEEE Trans. Reliab.* 45, 233-237.

[19] Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., and Teller, E. (1953), Equations of state calculations by fast computing machines, *J. Chem. Physics,* **21**, 1087-1091.

[20] Miyakawa, M. (1984), Analysis of incomplete data in a competing risks model, *IEEE Trans. Reliab.* **33**, 293-296.

[21] Mukhopadhyay, C., and Basu, A.P. (1997), Bayesian analysis of incomplete time and cause of failure data, *J. Statist. Plann. Inference*, **59**, 79-100.

[22] Reiser, B., Flehinger, B. J., and Conn, A. R. (1996), Estimating component-defect probability form masked system success/failure data, *IEEE Trans. Reliab.* **45**, 238-243.

[23] Reiser, B., Guttman, I., Lin, D.K.J., Guess, F.M., and Usher, J.S. (1995), Bayesian inference for masked system lifetime data, *Applied Statistics,* **44**, 79-90.

[24] Sinha, B. and Zellner, A. (1990), A note on the prior distributions of Weibull parameters, *SCIMA,* **19**, 5-13.

[25] Sinha, S. (1986), *Reliability and Life Testing.* Wiley, New York.

[26] Tanner, M.A., and Wong, W.H. (1987), The calculation of posterior distributions by data augmentation, *J. Amer. Statist. Assoc.* **82**, 528-550.

[27] Usher, J. (1996), Weibull component reliability-prediction in the presence of masked data, *IEEE Trans. Reliab.* **45**, 229-232.

[28] Usher, J. and Hodgson, T. (1988), Maximum likelihood analysis of component reliability using masked system life-test data, *IEEE Trans. Reliab.* **37**, 550-555.

[29] Yang, T., and Kuo, L. (1999), Bayesian computation for the superposition of nonhomogeneous Poisson processes, *Canad. J. Statist.* In Press.

Table 1: Simulated Data

Set of Labels	System's Failure Time
$s_i = \{1\}$	3.4, 10.9, 71.5, 56.9, 41.6, 41.6, 38.6, 4.6, 8.5. 27.8, 2.9, 13.3, 37.4, 57.5, 168.1, 123.9, 136.3, 6.5, 21.4, 135.4, 93.1, 14.3, 10.8, 26.1, 95.3, 68.7, 22.3, 13.3, 73.4, 35.2, 28.0, 21.1
$s_i = \{2\}$	2.5, 212.2, 147.3, 63.6, 45.6, 4.2, 123.1, 10.5, 2.6, 27.4
$s_i = \{1, 2\}$	68.6, 2.3, 5.4, 8.8, 63.7, 325.2, 216.4, 132.0

Table 2: Priors, Posterior Means, and Predictive Likelihood

Model	Full Exponential	Reduced Exponential
	$\theta_i \sim$ *Beta* $(1,1)$ for $i = 1,2$	$\theta \sim$ *Beta* $(1,1)$
Prior	$\lambda_i \sim \Gamma(1, 0.001)$ for $i = 1,2$	$\lambda_i \sim \Gamma(1, 0.001)$ for $i = 1,2$
log Predictive Likelihood	-238.52	-240.47
Posterior Means	$\hat{\theta}_1 = 0.866, \hat{\theta}_2 = 0.718$	$\hat{\theta} = 0.827$
Using Gibbs	$\hat{\lambda}_1 = 0.0125, \hat{\lambda}_2 = 0.0050$	$\hat{\lambda}_1 = 0.0131, \hat{\lambda}_2 = 0.0044$
Estimates	$\hat{\theta}_1 = 0.889, \hat{\theta}_2 = 0.714$	$\hat{\theta} = 0.840$
Using EM	$\hat{\lambda}_1 = 0.0121, \hat{\lambda}_2 = 0.00471$	$\hat{\lambda}_1 = 0.0128, \hat{\lambda}_2 = 0.00401$
$\hat{p}(k_i = 1 \| t_i, s_i = \{1, 2\})$	0.53	0.75

Table 3: Modified Dinse Data

NRVD Status	Set of Labels	Time to Death
Incidental	$s_i = \{1\}$	231, 444, 468, 473, 527, 550, 559, 593, 595, 596, 600, 603, 610, 650, 655, 660, 715, 720, 752, 765, 783, 785, 794, 811, 832, 838, 856, 859, 870, 883, 891, 896, 897, 904, 931, 952, 975, 978, 991, 998, 1005, 1023, 1026, 1053
Fatal	$s_i = \{2\}$	500, 591, 713, 751, 778, 784, 786, 796
Unknown	$s_i = \{1,2\}$	593, 735, 816, 848, 850, 1046

Table 4: Priors, Posterior Means, and Predictive Likelihoods for the Modified Dinse Data

Model	priors	Posterior Means	log Predictive Likelihood
Full Exponential	$\theta_i \sim$ *Beta* $(1,1)$ for $i = 1,2$ $\lambda_i \sim \Gamma(1, .000001)$ for $i = 1,2$	$\hat{\theta}_1 = 0.914, \hat{\theta}_2 = 0.727$ $\hat{\lambda}_1 = 0.00109, \hat{\lambda}_2 = 0.000264$	-484.7
Reduced Exponential	$\theta \sim$ *Beta* $(1,1)$ $\lambda_i \sim \Gamma(1, .000001)$ for $i = 1,2$	$\hat{\theta} = 0.883$ $\hat{\lambda}_1 = 0.00113, \hat{\lambda}_2 = 0.000225$	-487.0
Full Weibull	$\theta_i \sim$ *Beta* $(1,1)$ for $i = 1,2$ $\lambda_i \sim \Gamma(1, .000001)$ for $i = 1,2$ $\rho_i \sim U(0,4)$ for $i = 1,2$	$\hat{\theta}_1 = 0.876, \hat{\theta}_2 = 0.853$ $\hat{\lambda}_1 = 0.113E-08, \hat{\lambda}_2 = 0.000576$ $\hat{\rho}_1 = 3.34, \hat{\rho}_2 = 1.21$	-474.6
Reduced Weibull	$\theta \sim$ *Beta* $(1,1)$ $\lambda_i \sim \Gamma(1, .000001)$ for $i = 1,2$ $\rho_i \sim U(0,4)$ for $i = 1,2$	$\hat{\theta} = 0.883$ $\hat{\lambda}_1 = 0.555E-08, \hat{\lambda}_2 = 0.000395$ $\hat{\rho}_1 = 3.27, \hat{\rho}_2 = 1.28$	-475.4

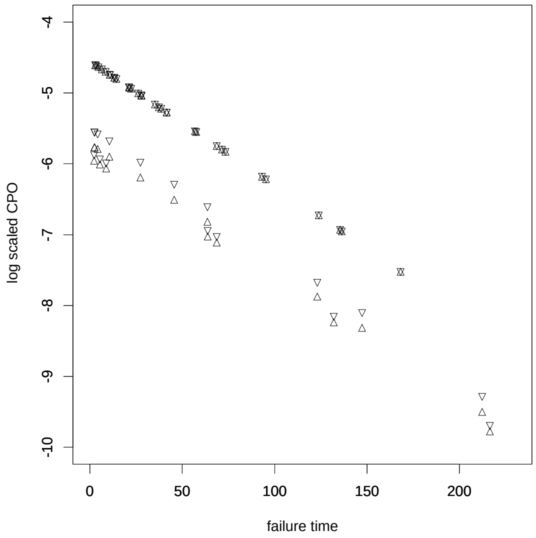

Figure 1: CPO Comparison for the Simulated Data. (∇ denotes the full model and $\triangle$ denotes the reduced model.)

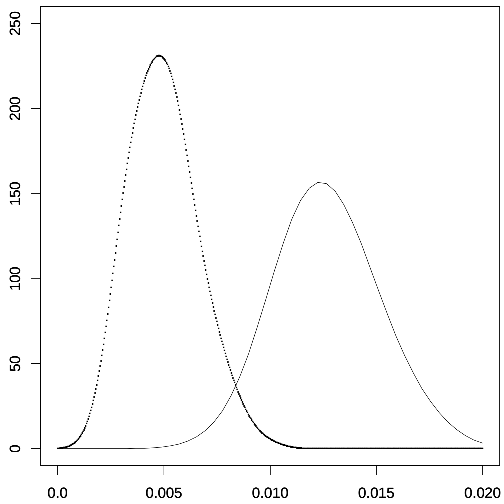

Figure 2: Density Estimates for λ_1 and λ_2 (the solid line denotes the density estimate for λ_1 and the dotted line denotes the density estimate for λ_2)

.