Bayesian Analysis of Masked Series System Lifetime Data with log-Normal Component Lives

Chiranjit Mukhopadhyay
Department of Statistics
Case Western Reserve University
Cleveland, OH 44106, USA
e-mail: chiranjit.mukhopadhyay@case.edu

Sanjib Basu
Division of Statistics
Northern Illinois University
Dekalb, IL 60115, USA
e-mail: basu@math.niu.edu

Abstract

The problem of analyzing series system lifetime data with masked or partial information on cause of failure is recent, compared to that of the standard competing risks model. Here it is assumed that the components of a system have independent log-Normal life distributions and a Bayesian analysis is presented for such models. The masking probabilities are not subjected to the symmetry assumption and independent Dirichlet priors are used to marginalize these nuisance parameters. Independent Normal-Gamma priors are employed for the component lifetime parameters. The Gibbs sampling scheme developed for the required analysis can also be used to provide a Bayesian analysis of data arising from the conventional competing risks model of independent log-Normals, which interestingly has so far remained by and large neglected in the literature. The developed methodology is deployed to analyze a masked lifetime data of PS/2 computer systems.

Key Words: Censoring; Competing Risks; Dirichlet Distribution; Gibbs Sampling; Masking Probabilities; Normal-Gamma Prior; Reliability.

3On leave from Department of Management Studies, Indian Institute of Science, Bangalore 560 012, INDIA, e-mail: cm@stats.mgmt.iisc.ernet.in
1 Introduction

Consider the masked series system lifetime data in Reiser et al. (1995). 682 PS/2 computer systems are put on life-test, 8 of which failed, and the lifetime of the remaining 674 are censored. A system can fail due to malfunction of either the 1) Motherboard, or the 2) Disc Drives, or the 3) Power Supply. The overall system fails at the first occurrence of failure of any one of these three components, but not from simultaneous failures of two or more of these components.

Out of the 8 systems which failed, the exact cause of failure of 3 systems are not known. Their causes of failure are known only to belong to a subset (of cardinality $> 1$) of $\{1,2,3\}$. The data and details of the reasons behind such observations are provided in Reiser et al. (1995). When one has such partial information on the cause or the component responsible for a system failure, the phenomenon has come to be known as masking in the literature. A brief review of the masking literature is available in §1.3 of Mukhopadhyay (2004).

Reiser et al. (1995) analyze the data set from a Bayesian approach, assuming that the lifetimes of the three components are mutually independent and have exponential distributions. They employ the standard independent non-informative priors for the model parameters. Basu et al. (1999) provide a Bayesian analysis of the same data set under the independent Weibull model with non-informative priors on the scale parameters and discrete priors on the shape parameters. Basu et al. (2003) further provide a Bayesian analysis of the data set under the assumption of the component lifetimes having either extreme value (Weibull) or log-Normal distributions. They employ general log-concave priors on the parameters for their analysis. Mukhopadhyay (2004) provides a maximum likelihood analysis of the same data set with the independent Weibull component lifetime model without the so-called symmetry assumption.

The Bayesian analyses mentioned above either explicitly or tacitly subject the masking probabilities to the unrealistic symmetry assumption, which may be described as follows. A masked datum on a system failure consists of the log-time at which the failure has occurred, say $T$, accompanied by a Minimum Random Subset (MRS) (vide. Guess et al. (1991)) containing the masked cause of the system failure, say $M$. Development of a probabilistic model for this observed datum leads to the consideration of a vector of masking probabilities (see §1.2 of Mukhopadhyay (2004) for example). The masking probabilities are nothing but the conditional probability distributions of $M$ given $T$ and the exact cause of a system failure, say $K$. $K$ is a discrete random variable which assumes values in $\{1, \ldots, J\}$, where $J$ denotes the number of components in a system. The symmetry assumption imposes the restriction that $\forall A \subseteq \{1, \ldots, J\}$, the masking probability say, $\mu_{t,j}(A) = \text{Prob}(M = A | T = t, K = j)$ is a constant free of both $j$ and $t$, $\forall j \in \{1, \ldots, J\}$ and $\forall t > 0$.

The symmetry assumption enables one to factor the likelihood function into product of two components - one consisting of only the masking probabilities and the other comprising solely the lifetime parameters. Consequently inference about the lifetime parameters, the primary parameters of interest, can proceed by completely ignoring the nuisance parameters viz. the masking probabilities. However this is done purely for mathematical convenience without much practical justification. Even if there is one, there must be some way of testing such a strong assumption. Furthermore, a proper consideration of the probabilistic model for this problem cannot avoid the additional random entity, the MRS $M$, and thus the masking probabilities (vide. Appendix A of Mukhopadhyay (2004)).
Relaxation of the symmetry assumption has been considered by Dinse (1986), Lin and Guess (1994), Guttman et.al. (1995), Flehinger et.al. (1996, 1998, 2002), Kuo and Yang (2000) and Mukhopadhyay (2004). All of these are frequentist work with the exception of the ones by Guttman et.al. (1995) and Kuo and Yang (2000). Guttman et.al. (1995) make the assumption of proportional masking probabilities and provide a Bayesian analysis for two independent exponentials. Kuo and Yang (2000) provide a Bayesian analysis for $J = 2$ independent exponentials and Weibulls. To the authors’ knowledge, no Bayesian analysis is available in the literature without the symmetry condition for log-Normally distributed component lifetimes.

This article provides a Bayesian analysis of masked series system lifetime data without the symmetry assumption when the component lifetimes have independent log-Normally distributed distributions. Bayesian analysis of general location-scale models developed in Basu et.al. (2003) includes the model considered in this article as a special case. But Basu et.al. (2003) assume the crucial symmetry condition, which considerably simplifies the analysis, which this article does not. The Bayesian methodology, developed in this article, also differs from all other Bayesian works in this problem such as Berger and Sun (1993), Mukhopadhyay and Basu (1993, 1997), Basu et.al. (1999, 2003) and Kuo and Yang (2000) in its data imputation step. The data augmentation technique proposed in this article will also facilitate Bayesian analysis of the conventional competing risks models with dependent component lifetimes like multivariate log-Normal, or when there is no so-called closed-form expression for the component survival functions, such as the Normal/log-Normal model considered in this article or other models like Gamma or generalized Gamma (vide. §5.3 of Lawless (1982)). Interestingly in such cases, which though are pristine, no Bayesian analysis is still available in the literature for the conventional competing risks models. This article would provide a (may be an alternative) way of carrying out a Bayesian analysis of the conventional competing risks model in such cases.

The article is organized as follows. Section 2 first introduces the model for masked system life data and the likelihood function. §2.1 then details the auxiliary or latent variables (“pseudo data”) that are augmented to the observed data in this case. §2.2 admits the assumed prior distributions. §2.3 provides the required conditional distributions for the imputation of the “pseudo data”, and the conditional posteriors of the model parameters for the Gibbs sampling. A reference Bayesian analysis of the data set in Reiser et.al. (1995), under the assumption that the log-lifetimes of the three constituent components of the PS/2 system are independent and Normally distributed, is presented in section 3. Section 4 concludes the article by summarizing its contributions and directions of possible future research.

2 Bayesian Analysis

Let $X_j$ denote the log-lifetime of the $j$-th component of a series system $j = 1, \ldots, J$. It is assumed that the lifetimes of the $J$ components have independent log-Normal distributions i.e. $X_j \sim N(\theta_j, \sigma_j^2)$ and $X_1, \ldots, X_J$ are mutually independent. Let $X = (X_1, \ldots,$

2In some of these cases, like for example the independent log-Normal model considered in this article, a Bayesian analysis of the conventional competing risks model may be carried out using the recently developed method by Basu et.al. (2003) though.
\(X_j\), \(\tau_j=1/\sigma_j^2\), \(\theta=(\theta_1, \ldots, \theta_J)\) and \(\tau=(\tau_1, \ldots, \tau_J)\). Then by the definition of a series system the overall system log-lifetime \(T=\text{Minimum}\{X_1, \ldots, X_J\}\). Thus \(T\) has the survival function \(F_T(t|\theta, \tau) = \prod_{j=1}^{J} \Phi(\sqrt{\tau_j}(t - \theta_j))\) and probability density function \(f_T(t|\theta, \tau) = \sum_{j=1}^{J} \{\sqrt{\tau_j} \phi(\sqrt{\tau_j}(t - \theta_j)) \prod_{k \neq j} \Phi(\sqrt{\tau_k}(t - \theta_k))\}\), where \(\phi(\cdot)\) and \(\Phi(\cdot)\) respectively denote the standard Normal density and survival functions. Recall that the random variable \(t\) denotes the component among \(\{1, \ldots, J\}\) which is the cause of a system failure. Then the joint density of \((K, T)\) is given by:

\[
p(k, t|\theta, \tau) = \sqrt{\tau_k} \phi(\sqrt{\tau_k}(t - \theta_k)) \prod_{j \neq k} \Phi(\sqrt{\tau_j}(t - \theta_j))
\]

For a failed system, although its log-failure time \(T\) is known somewhat accurately, the associated cause of failure \(K\) may be masked. In general, it is only known to belong to \(A \subseteq \{1, \ldots, J\}\). That is, the observation on a failed system with the observed log-failure time \(t\) and the observed MRS \(A\), is written as \((M=A, T=t)\). Note that \(M=A \Rightarrow K \in A\). Thus for writing the likelihood function, one needs to introduce \(\mu_{t,j}(A) = \text{Prob}(M=A|T=t, K=j)\).

Here it is assumed that \(\mu_{t,j}(A) = \mu_j(A) \forall A \subseteq \{1, \ldots, J\}, \forall j \in \{1, \ldots, J\}\) and \(\forall t > 0\), where \(\mu_j(A) = \text{Prob}(M=A|K=j)\). This assumption has also been made by Flehinger et.al. (1996, 1998, 2002), Mukhopadhyay (2004) and in parts of their articles by Dinse (1986) and Kuo and Yang (2000). Analogous to the Remarks in Mukhopadhyay (2004) here also it should be remarked that even without this assumption it is possible to carry out a Bayesian analysis of the lifetime parameters. But in that case for small samples the posterior analysis becomes extremely sensitive to the assumed prior on the \(\mu_{t,j}(A)\)’s. Thus to avoid such an overwhelming dependence on the prior, the nuisance parameter space of the masking probabilities is kept restricted to a finite dimension with the help of this assumption.

Now since \(\mu_j(A)\)’s are conditional probabilities, they are required to satisfy certain constraints which are as follows.

\[
\sum_{A \in A_j} \mu_j(A) = 1 \quad \text{with} \quad 0 \leq \mu_j(A) \leq 1 \quad \forall A \in A_j
\]
and
\[
\mu_j(A) = 0 \quad \forall A \in A_j^c = A - A_j
\]

where \(A_j=\{A \in A : j \in A\}\) for a fixed \(j \in \{1, \ldots, J\}\) and \(A\) is the collection of all \(2^J-1\) possible non-empty subsets of \(\{1, \ldots, J\}\). Let \(\mu_j = \{\mu_j(A)\}_{A \in A_j}\) and \(\mu = \{\mu_j\}_{j=1}^{J}\). It should be mentioned that here the interest lies in inference about the lifetime parameters \((\theta, \tau)\), and not in the nuisance parameter \(\mu\) of masking probabilities. The nuisance parameter \(\mu\) arises for a probabilistic model of the masking problem. However as it is evident from Proposition 1 in §2.3 below, whether one is interested in drawing inference about this nuisance parameter \(\mu\) or not, one has to generate these masking probabilities in the Gibbs sampling. Thus inference can also be made about the masking probabilities at the end of the Gibbs sampling if one wishes to do so.

Now suppose \(N\) \(J\)-component series systems are put on life-test and \(n\) of them failed. For \(i = 1, \ldots, n\) let \(T^i\) and \(t^i\) respectively denote the random variable and its observed value of the log-failure time of the \(i\)-th failed system. Similarly for \(r = 1, \ldots, N-n\) let \(T^{n+r}\) and \(t^{n+r}\) respectively denote the random variable representing the future log-failure time and
the observed log-censoring time of the \( r \)-th running system. That is \( N - n \) systems have not failed within their stipulated times and the \( r \)-th one of such systems is still running at time \( e^{(n+r)} \), i.e. \( T^{n+r} > t^{n+r} \). Finally let \( M^i \) and \( A^i \) respectively denote the MRS and the observed masking set of the \( i \)-th failure. Thus the observed data is given by,

\[
D = \{ M^1 = A^1, \ldots, M^n = A^n, T^1 = t^1, \ldots, T^n = t^n, T^{n+1} > t^{n+1}, \ldots, T^N > t^N \}
\]

Let \( A^i = (A^1, \ldots, A^n) \), \( t = (t^1, \ldots, t^n) \), and \( t^+ = (t^{n+1}, \ldots, t^N) \). Then the likelihood function of the observed data \( D \) is given by,

\[
L_D(\mu, \theta, \tau | A, t, t^+) \propto \prod_{i=1}^{n} \sum_{j \in A^i} \{ \mu_j(A^i)p(j, t^i|\theta, \tau) \} \prod_{r=1}^{N-n} \prod_{j=1}^{J} \Phi(\sqrt{\tau_j}(t^{n+r} - \theta_j)) \tag{3}
\]

where \( p(j, t^i|\theta, \tau) \)'s are as in (1).

### 2.1 Pseudo Data

In this problem, in the literature, the auxiliary variables that are imputed are \( K^1, \ldots, K^n \), where \( K^i \) is the possibly unobserved exact cause of the \( i \)-th failure. This imputation is essentially suggested for tackling the problem of masking, by reducing it down to a standard competing risks model. This is based on the expectation that a straightforward Bayesian analysis is available for the conventional competing risks model, which is typically the case if the survival functions of the \( X_j \)'s have closed form expressions. As in the case of distinct Weibulls for instance, censoring and similarly the \( J \) competing risks do not pose any additional problem for the generation of observations from the conditional marginals in the Gibbs sampling of the conventional competing risks model vide. Berger and Sun (1993), Basu et al. (1999, 2003) and Kuo and Yang (2000).

In the present independent distinct log-Normal case however, along with masking, censoring and thus the \( J \) competing risks pose additional problems in the Gibbs sampling, because of the nature of the survival function. Thus here instead of augmenting the observed data \( D \) with just \( K^1, \ldots, K^n \), the Bayesian analysis is facilitated by augmenting \( D \) with the log-lifetimes of all the constituent components of both the censored and uncensored systems. Note that although it is practically impossible to observe these component log-lifetimes, they nonetheless theoretically exist by the virtue of the assumed competing risks model. Also note that this information about the component lifetimes automatically determine the values of \( K^1, \ldots, K^n \). In this sense the proposed data augmentation technique is more general than the one suggested in the literature. It should be remarked that, as mentioned in §1, this type of data augmentation will also be useful for a Bayesian analysis of not only the masking problem, but also the standard competing risks model with other types of component life distributions as well, which are as yet unavailable in the literature.

Thus for \( i = 1, \ldots, n \) let \( X^i = (X^i_1, \ldots, X^i_J) \) denote the random vector of theoretical log-lifetimes of the \( J \) components of the \( i \)-th failed system. Note that all but one of these \( X^i_j \)'s are latent. That is it is neither possible nor can one observe all of these \( X^i_j \)'s. When the \( i \)-th system fails one only observes their minimum. Now with the convention that \( 0^0 = 1 \), the quantum of contribution to the likelihood of the data-point \((M^i = A^i, X^i = x^i)\) is
proportional to $\prod_{j=1}^{J} \mu_j(A_j)^{\#(x_j=t')} \tau_j^{1/2} e^{-\frac{1}{2} \tau_j (x_j^r - \theta_j)^2}$, where $I[.]$ is the indicator function and $x^i_j = (x^i_1, \ldots, x^i_J)$.

Similarly for the $r$-th running system, for which the system lifetime is censored, introduce the latent vector of component log-lifetimes $X^{n+r} = (X_1^{n+r}, \ldots, X_j^{n+r})$, $r=1, \ldots, N-n$. The quantum of contribution to the likelihood of this data-point $(X^{n+r} = x^{n+r})$ is proportional to $\prod_{j=1}^{J} \tau_j^{1/2} e^{-\frac{1}{2} \tau_j (x_j^{n+r} - \theta_j)^2}$, where $x^{n+r} = (x_1^{n+r}, \ldots, x_J^{n+r})$.

Note that the Laplacian non-informative prior on $x_j$ is $\frac{1}{\tau_j} e^{-\frac{1}{\tau_j} |x_j - \theta_j|}$.

Then the likelihood function of this completed data $(X^{n+r} = x^{n+r})$ is proportional to $\prod_{j=1}^{J} \tau_j^{1/2} e^{-\frac{1}{2} \tau_j (x_j^{n+r} - \theta_j)^2}$, where $x^{n+r} = (x_1^{n+r}, \ldots, x_J^{n+r})$.

Note that the $X^i_j$’s are mutually independent $\forall i=1, \ldots, N$. Thus the augmented data are $X^i = x^i$, $i=1, \ldots, n$ and $X^{n+r} = x^{n+r}$, $r=1, \ldots, N-n$. Let $X = \{X^1, \ldots, X^n, X^{n+1}, \ldots, X^N\}$ and the complete data

\[ C = \left\{ M^1 = A^1, \ldots, M^n = A^n, X^1 = x^1, \ldots, X^n = x^n, X^{n+1} = x^{n+1}, \ldots, X^N = x^N \right\}. \]

Then the likelihood function of this completed data $C$ is given by:

\[ L_C(\mu, \theta, \tau | A, X) \propto \prod_{j=1}^{J} \left\{ \prod_{A \in A_j} \mu_j(A)^{n_j(A)} \right\} \tau_j^{N/2} \exp \left( -\frac{N}{2} \tau_j \left\{ (\theta_j - \overline{x}_j)^2 + s_j^2 \right\} \right) \]

where $n_j(A) = \# \left\{ i \in \{1, \ldots, n\} : A^i = A \& x^i_j = t' \right\}$, $\overline{x}_j = \frac{1}{N} \sum_{i=1}^{N} x^i_j$ and $s_j^2 = \frac{1}{N} \sum_{i=1}^{N} (x^i_j - \overline{x}_j)^2$.

### 2.2 Prior Distributions

To keep computations tractable, we will work with independent conjugate family of priors for both $\mu$ and $(\theta, \tau)$ for the likelihood in (4). As is evident from (4) it then suffices to impose $J$ independent priors on $\mu_j$ and $(\theta_j, \tau_j)$ for $j=1, \ldots, J$.

We first address the problem of prior specification for $\mu_j$. Note that $\mu_j$ has to satisfy the constraints given in (2). That is $\mu_j$ must be such that $\{\mu_j(A)\} \forall A \in A_j$ is a probability distribution on $A_j$. Let $\rho = 2^{J-1}$, denote the number of sets in $A_j$. A prior on $\mu_j$ which satisfies the constraint is a $\rho$-parameter Dirichlet distribution with parameter vector $\alpha_j = \{\alpha_j(A)\} \forall A \in A_j$. Now since it is further assumed that the $\mu_j$’s are also independent apriori,

\[ \pi(\mu) \propto \prod_{j=1}^{J} \prod_{A \in A_j} \mu_j(A)^{\alpha_j(A)-1} \quad 0 \leq \mu_j(A) \leq 1, \quad \sum_{A \in A_j} \mu_j(A) = 1 \] (5)

Note that the Laplacian non-informative prior on $\mu_j$ can be modeled by assuming $\alpha_j(A) = 1$ $\forall A \in A_j$, while the Jeffreys’ and Haldane’s non-informative priors are modeled by assuming $\alpha_j(A) = 1/2$ and $\alpha_j(A) = 0$ $\forall A \in A_j$ respectively.

Now for $(\theta_j, \tau_j)$ we assume the standard conjugate Normal-Gamma prior for the mean and the precision. That is given $\tau_j$ it is assumed that $\theta_j | \tau_j \sim N(\eta_j, \frac{1}{\tau_j \psi_j})$ and $\tau_j \sim$ Gamma($\beta_j, \lambda_j$). Now since it is further assumed that $(\theta_1, \tau_1), \ldots, (\theta_J, \tau_J)$ are mutually independent *apriori*,

\[ \pi(\theta, \tau) \propto \prod_{j=1}^{J} \Gamma_j^{\beta_j/2} \exp \left( -\tau_j \left\{ \lambda_j + \frac{1}{2} \psi_j (\theta_j - \eta_j)^2 \right\} \right) \quad -\infty < \theta_j < \infty \quad \tau_j > 0 \] (6)
Jeffreys’ prior for \((\theta_j, \tau_j)\) is given by \(\frac{1}{\sqrt{\tau_j}}\), which can be modeled by choosing \(\beta_j=0\), \(\lambda_j=0\) and \(\psi_j=0\) in (6) \(\forall j\). An alternative location-scale invariant non-informative prior for \(N(\theta_j, \sigma_j^2)\) is given by \(\frac{1}{\sigma_j}\) (Berger (1985) p.88), which after transforming to the current parameterization yields \(\frac{1}{\tau_j}\) \(\forall j\), which can be modeled by choosing \(\beta_j= -\frac{1}{2}\), \(\lambda_j=0\) and \(\psi_j=0\) in (6) \(\forall j\).

2.3 Gibbs Sampling

The posterior features of \((\theta, \tau)\) are extracted via the ubiquitous Gibbs Sampling. In order to implement the Gibbs sampler, one needs the following conditional distributions (denoted generically by \(\pi(\cdot|\cdot)\) in all cases, where the subscript of \(\pi\) indicates the random variable (vector), whose conditional distribution is being sought):

1. \(\pi_X(x^i|t, A, t^+, \mathcal{X}^{(-i)}, \mu, \theta, \tau)\) where \(\mathcal{X}^{(-i)} = \{X^1, \ldots, X^{i-1}, X^{i+1}, \ldots, X^N\}\) for \(i = 1, \ldots, N\).

2. \(\pi\mu_j(\cdot|t, A, t^+, \mathcal{X}, \mu_{(-j)}, \theta, \tau)\) where \(\mu_{(-j)} = (\mu_1, \ldots, \mu_{j-1}, \mu_{j+1}, \ldots, \mu_J)\) for \(j = 1, \ldots, J\).

3. \(\pi\theta_j(\cdot|t, A, t^+, \mathcal{X}, \mu, \theta_{(-j)}, \tau)\) where \(\theta_{(-j)} = (\theta_1, \ldots, \theta_{j-1}, \theta_{j+1}, \ldots, \theta_J)\) for \(j = 1, \ldots, J\).

4. \(\pi\tau_j(\cdot|t, A, t^+, \mathcal{X}, \mu, \theta, \tau_{(-j)})\) where \(\tau_{(-j)}=(\tau_1, \ldots, \tau_{j-1}, \tau_{j+1}, \ldots, \tau_J)\) for \(j = 1, \ldots, J\).

Proposition 1 \(\forall i = 1, \ldots, n\)

\[
\pi_X(x^i|t, A, t^+, \mathcal{X}^{(-i)}, \mu, \theta, \tau) = \begin{cases} 
\sum_{j=1}^{J} \left\{ q_j^i(\mu, \theta, \tau, \tau_j) \prod_{k \neq j} \sqrt{\tau_k} \phi(\sqrt{\tau_k}(x_k - \theta_k)) / \Phi(\sqrt{\tau_k}(t_k - \theta_k)) \right\} & \text{if Minimum}\{x_1^i, \ldots, x_j^i\} = t^i \\
0 & \text{otherwise}
\end{cases}
\]

where \(q_j^i(\mu, \theta, \tau, \tau_j) = \begin{cases} \mu_j(A^i)p(j, t^i|\theta, \tau) / \sum_{k \in A^i} \mu_k(A^i)p(k, t^i|\theta, \tau) & \text{if } j \in A^i \\
0 & \text{otherwise}
\end{cases}\).

Proof: Let \(x^i\) be such that Minimum\{\(x_1^i, \ldots, x_j^i\}\} = t^i and Arg. Min. \{\(x_1^i, \ldots, x_j^i\}\} \in A^i 1, \ldots, J\) Then by mutually exclusive disjointification of the event \(M^i = A^i, T^i = t^i\) as \(\cup_{j \in A^i}(M^i = A^j, T^i = t^k, K^i = j)\), (1) and the definition of \(\mu_j(A^i)\),

\[
\pi_X(x^i|t, A, t^+, \mathcal{X}^{(-i)}, \mu, \theta, \tau) = \frac{1}{\sum_{j \in A^i} \mu_j(A^i)p(j, t^i|\theta, \tau)} \sum_{j \in A^i} \left\{ \mu_j(A^i) \sqrt{\tau_j} \phi(\sqrt{\tau_j}(x_j^i - \theta_j)) \prod_{k \neq j} \sqrt{\tau_k} \phi(\sqrt{\tau_k}(x_k^i - \theta_k)) \right\}
\]

\[
= \sum_{j=1}^J \left\{ q_j^i(\mu, \theta, \tau, \tau_j) \prod_{k \neq j} \sqrt{\tau_k} \phi(\sqrt{\tau_k}(x_k^i - \theta_k)) / \Phi(\sqrt{\tau_k}(t_k - \theta_k)) \right\} .
\]

The last equality follows from the definition of \(q_j^i(\mu, \theta, \tau)\) and (1). Q.E.D.
Note that \( q_j^i(\mu, \theta, \tau) = P(K^i = j|M^i = A^i, T^i = t^i, \mu, \theta, \tau) \), and 
\[
\frac{\sqrt{\tau_k \Phi(\sqrt{\tau_k (t^i_k - \theta_k)})}}{\Phi(\sqrt{\tau_k (t^i - \theta_k)})}
\]
is the conditional density of \((X_k|X_k > t^i, \theta, \tau)\). Thus the conditional density of \(X^i\) given \(t, A, t^+, X^{(-i)}, \mu, \theta, \tau\) derived in Proposition 1 above can be interpreted as a mixture of \(J\) densities, where each of these \(J\) densities is a product of \(J - 1\) truncated Normals, each truncated at \(t^i\), with mixing proportions \(q_j^i(\mu, \theta, \tau)\). Hence simulating observations from \(\pi_{X^i}(x^i|t, A, t^+, X^{(-i)}, \mu, \theta, \tau)\) is easily accomplished as follows. First generate a \(K^i\) taking values in \(\{1, \ldots, J\}\) with respective probabilities \(\{q_1^i(\mu, \theta, \tau), \ldots, q_J^i(\mu, \theta, \tau)\}\). Denote this generated value of \(K^i\) by \(j\). Fix \(x^i_j\) at \(t^i\), and for \(k \neq j\) keep generating observations from \(N(\theta_k, 1/\tau_k)\) until it is greater than \(t^i\), and thus yielding \(x^i_k\), independently of each other \(\forall k \neq j\). It is easy to verify that the joint density of \(X^i\), thus generated, is identical to the one provided in Proposition 1.

For the conventional competing risks model, in which the exact reasons of failures are known for all the failed systems, \(\forall i = 1, \ldots, n\), \(q_j^i(\mu, \theta, \tau) = \begin{cases} 1 & \text{if } j = K^i \\ 0 & \text{otherwise} \end{cases}\). In such a situation, the value of the \(K^i\)'s being exactly known, the first sub-step of imputing the \(K^i\)'s is not required, and one can directly proceed to simulate the values of the \(x^i\)'s via the rejection method outlined above.

Observations from \(\pi_{X^{n+r}}(x^{n+r}|t, A, t^+, X^{(-i)}, \mu, \theta, \tau)\) for \(r = 1, \ldots, N - n\) are generated using a simple rejection algorithm (for example see Smith and Roberts (1993)). For \(r = 1, \ldots, N - n\) keep generating independent observations from \(N(\theta_j, 1/\tau_j)\) to form the \(J\)-tuple \((x_1^{n+r}, \ldots, x_J^{n+r})\), until \(\text{Minimum}\{x_1^{n+r}, \ldots, x_J^{n+r}\} > t^{n+r}\). Note that this is equivalent to keeping on generating \(x_j^{n+r}\) from \(N(\theta_j, 1/\tau_j)\) till it is greater than \(t^{n+r}\), independently \(\forall j = 1, \ldots, J\).

Imputation of the pseudo-data \(X\) thus taken care of, we next proceed to generate the values of the nuisance parameter \(\mu\) and the lifetime parameter \(\theta, \tau\) from their respective conditional distributions with the help of the following Propositions.

**Proposition 2** \(\forall j = 1, \ldots, J\) \(\mu_j|t, A, t^+, \theta, \tau\) is a \(\rho\)-parameter Dirichlet with parameter vector \(\{n_j(A) + \alpha_j(A)\}_{A \in A_j}\).

**Proposition 3** \(\forall j = 1, \ldots, J\) \(\theta_j|t, A, t^+, \mu, \theta_{(-j)}, \tau \sim N\left((\sum_{i=1}^N x_j^i + \eta_j \psi_j)/(N + \psi_j), 1/(\tau_j(N + \psi_j))\right)\).

**Proposition 4** \(\forall j = 1, \ldots, J\) \(\tau_j|t, A, t^+, \mu, \theta, \tau_{(-j)} \sim \text{Gamma}(N/2 + \beta_j + 1/2, (1/2)(N + \psi_j)\theta_j^2 - (\sum_{i=1}^N x_j^i + \eta_j \psi_j)\theta_j + (1/2)(\sum_{i=1}^N (x_j^i)^2 + \eta_j^2 \psi_j) + \lambda_j)\).

Proposition 2 can be proved by multiplying the likelihood in (4) by the prior in (5), while Propositions 3 and 4 can be proved by multiplying the likelihood in (4) by the prior in (6), and trivial algebra, and hence the proofs are omitted.

Given a set of values for \(\mu, \theta, \tau\), say \(\mu^{(t)}, \theta^{(t)}, \tau^{(t)}\), the Gibbs sampler first imputes \(X\) following the algorithm described in the paragraph following the proof of Proposition 1. Note that this first requires generation of \(K^1, \ldots, K^n\), which determine the values of \(n_j(A)\)'s \(\forall j = 1, \ldots, J\) \(\forall A \in A\). Thus \(\mu^{(t+1)}\) can be generated next using Proposition 2. The sufficient
statistics of the imputed values of $X$ are then used as in Propositions 3 and 4 to generate the next value $\left(\theta^{(t+1)}, \tau^{(t+1)}\right)$ of $(\theta, \tau)$ in the chain.

3 Data Analysis

A reference Bayesian analysis of the data set in Reiser et al. (1995) is presented in this section under the assumption that the log-lifetimes of the Motherboard, Disc Drives and the Power Supply are mutually independent and have Normal distributions with $\theta_1, \theta_2, \theta_3$ and $\sigma_1, \sigma_2, \sigma_3$ as their respective means and standard deviations. Jeffreys’ prior, as mentioned §2.2, is used for this reference Bayesian analysis for both $\mu$ and $(\theta, \sigma)$, where $\mu = (\mu_1, \mu_2, \mu_3)$, $\theta = (\theta_1, \theta_2, \theta_3)$ and $\sigma = (\sigma_1, \sigma_2, \sigma_3)$.

Note that the masking set $\{1, 2\}$ is absent in the data set. Thus as in Mukhopadhyay (2004), the masking probabilities of this masking set viz. $\mu_1(\{1, 2\})$ and $\mu_2(\{1, 2\})$ are assumed to be 0 and dropped from further consideration from $\mu$, the vector of unknown masking probabilities. As mentioned §2.2, for the Jeﬀreys’ prior, all other $\alpha_j(A)$’s are set equal to $1/2$ $\forall A \in A_j$ for $j = 1, 2, 3$ in (5), and $\beta_j, \lambda_j$ and $\psi_j$ all set equal to 0 in (6) $\forall j = 1, 2, 3$.

Gibbs samples of size 10000 are obtained from $\pi(\mu, \theta, \sigma|D)$. Strictly speaking, the samples are obtained from $\pi(\mu, \theta, \tau|D)$ and then $\sigma_j$ is taken as $1/\sqrt{\tau_j}$. But note that this is equivalent to obtaining samples from $\pi(\mu, \theta, \sigma|D)$. Each one of the 10000 observations on $(\mu, \theta, \sigma)$ is an end product of 500 Gibbs iterations, where the previous observation is taken as the starting value of an iteration. That is, a chain of $10000 \times 500$ values of $(\mu, \theta, \sigma)$ are generated using C codes, and every 500-th value in the chain is regarded as a random observation from the respective joint posteriors.

Time series plots of the generated values of $\mu_j(A)$’s, $\theta_j$’s and $\sigma_j$’s suggested convergence of the MCMC. Marginal posterior densities of the $\theta_j$’s are computed using Gaussian kernel density estimates, and are plotted in Figure 1. Similarly the marginal posterior densities of $\sigma_j$’s are plotted in Figure 2. Table 1 provides the posterior mean, median, mode, standard deviation, and 95% Highest Posterior Density Credible Set (HPDCS) of $\theta$ and $\sigma$, computed from their respective marginal posterior densities.

Posterior means of the survival functions of the three components viz. Motherboard (MB), Disc Drives (DD) and Power Supply (PS) and the PS/2 system as a whole are plotted in Figure 3. From these plots it appears that in the log-run it is the Disc Drives which is most susceptible to failure, while the reliabilities of the Motherboard and Power Supply are more or less comparable to one another. As a result the PS/2 system reliability seems to be most influenced by the Disc Drives. This finding also conforms to the pattern observed in the raw data, where at least one of the causes of later failures is attributed to the Disc Drives. However since most of the failures are experienced early in its life of a PS/2 system, it is imperative that one compares the component hazard-rates during this infant mortality period.

Thus the posterior means of the component hazard rates are plotted in Figure 4. This figure reveals the story inherently latent in the observed data set. At the very early stage of an IBM PS/2’s life, it is the Motherboard which is most susceptible to failure, then comes the Power Supply, and in these early stages, Disc Drives are the most reliable components. But
as time progresses, though the Motherboard still remains the most likely culprit for failure, the Disc Drives slowly takes over the Power Supply in terms of instantaneous probability of failure. Then eventually the Disc Drives become most prone to failure, followed by the Power Supply, and in the long run it appears that the Motherboard is the most reliable component after all.

This finding is qualitatively similar to that of Mukhopadhyay (2004) though the quantitative values are quite different. The Disc Drive is detected as the most failure-prone component also by Basu et al. (1999, 2003). In contrast, the analysis in Reiser et al. (1995) concludes that the Disc Drives are as reliable as the Motherboards, with Power Supply being more reliable than the two.

So far nothing has been stated about the masking probabilities. Though these are nuisance parameters, they are nonetheless needed in estimating the diagnostic probabilities. That is for example consider the third failed system in the Reiser et al. (1995) data set. Its cause of failure is masked with MRS=$\{1,3\}$ and in retrospect one is interested in estimating the probability of its cause of failure being 1 or 3, which are respectively given by $q_3^1(\mu, \theta, \tau)$ and $q_3^2(\mu, \theta, \tau)$, where $q_i^j(\mu, \theta, \tau)$’s are as defined in Proposition 1.

Thus before turning the attention towards estimation of the diagnostic probabilities, one first needs to settle the issue of posterior inference of the masking probabilities. Marginal posterior densities of $\mu_1(A)$’s, $\mu_2(A)$’s and $\mu_3(A)$’s are provided respectively in Figures 5, 6 and 7. Their posterior summary measures like mean, median, mode, standard deviation, and 95% Highest Posterior Density Credible Set (HPDCS) are provided in Table 2. Note that $\mu_1(\{1,2\})$ and $\mu_2(\{1,2\})$ are absent from the analysis for the reason mentioned earlier in this section. It should be remarked that this marginal Bayesian inference of the masking probabilities are far more satisfactory than their profile likelihood based frequentist analysis provided in Mukhopadhyay (2004) vide. Table 2 and Figures 1 through 6 of Mukhopadhyay (2004).

Now coming back to the problem of estimating the diagnostic probabilities, note that there are three systems in the data set, whose cause of failures are masked viz. systems 3, 4, and 6 with respective MRS $\{1,3\}$, $\{1,2,3\}$ and $\{2,3\}$. Thus posterior inference of their diagnostic probabilities can be carried out by studying the posterior distributions of the respective $q_i^j(\mu, \theta, \tau)$’s. However the posterior mean of $q_i^j(\mu, \theta, \tau)$ can be estimated as the relative frequency of $K^i = j$ in the generated Gibbs sample. The posterior means of these diagnostic probabilities, thus computed, for the three systems with masked cause of failure are presented as a barchart in Figure 8.

4 Conclusions

A Bayesian analysis of masked series system lifetime data with log-Normal component lifetimes has been presented in this article. The masking probabilities are not subjected to the symmetry condition. A general parametric frequentist analysis of this problem without the symmetry assumption is available in Mukhopadhyay (2004). While a Bayesian analysis with component lifetimes belonging to general location-scale families of distributions under the symmetry assumption are provided by Basu et al. (2003). Bayesian analysis without the symmetry assumption with Weibull distributed component lifetimes is available in Kuo &
Yang (2000). This article attempted to fill the research gap when the component lifetimes follow the other popular location-scale family of log-Normal distributions, without the symmetry assumption. Though a frequentist analysis of this model can be carried out following the methodology developed in Mukhopadhyay (2004), a Bayesian analysis of this model is covered neither by Basu et al. (2003) nor Kuo & Yang (2000).

A second contribution of this article lies in its idea of augmenting the observed data set with latent variables of component lifetimes. While so far in the literature, suggestions have been made in imputing only the unknown causes of failures, here it is taken one more step forward and it is proposed that lifetimes of all the component lifetimes be generated as auxiliary variables. This facilitates the Bayesian analysis of not only the masked data, but that arising from the traditional competing risks model with no closed-form expressions for the survival functions as well. Interestingly the only method presently available in the literature for this old problem is the one recently developed by Basu et al. (2003).

With both the frequentist and Bayesian analysis of masked series system lifetime data for the popular component lifetime models being under their way, some remarks should be made about non-parametric or distribution-free approaches to this problem. Both Dinse (1982) and Miyakawa (1984) have considered this approach for simple cases under the tacit symmetry assumption. Flehinger et al. (1996, 1998) have relaxed the symmetry condition in their non-parametric considerations. But in one (1996) they do not have $T$ and in the other they assume the component hazard rates to be proportional. Thus a general non-parametric solution to this problem, be it frequentist or Bayesian, is still awaited in the literature.

REFERENCES


Table 1: Posterior Summary Measures of Lifetime Parameters

<table>
<thead>
<tr>
<th>Parameters</th>
<th>$\theta_1$</th>
<th>$\sigma_1$</th>
<th>$\theta_2$</th>
<th>$\sigma_2$</th>
<th>$\theta_3$</th>
<th>$\sigma_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>72.5063</td>
<td>24.4301</td>
<td>32.6263</td>
<td>10.4238</td>
<td>61.725143</td>
<td>19.0444</td>
</tr>
<tr>
<td>Median</td>
<td>58.0595</td>
<td>19.6234</td>
<td>28.1892</td>
<td>8.6600</td>
<td>46.8665</td>
<td>14.7273</td>
</tr>
<tr>
<td>Mode</td>
<td>42.0420</td>
<td>14.0140</td>
<td>23.0630</td>
<td>6.0066</td>
<td>34.2340</td>
<td>10.1700</td>
</tr>
<tr>
<td>95% HPDCS</td>
<td>(8.7409, 179.5800)</td>
<td>(1.8018, 62.3645)</td>
<td>(9.9700, 67.9704)</td>
<td>(2.2523, 23.4723)</td>
<td>(4.5045, 171.1487)</td>
<td>(1.8418, 51.3368)</td>
</tr>
</tbody>
</table>

Table 2: Posterior Summary Measures of Masking Probabilities

<table>
<thead>
<tr>
<th>Parameters</th>
<th>$\mu_1({1})$</th>
<th>$\mu_1({1, 2})$</th>
<th>$\mu_1({1})$</th>
<th>$\mu_1({1, 2, 3})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.6140</td>
<td>-</td>
<td>0.2156</td>
<td>0.1709</td>
</tr>
<tr>
<td>Median</td>
<td>0.6294</td>
<td>-</td>
<td>0.1637</td>
<td>0.1114</td>
</tr>
<tr>
<td>Mode</td>
<td>0.7017</td>
<td>-</td>
<td>0.0311</td>
<td>0.0227</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.2348</td>
<td>-</td>
<td>0.2006</td>
<td>0.1813</td>
</tr>
<tr>
<td>95% HPDCS</td>
<td>(0.1740, 1.0)</td>
<td>-</td>
<td>(0.0, 0.6311)</td>
<td>(0.0, 0.5568)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameters</th>
<th>$\mu_2({2})$</th>
<th>$\mu_2({1, 2})$</th>
<th>$\mu_2({2, 3})$</th>
<th>$\mu_2({1, 2, 3})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.5532</td>
<td>-</td>
<td>0.2397</td>
<td>0.2077</td>
</tr>
<tr>
<td>Median</td>
<td>0.5561</td>
<td>-</td>
<td>0.1982</td>
<td>0.1595</td>
</tr>
<tr>
<td>Mode</td>
<td>0.5661</td>
<td>-</td>
<td>0.0516</td>
<td>0.0316</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.2313</td>
<td>-</td>
<td>0.2033</td>
<td>0.1922</td>
</tr>
<tr>
<td>95% HPDCS</td>
<td>(0.1316, 0.9731)</td>
<td>-</td>
<td>(0.0, 0.6556)</td>
<td>(0.0, 0.6038)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameters</th>
<th>$\mu_3({3})$</th>
<th>$\mu_3({1, 3})$</th>
<th>$\mu_3({2, 3})$</th>
<th>$\mu_3({1, 2, 3})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.3714</td>
<td>0.2447</td>
<td>0.2021</td>
<td>0.1825</td>
</tr>
<tr>
<td>Median</td>
<td>0.3423</td>
<td>0.1963</td>
<td>0.1430</td>
<td>0.1206</td>
</tr>
<tr>
<td>Mode</td>
<td>0.2284</td>
<td>0.0388</td>
<td>0.0299</td>
<td>0.0267</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.2331</td>
<td>0.2135</td>
<td>0.1997</td>
<td>0.1917</td>
</tr>
<tr>
<td>95% HPDCS</td>
<td>(0.0, 0.8319)</td>
<td>(0.0, 0.6778)</td>
<td>(0.0, 0.6226)</td>
<td>(0.0, 0.5966)</td>
</tr>
</tbody>
</table>
Figure 1: Marginal Posterior Densities of $\theta_j$'s

Figure 2: Marginal Posterior Densities of $\sigma_j$'s

Figure 3: Survival Functions

Figure 4: Component Hazard Rates
Figure 5: Marginal Posterior Densities of $\mu_1(A)$

Figure 6: Marginal Posterior Densities of $\mu_2(A)$

Figure 7: Marginal Posterior Densities of $\mu_3(A)$

Figure 8: Posterior Means of Diagnostic Probabilities